

УДК 004.031.42

ОПТИМИЗАЦИЯ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ЭЛЕКТРОННОГО БИЗНЕСА

Бельков Д.В., Незамова Л.В., Шардакова Л.Ю.

Донецкий национальный технический университет
кафедра вычислительной математики и программирования

E-mail: belkov@telenet.dn.ua

Аннотация

Бельков Д.В., Незамова Л.В., Шардакова Л.Ю. Оптимизация распределенной информационной системы электронного бизнеса. Статья посвящена разработке динамического метода оптимизации распределенной информационной системы электронного бизнеса. Предложена система управления, которая позволяет перераспределить функциональные модули среди серверов в случае дисбаланса нагрузки. Это обеспечивает повышение эффективности функционирования системы за счет уменьшения времени отклика и сбалансированности нагрузки.

Общая постановка проблемы

Интернет предоставляет потенциально большие возможности электронного бизнеса, что делает актуальной разработку новых распределенных информационных систем. На этапе создания системы требуется разместить ее функциональные модули среди серверов. Возникает оптимизационная задача, в которой множество вершин графа $G(V,E)$ необходимо разделить на p несвязанных подмножеств $S_1, S_2, \dots, S_i, S_p$, где p - количество серверов: $\cup_i S_i = V$, $S_i \cap S_j = 0$, $i \neq j$. При эффективной стабильной работе системы загрузка серверов должна быть сбалансированной, а время отклика сети – минимальным. Для балансировки загрузки в каждом из подмножеств суммарный вес вершин должен быть примерно равным: $|S_i| \approx |V| / p$. Чтобы минимизировать время отклика сети количество обрезанных ребер должно быть минимальным. Данная задача может быть решена с помощью методов, описанных в работе [1]. Однако эти методы являются статическими, при их использовании возможны задержки и блокировки выполнения функций, заранее распределенных по серверам. Для преодоления этого недостатка необходимо разработать динамический метод оптимизации системы, который допускает миграцию функций по серверам в зависимости от сетевой нагрузки.

Данная статья посвящена разработке динамического метода оптимизации распределенной системы электронного бизнеса. В этом методе используется система управления балансировкой загрузки серверов, которая при обнаружении дисбаланса загрузки должна переназначить функциональные модули среди серверов. Это обеспечивает повышение эффективности функционирования системы за счет уменьшения времени отклика и балансировки загрузки. Целью статьи является построение структурной схемы системы управления балансировкой загрузки серверов распределенной системы. В работе необходимо решить следующие основные задачи: построить структурную схему системы управления балансировкой загрузки серверов, определить вероятностный закон распределения фазовых координат.

Структурная схема системы управления балансировкой загрузки

Распределенную информационную систему будем рассматривать как объект, балансировкой загрузки которого можно управлять. Распределение загрузки объекта в состоянии s характеризуется вектором $X^s = \{X_1^s, X_2^s, \dots, X_i^s, \dots, X_n^s\}^T$, который назовем вектором состояния объекта управления, а его элементы – фазовыми координатами. В данной работе используются условные обозначения, приведенные в таблице 1.

