

Математическая модель оптимального размещения информационных ресурсов по узлам распределённой информационной системы на базе трехуровневой клиент-серверной архитектуры без учета влияния блокировок

А. Н. Скоба, Айеш Ахмед Нафеа Айеш (Ирак)

*Южно-Российский государственный политехнический университет (НПИ)
им.М. И. Платова, Новочеркасск*

Аннотация: В данной статье, с использованием аппарата замкнутых экспоненциальных сетей массового обслуживания (СеМО), решена задача оптимального размещения распределённой базы данных (РБД) в распределённой информационной системе на базе локальной вычислительной сети (ЛВС) с использованием трехуровневой клиент-серверной архитектуры по критерию минимума среднего времени реакции системы на запросы пользователей. Приведены результаты численных экспериментов.

Ключевые слова: распределённая информационная система, тонкий-клиент, сервер-приложений, сервер баз данных, концептуальная модель, матрица переходных вероятностей, интенсивности обслуживания, стационарная вероятность, уравнение глобального баланса, среднее время реакции системы на запросы “тонких-клиентов”.

Математическая постановка задачи. Согласно [1] задача ставится следующим образом. Имеется распределённая информационная система (РИС) реализованная в среде локальной вычислительной сети (ЛВС), включающая: множество клиентских компьютеров (клиентских приложений) $-A=\{A_1, \dots, A_s, \dots, A_n\}$; множество серверов положений (СП) $-SP=\{SP_1, \dots, SP_s, \dots, SP_n\}$; множество серверов баз данных - $DB=\{DB_1, \dots, DB_s, \dots, DB_n\}$; множество интенсивностей формирования запросов клиентскими приложениями к программному обеспечению делового анализа $-A=\{\lambda_1, \dots, \lambda_s, \dots, \lambda_n\}$; множество запросов клиентских приложений на инициализацию соответствующих бизнес-приложений $-Q=\{Q_1, \dots, Q_b, \dots, Q_q\}$; множество бизнес-приложений $-r=\{r_1, \dots, r_b, \dots, r_q\}$; множество отношений (распределённая база данных-РБД) $-R=\{R_1, \dots, R_j, \dots, R_d\}$; множество объемов отношений $-V=\{V_1, \dots, V_j, \dots, V_d\}$; скорости считывания данных в СП $-VSP=\{VSP_1, \dots, VSP_s, \dots, VSP_n\}$; скорости записи данных в СП $-DSP=\{DSP_1, \dots, DSP_s, \dots, DSP_n\}$; скорости считывания данных в серверах БД $-VDB=\{$

$VDB_1, \dots, VDB_s, \dots, VDB_n$; скорости записи данных в серверах БД $-DDB = \{DDB_1, \dots, DDB_s, \dots, DDB_n\}$; производительность процессов СП $-PSP = \{PSP_1, \dots, PSP_s, \dots, PSP_n\}$; производительность процессоров серверов баз данных $-PDB = \{PDB_1, \dots, PDB_s, \dots, PDB_n\}$; скорость передачи файлов по каналу связи $- \theta$; постоянная задержка при передаче данных по каналу связи $- \theta_0$;

постоянная задержка при обработке данных в СП $- \alpha_1$; постоянная задержка при обработке данных в сервере БД $- \alpha_2$; матрица вероятностей формирования запросов клиентскими приложениями $-$

$F = \|A_{s,l}\| = \|f_{sl}\|, (s = \overline{1, n}, l = \overline{1, q})$, где элемент f_{sl} представляет собой вероятность того, что s -й тонкий клиент сформировал l -й запрос, причём $\sum_{l=1}^q f_{sl} = 1, (s = \overline{1, d})$;

матрица $\Omega = \|\omega_{l,j}\| = \|\omega_{ij}\|, (l = \overline{1, q}, j = \overline{1, s})$, где

$$\omega_{ij} = \begin{cases} 1, \text{ если } l\text{-й запрос инициализирует } j\text{-е бизнес-приложение,} \\ 0, \text{ в противном случае;} \end{cases}$$

матрица распределения бизнес-приложений по СП $-$

$Y = \|y_{r_j, SP_i}\| = \|y_{ji}\|, (j = \overline{1, s}, i = \overline{1, n})$, где

$$y_{ji} = \begin{cases} 1, \text{ если } r_j \in SP_i, \\ 0 - \text{ в противном случае, причём } \sum_{i=1}^n y_{ji} = 1, (j = \overline{1, n}); \end{cases}$$

матрица объемов считываемой информации $B = \|b_{r_j, R_j}\| = \|b_{ij}\|, (i = \overline{1, s}, j = \overline{1, d})$, где b_{ij} – объем считываемой информации по SQL запросу, сформированного бизнес-приложением r_j к отношению R_j ;

$$\delta(B) = \|\delta(b_{r_j, R_j})\| = \|\delta_{ij}\|, (i = \overline{1, s}, j = \overline{1, d}), \text{ где}$$

$$\delta_{ij} = \begin{cases} 1, \text{ если бизнес-приложение } r_j \text{ требует для своего выполнения} \\ \text{отношения } R_j, \\ 0, \text{ в противном случае;} \end{cases}$$

матрица объемов информации, полученной после процессорной обработки бизнес-приложениями - $\bar{B} = \|\bar{b}_{r_j}\| = \|\bar{b}_{ij}\|, (i = \overline{1, s}, j = \overline{1, d})$, где \bar{b}_{ij} – объем информации получаемой после процессорной обработки бизнес-приложения r_j , которое обращалось к отношению R_j ; матрица распределения отношений по серверам БД – $X = \|x_{R_j, DB_k}\| = \|x_{jk}\|, (j = \overline{1, d}, k = \overline{1, n})$, где

$$x_{jk} = \begin{cases} 1, \text{ если } R_j \in DB_k, \\ 0 - \text{ в противном случае, причем } \sum_{k=1}^n x_{jk} = 1, (j = \overline{1, d}). \end{cases}$$

Допущения модели.

- 1) Число СП совпадает с числом серверов БД, т.е. $|SP| = |DB|$;
- 2) рассматривается случай, когда для инициализации бизнес-приложения $r_j (j = \overline{1, s})$ требуется отношения $R_k (k = \overline{1, d})$ такие, что $r_j \in SP_i \Leftrightarrow R_k \in DB_i, (i = \overline{1, n})$.

Требуется найти такое размещение РБД по серверам баз данных, что бы среднее время реакции РИС на запросы «тонких клиентов» было бы минимально.

Концептуальная модель. В работе [1] была представлена оригинальная, отсутствующая в теории массового обслуживания, концептуальная модель функционирования РИС на базе трёхуровневой клиент-серверной архитектуры, включающая: прибор – P_0 , моделирующий работу канала передачи данных канала; буферные памяти канала, предназначенные для хранения запросов пользователей – $BPC_1, \dots, BPC_s, \dots, BPC_n$; приборы – $DB_1, \dots, DB_s, \dots, DB_n$, моделирующие работу серверов БД; буферные памяти серверов БД – $BDB_1, \dots, BDB_s, \dots, BDB_n$; приборы – $SP_1, \dots, SP_s, \dots, SP_n$, - моделирующие работу СП; буферные памяти СП – $BSP_1, \dots, BSP_s, \dots, BSP_n$. Концептуальная модель РИС представлена на рис. 1.