Таблица 1 – Условные обозначения

n	Количество серверов информационной системы
μ_i (бит/с)	Быстродействие i -го сервера, $i=1,2,\dots,n$
m_i	Количество запросов к функциям, выполняемым i -м сервером, $i=1,2,\dots,n$
λ_{ir} (c^{-1})	Интенсивность запросов к функции r -го типа, выполняемой i -м сервером, $r=1,2,\dots, m_i$, $i=1,2,\dots,n$
ν_{ir} (бит)	Объем запроса к функции r -го типа, выполняемой i -м сервером, $r=1,2,\dots, m_i$, $i=1,2,\dots,n$
λ_i (бит/с)	Суммарная интенсивность запросов к функциям, выполняемым i -м сервером, $i=1,2,\dots,n$ $\lambda_i = \sum_{r=1}^{m_i} \lambda_{ir} \nu_{ir}$
X_i	Коэффициент загрузки i -го сервера, $i=1,2,\dots,n$ $X_i = \lambda_i / \mu_i$, $0 < X_i < 1$
X_{\max}	Максимальное значение X_i : $X_{\max} = \max(X_i)$, $i=1,2,\dots,n$
\bar{X}	Среднее значение X_i : $\bar{X} = \sum_{i=1}^n X_i / n$
c	Среднеквадратичное отклонение значений X_i : $\sigma = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / n}$
β	Степень дисбаланса загрузки: $\beta = c / X_{\max}$, $0 < \beta < 1$
s	Номер состояния объекта управления, $s=0,1,\dots$
q	Номер момента измерения значений X_i^s , $q=0,1,\dots$
Ω^s	Дискретный процесс балансировок загрузки, $s=0,1,\dots$
θ^s	Случайный интервал времени между балансировками загрузки, $s=0,1,\dots$
L^s	n -мерный дискретный процесс изменения фазовых координат во времени, $s=0,1,\dots$
I_q^s	n -мерный регулярный дискретный процесс измерения фазовых координат, $s=0,1,\dots$, $q=0,1,\dots$
Δ	Интервал времени между моментами измерения фазовых координат.
ε	Невязка: $\varepsilon = \beta^s - \beta^{s-1}$
δX^s	Вектор изменения фазовых координат с целью балансировки загрузки: $\delta X_i^s = X_i^{s+1} - X_i^s$

Индикатором дисбаланса загрузки серверов является параметр β . Чем меньше значение β , тем более равномерно распределена нагрузка.

Процесс Ω^s балансировки нагрузки переводит объект управления из одного состояния в другое. Между моментами балансировки состояние объекта остается неизменным. Поскольку переходы от состояния к состоянию происходят в случайные моменты времени, то объект управления имеет случайную структуру.

В каждом состоянии выполняется регулярный дискретный процесс I_q^s измерения фазовых координат. Моменты изменения состояний и моменты измерения фазовых координат под управлением процессов Ω^s и I_q^s показаны на рисунке 1. В течение случайного интервала θ^s между моментами перехода из состояния s в состояние $s+1$ регулярно с интервалом Δ выполняются измерения фазовых координат.

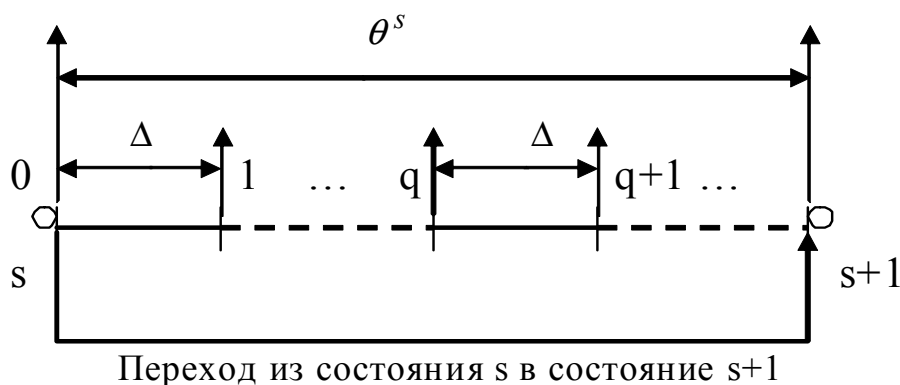


Рисунок 1 - Моменты изменения состояний и моменты измерения фазовых координат

В данной работе предлагается система управления балансировкой нагрузки, структурная схема которой показана на рисунке 2. На рисунке 3 показана укрупненная структурная схема системы управления.