$\dots, i_{2n+s+1,s}, \dots, i_{2n+2,n}, \dots, i_{2n+s+1,1}, \dots, i_{2n+s+1,s}, \dots, i_{2n+s+1,n}, \dots, i_{3n+1,1}, \dots, i_{3n+1,s}, \dots, i_{3n+1,n})$
 $\}, i = \overline{1, s}$, где

$$i_{1s} = \begin{cases} 1, \text{ если } s\text{-й пользователь (тонкий клиент, } s = \overline{1, n}) \text{ находится в} \\ \text{активном состоянии (формирует запрос на инициализацию} \\ \text{бизнес - приложения),} \\ 0 - \text{ в противном случае;} \end{cases}$$

$\{i_{sr}, s = \overline{2, n+1}, r = \overline{1, n}\}$ – описывает очереди к каналу и состояние канала, где i_{sr} – количество запросов r -го пользователя (тонкого клиента) в s -ой буферной памяти канала и на обслуживании в канале;

$\{i_{sr}, s = \overline{n+2, 2n+1}, r = \overline{1, n}\}$ – описывает очереди запросов к серверам приложений и состояния серверов приложений, где i_{sr} – количество запросов r -го пользователя в буферной памяти s -го сервера приложений и на обслуживании в s -ом сервере приложений;

$\{i_{sr}, s = \overline{2n+2, 3n+1}, r = \overline{1, n}\}$ – описывает очереди к серверам баз данных и состояние серверов баз данных, где i_{sr} – количество сообщений r -го пользователя в буферной памяти s -го сервера баз данных и на обслуживании в s -ом сервере баз данных.

При этом имеют место следующие ограничения:

$$1) \sum_{k=1}^{3n+1} \sum_{r=1}^n i_{kr} = n \cdot r \cdot d;$$

$$2) \sum_{k=1}^{3n+1} i_{kr} = r \cdot d.$$

Представляющие интерес характеристически РИС определяются стационарными вероятностями состояний сети. Пусть $P(\bar{i})$ – стационарная вероятность того, что сеть находится в состоянии \bar{i} , где $\bar{i} = (i_{11}, \dots, i_{1r}, \dots, i_{1n}, \dots, i_{3n+1,1}, \dots, i_{3n+1,r}, \dots, i_{3n+1,n})$. В работах [2-4] было показано, что процесс изменения состояний такой сети описывается однородным регулярным

марковским процессом, и уравнение глобального баланса для стационарного режима функционирования такой сети, согласно [5-6] будет иметь вид:

$$\sum_{k=1}^{3n+1} \sum_{r=1}^n P(\bar{i}) \mu_{kr} = \sum_{l=1}^{3n+1} \sum_{k=1}^{3n+1} \sum_{r=1}^n P(\bar{i} + \bar{1}_{lr} - \bar{1}_{kr}) \mu_{lr} P_{lk}(r),$$

где $\mu_{sr}(r = \overline{1, n}, s = \overline{1, 3n+1})$ – интенсивность обслуживания в s -м центре сообщения r -го пользователя; $P_{lk}(r)$, ($l = \overline{1, 3n+1}, k = \overline{1, 3n+1}, r = \overline{1, n}$) – вероятность того, что сообщение r -го пользователя после обслуживания в l -м центре попадёт в k -й центр; $\bar{1}_{sr}(s = \overline{1, 3n+1}, r = \overline{1, n})$ – вектор, в s -ой координате которого на r -ом месте стоит 1, а все остальные значения равны нулю.

Подробный расчет стационарных вероятностей состояний сети $P(\bar{i})$ приведен в работах [2 - 4].

Конструирование элементов матриц переходных вероятностей $P = \|P_{ik}(r)\|, (r = \overline{1, n}; i, k = \overline{1, 3n+1})$ и интенсивностей обслуживания $\mu_{sr}(s = \overline{1, 3n+1}, r = \overline{1, n})$ представлено в работах [1-3].

Расчёт среднего времени реакции системы на запросы «тонких клиентов» производится по формуле:

$$\bar{T} = \left(\frac{1}{\sum_{s=1}^n \lambda_s} \right) \sum_{s=1}^n \lambda_s \bar{T}_s,$$

где $\lambda_s(s = \overline{1, n})$ – интенсивность формирования запросов s -м «тонким клиентом»; $\bar{T}_s(s = \overline{1, n})$ – среднее время реакции системы на запрос s -го

«тонкого клиента». Величину \bar{T}_s определим, как $\bar{T}_s = \frac{\bar{N}_s}{\bar{\lambda}_s}$, где

$\bar{N}_s(s = \overline{1, n})$ – среднее количество запросов s -го «тонкого клиента»;

$\bar{\lambda}_s(s = \overline{1, n})$ – средняя интенсивность формирования запросов s -м «тонким

клиентом». Величины \bar{N}_s и $\bar{\lambda}_s$ определим, как: $\bar{N}_s = 1 - P_s(d)$, $\bar{\lambda}_s = \lambda_s P_s(d)$, где $P_s(d)$ – вероятность того, что s -й «тонкий клиент» находится в активном состоянии (формирует запрос). Расчёт величины $P_s(d)$ аналогичен расчёту величины $P_s(1)$ для информационной системы на базе файл-серверной архитектуры с простым типом запросов [2].

Аналогично работе [3], расчёт величины \bar{T} также по существу сводится к расчёту нормализующей константы $G(N_1, \dots, N_n)$, для вычисления которой был использован рекуррентный метод Бузена [7,8].

Результаты численных экспериментов. В работе [2] представлен разработанный авторами эвристический алгоритм решения задачи об оптимальном размещении распределённой базы данных (РБД) по узлам ЛВС по критерию минимума среднего времени реакции системы на запросы пользователей. Данный алгоритм был программно реализован на языке C# при следующих исходных данных: скорость считывания в s -м узле $\{VSP_s, VDB_s\} \in [60000; 100000] \text{KB/сек}, s = \overline{1, n}$; скорость записи в оперативную память s -го узла $\{DSP_s, DDB_s\} \in [1 \times 10^7; 3 \times 10^7] \text{KB/сек}, s = \overline{1, n}$; производительность процессора s -го узла $\{PSP_s, PDB_s\} \in [2,5 \times 10^9; 3,5 \times 10^9] \text{операций/сек}, s = \overline{1, n}$; скорость передачи данных по каналу связи $\theta \in [1000; 10000] \text{KB/с}$; постоянная задержка при передаче по каналу $\theta_0 = 3 \times 10^{-4} \text{с}$; постоянная задержка при обработке в узле $\alpha_1 = \alpha_2 = 3 \times 10^{-6}$; объём j -го отношения $V_j \in [50000; 150000] \text{KB}, j = \overline{1, d}$; объём считываемой информации $b_{lj} \in [1; 1000] \text{KB} (l = \overline{1, q}, j = \overline{1, d}), j = \overline{1, d}$ по l -му запросу на чтение из j -го отношения, по l -му запросу на чтение из j -го отношения; $\bar{b}_{lj} \in [1; 500] \text{KB} (l = \overline{1, q}, j = \overline{1, d})$ – объём информации, получаемый после процессорной обработки по l -му запросу на чтение из j -го отношения.

Результаты численного моделирования при условии отсутствия блокировок в системе приведены в таблице № 1.

Таблица №1

Результаты машинных экспериментов

Размерность задачи $n \times d \times q$	Начальное значение	Число итераций МПП	Значение $\bar{T}^{(r)}$	Время решения задачи МПП, с	Число итераций ЭА	Значение $\bar{T}^{(n)}$	Время решения задачи ЭА, с
3x4x5	4,1786	81	2,6875	6,13	3	2,8603	3,43
6x8x10	2,6754	6 ⁸	1,2348	156	6	1,2023	23,01
8x13x15	4,2876	8 ¹³	-	-	11	4,5498	48,22
10x15x20	3,5602	10 ¹⁵	-	-	26	0,6523	103,5

Здесь МПП – метод полного перебора; ЭА – эвристический алгоритм; $\bar{T}^{(r)}$ – среднее время реакции системы для оптимального размещения РБД, полученного МПП; $\bar{T}^{(n)}$ – среднее время реакции системы для оптимального размещения РБД, полученного ЭА.