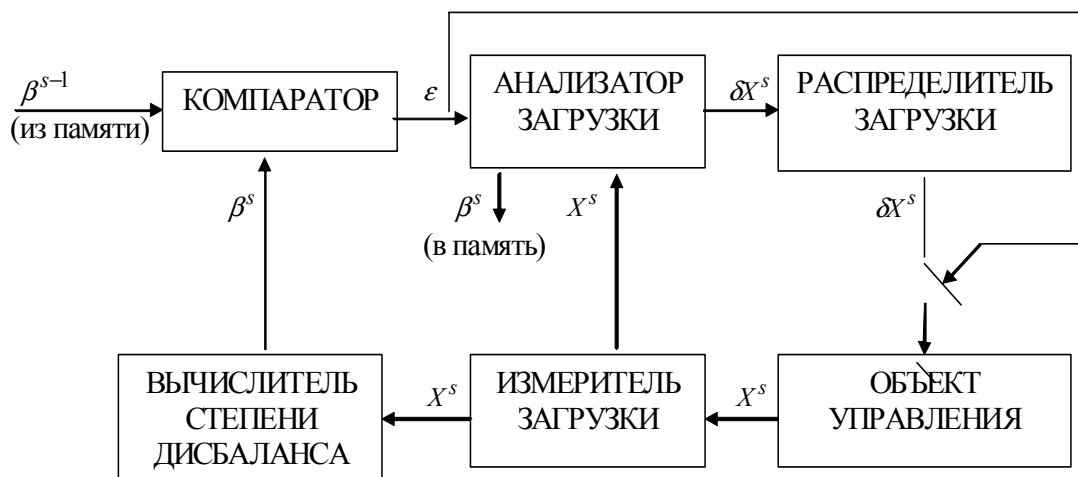


Рисунок 2 - Структурная схема системы управления балансировкой нагрузки

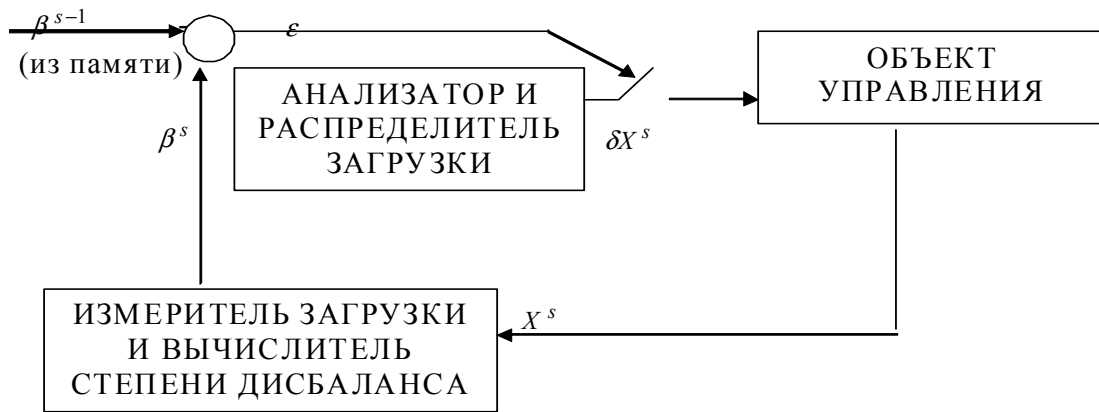


Рисунок 3 - Укрупненная структурная схема системы управления балансировкой загрузки

Измеритель загрузки в состоянии s выполняет измерение значений коэффициентов загрузки серверов. Результатом его работы является вектор $X^s = \{X_1^s, X_2^s, \dots, X_i^s, \dots, X_n^s\}^T$, который подается на вход вычислителя степени дисбаланса и на первый вход анализатора загрузки.

Вычислитель степени дисбаланса определяет параметр β^s по формуле $\beta^s = \sigma^s / X_{\max}^s$. Значение β^s поступает на вход компаратора. На другой его вход из памяти поступает значение β^{s-1} . На выходе компаратора формируется невязка ε : $\varepsilon = \beta^s - \beta^{s-1}$. Значение ε подается на второй вход анализатора загрузки.

Анализатор загрузки принимает решение о необходимости процесса балансировки в состоянии s . Если $\varepsilon < 0$, следовательно, распределение загрузки в состоянии s более равномерно, чем в состоянии $s - 1$. В этом случае значение β^{s-1} заменяется в памяти значением β^s и балансировка не выполняется. Если $\varepsilon = 0$, то распределение загрузки не ухудшилось и балансировка не выполняется. Если $\varepsilon > 0$, то распределение загрузки ухудшилось и в состоянии s необходимо выполнить балансировку загрузки. В этом случае анализатором загрузки формируются векторы X^{s+1} и ΔX^s . Вектор ΔX^s подается на вход распределителя загрузки, который выполняет перестановку функциональных модулей между серверами с целью загрузки серверов в соответствии с вектором X^{s+1} . Это приводит к более равномерной загрузке серверов.