Разработанная модель оптимального размещения РБД по узлам ЛВС на базе трёхуровневой клиент-серверной архитектуры также может быть использована при внедрении интегрированных информационно-справочных систем для различных предметных областей.

Литература

1. Скоба А.Н., Айеш Ахмед Нафеа Айеш. Математическая модель функционирования распределённой информационной системы на базе трёхуровневой клиент-серверной архитектуры // Инженерный вестник Дона. 2017. № 4. URL: ivdon.ru/ru/magazine/archive/n4y2017/4482.

2. Скоба А.Н., Состина Е.В. Математическая модель оптимального размещения распределенной базы данных по узлам ЛВС на базе файло-серверной архитектуры // Инженерный вестник Дона, 2015, №2 URL: ivdon.ru/ru/magazine/archive/n2y2015/2881.

3. Скоба А.Н., Состина Е.В. Математическая модель оптимального размещения распределенной базы данных по узлам ЛВС на базе двухуровневой клиент-серверной архитектуры. // Инженерный вестник Дона, 2015, №2 URL: ivdon.ru/ru/magazine/archive/n2y2015/2882.

4. Скоба А.Н., Панфилов А.Н. Модель оптимального размещения информационных ресурсов по узлам распределенной информационной системы предприятия на базе двухуровневой архитектуры “клиент-сервер” с учетом влияния блокировок // Изв. вузов. Электромеханика. 2007. Т. 60, № 2. С. 77-84.

5. Жожикашвили В.А., Вишневский В.М. Сети массового обслуживания. Теория и применение к сетям ЭВМ. - М.: Радио и связь, 1988. - 192с.

6. Вишневский В.М. Теоретические основы проектирования компьютерных сетей.- М.: Техносфера, 2003.- 512 с.

7. Chakka R., Harrison P.G. A Markov modulated multi-server queue with negative customers –the MM CPP/GE/c/LG-queue // Acta Informatika/-2001.- v.37. pp. 785-799.

8. Buzen J.P. Computational Algorithms for Closed Queueing Networks with Exponential Servers. Commun. ACM. 1983. Vol.16, №9.pp.527-531.

References

1. Skoba A.N., Ajesh Ahmed Nafea Ajesh. Inzhenernyj vestnik Dona, 2017, №4 URL: ivdon.ru/ru/magazine/archive/n4y2017/4482.

2. Skoba A.N., Sostina E.V. Inzhenernyj vestnik Dona (Rus), 2015, №2 URL:ivdon.ru/ru/magazine/archive/n2y2015/2881.



3. Skoba A.N., Sostina E.V. Inzhenernyj vestnik Dona (Rus), 2015, №2 URL: ivdon.ru/ru/magazine/archive/n2y2015/2882.
4. Skoba A.N., Panfilov A.N. Izv. vuzov. Jelektromehaniika. 2017. T.60, №2. pp.77-84.
5. Zhzhikashvili V.A., Vishnevskiy V.M. Seti massovogo obsluzhivaniya. Teoriya i primeneniye k setyam EVM [Queueing networks. Theory and its network application]. M.: Radio isvyaz', 1988. 192 p.
6. Vishnevskiy V.M. Teoreticheskie osnovy proektirovaniya komp'yuternykh setey [Theoretical foundations of computer network design]. M.: Tekhnosfera, 2003. 512 p.
7. Chakka R., Harrison P.G. A Markov modulated multi-server queue with negative customers. The MM CPP/GE/c/LG-queue. Acta Informatika. 2001. v.37. pp.785-799.
8. Buzen J.P. Computational Algorithms for Closed Queueing Networks with Exponential Servers. Commun. ACM. 1983. Vol.16, №9. pp.527-531.