При синтезе системы управления балансировкой загрузки необходимо знать математическое описание объекта управления. Для построения его модели требуется определить вероятностный закон распределения фазовых координат. Поскольку быстродействие процессоров (значение μ_i) не изменяется во времени, то достаточно найти распределение величины λ_i .

Известно [2], что сетевой трафик обладает свойством самоподобия (self-similarity) и может моделироваться ON/OFF процессом. Источник запросов на обработку событий переключается между ON состоянием, в котором происходит передача данных с постоянной интенсивностью, и OFF состоянием молчания. Длины ON и OFF периодов независимы и имеют вероятностное распределение с тяжелым хвостом (heavy tailed). Интенсивность сетевого трафика для Ethernet и Internet может моделироваться с помощью вероятностного распределения Парето [3].

В данной работе предполагается, что суммарная интенсивность запросов на обработку событий i -м процессором λ_i подчиняется распределению Парето, параметрами которого

являются значения α и β : $F(x) = P(X \leq x) = \begin{cases} 0, & x < \beta \\ 1 - (\beta/x)^\alpha, & x \geq \beta, \alpha < 2 \end{cases}$, где $P(X \leq x)$ - вероятность того, что значение случайной величины X не превысит заданное число x .

Процесс L^s изменения фазовых координат во времени является n -мерным случайным процессом, который протекает в каждом состоянии объекта управления. На ход процесса L^s в случайные моменты времени оказывает влияние процесс Ω^s . Поскольку процесс L^s может привести к возникновению дисбаланса, то он, в свою очередь, влияет на процесс балансировок. Процесс L^s для i -го процессора показан на рисунке 4. В течение случайного времени θ^s между моментами перехода из состояния s в состояние $s+1$ происходит изменение фазовых координат под управлением процесса L^s .

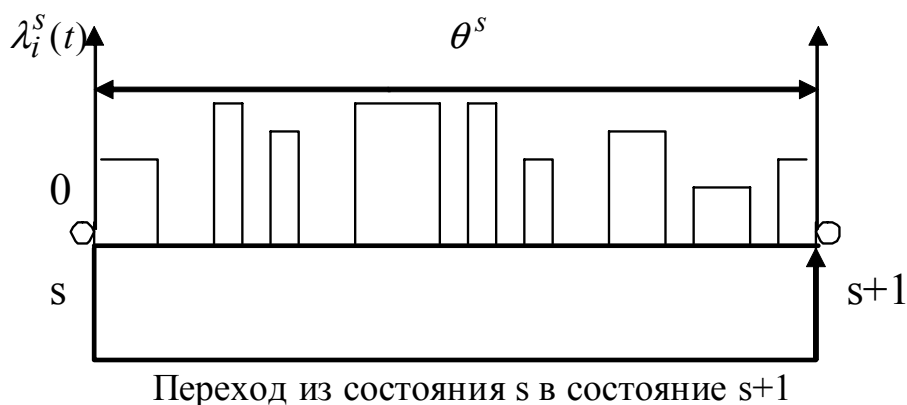


Рисунок 4 - Процесс L_i^s

Выводы

В работе получены следующие результаты:

1. предложен метод оптимизации распределенной системы электронного бизнеса на основе управления балансировкой загрузки ее серверов.
2. построена структурная схема системы управления балансировкой загрузки серверов.

Перспективными направлениями исследований являются: определение значений интервалов времени между моментами балансировок загрузки, определение оптимальных значений интервалов времени между моментами измерения фазовых координат.

Список литературы

1. Graph Partitioning Algorithms. Интернет-ресурс. – Режим доступа: [www/ URL: //www.stanford.edu/class/cs238](http://www.stanford.edu/class/cs238).
2. Мазалов В.В., Журавлев Д.Н. О методе кумулятивных сумм в задаче обнаружения изменения трафика компьютерных сетей. // Программирование, № 6. – 2002. – С. 55 – 61.
3. Кучук Г.А. Учет фрактальных свойств пульсирующего трафика. Интернет-ресурс. – Режим доступа: [www/ URL: //www.nestu.ru](http://www.nestu.ru).