

## **ОТЗЫВ**

**официального оппонента на диссертационную работу  
Шилова Сергея Николаевича на тему «Модели и алгоритмы  
балансировки нагрузки в кластерной системе с поддержкой  
механизма репликации», представленную на соискание  
ученой степени кандидата технических наук по специальности  
05.13.17 – «Теоретические основы информатики»**

### **Актуальность темы диссертационного исследования**

Развитие современных информационных технологий невозможно представить без теоретического исследования кластеризации компьютерных систем и практического использования его результатов. Актуальность данного направления исследований обусловлена возрастающей потребностью в эффективных и отказоустойчивых системах, обеспечивающих бесперебойную работу при обработке больших объемов информации. При этом необходимым компонентом для оптимального функционирования кластерной системы является специализированный программный комплекс, в котором одно из центральных мест занимает система балансировки нагрузки на обрабатывающие узлы. Применение распределенных хеш-таблиц для решения подобных задач является одним из наиболее эффективных подходов, позволяющим реализовать важнейшие преимущества кластерных систем, такие как масштабируемость и высокая доступность предоставляемого сервиса.

Один из наиболее успешных подходов к созданию распределенных хеш-таблиц, называемый консистентным хешированием, был предложен в работах D. Karger, E. Lehman, T. Leighton и др. (2001 г.), а в работе I. Stoica, R. Morris, D. Karger и др. (2001 г.) представлена его соответствующая реализация. Однако базовый вариант консистентного хеширования имеет ряд недостатков, что приводит к необходимости глубокого исследования данного направления и создания новых подходов к реализации распределенной хеш-таблицы как основы систем балансировки нагрузки.

Диссертационная работа Шилова Сергея Николаевича посвящена исследованию и разработке моделей и алгоритмов балансировки нагрузки в кластерной системе с поддержкой механизма репликации на основе технологии распределенной хеш-таблицы.

Для достижения цели исследования диссертантом в работе были поставлены и решены следующие задачи:

1. Разработка модели балансировки нагрузки на основе распределенной хеш-таблицы, соответствующей требованиям масштабируемости, быстродействия и равномерности распределения нагрузки.
2. Создание алгоритма поиска узла, ответственного за обработку элемента данных.
3. Создание алгоритмов репликации, повышающих стабильность, надежность и отказоустойчивость кластерной системы.
4. Разработка и исследование алгоритмов, созданных на основе предложенных моделей, реализация их в виде комплекса компьютерных программ, проведение статистических исследований для верификации и тестирования программных средств.

### **Общая оценка работы**

Диссертация состоит из введения, четырех глав, заключения, списка литературы, включающего 60 наименований, и 4 приложений. Общий объем диссертации составляет 143 страницы.

*Во введении* обоснована актуальность исследования, перечислены основные цели и задачи, определены научная новизна и практическая значимость.

*В первой главе* представлено понятие компьютерного кластера, приведена классификация компьютерных кластеров, описаны их преимущества по сравнению с одиночными ЭВМ. Дан обзор свойств

системы доменных имен, которая является одной из сфер применения разработанных в работе моделей и алгоритмов. Раскрыты понятия распределенной хеш-таблицы и консистентного хеширования, рассмотрены и проанализированы существующие реализации распределенных хеш-таблиц, описаны их преимущества и недостатки.

Проведенный анализ позволил автору получить четкую картину имеющихся методик и разработок в сфере распределенных хеш-таблиц и определить направление исследования.

*Во второй главе* диссертационной работы представлена разработанная модель балансировки нагрузки в кластерной системе. Описаны особенности системы доменных имен, такие как кэширование ресурсных записей. Приведено описание кластерной системы с точки зрения теории массового обслуживания, введено понятие выделенного обрабатывающего устройства, показано преимущество во времени обслуживания заявки на выделенном устройстве по отношению к случайному выбору устройств для обработки. Введено понятие таблицы вариантов распределения узлов кластерной системы, лежащей в основе разработанной модели балансировки нагрузки, представлены одноуровневая и двухуровневая модели организации таблиц вариантов распределения. Описан алгоритм поиска узла, ответственного за обработку заявки, проведена оценка временной сложности разработанных автором алгоритмов.

Особый интерес представляет разработанная автором оригинальная схема разбиения области значений базовой хеш-функции на области ответственности узлов кластерной системы. Данный подход позволил автору ввести таблицу вариантов распределения узлов для компактного хранения схемы разбиения и быстрого поиска узла, ответственного за обработку входящего запроса. Отдельно стоит отметить константную временную сложность алгоритма поиска ответственного узла, что крайне важно для общей производительности кластерной системы при масштабировании.

Также обращает на себя внимание тот факт, что автору удалось исключить работу с постоянным запоминающим устройством в процессе функционирования алгоритмов, ограничиваясь работой с оперативной памятью при обращении к таблице вариантов распределения, что положительно влияет на быстродействие программного комплекса балансировки нагрузки. Данное свойство позволяет рекомендовать предложенную модель для применения в кластерных системах, где большое значение имеет минимизация времени обработки запросов.

*В третьей главе* представлено понятие репликации данных, описаны виды репликации. На основе построенной модели балансировки нагрузки разработаны два алгоритма репликации: первый направлен на создание одной резервной копии для элементов данных, соответствующие ключи которых попадают в отрезки пересечения областей ответственности узлов, второй позволяет иметь произвольное четное число резервных копий на физически различных узлах кластерной системы. Оценена временная сложность разработанных алгоритмов.

Важным свойством двух предложенных автором алгоритмов репликации является их константная временная сложность по отношению к количеству обрабатываемых узлов. Таким образом, с ростом числа узлов данные алгоритмы не будут иметь возрастающего негативного влияния на общую производительность кластерной системы.

*Четвертая глава* посвящена тестированию и верификации программного комплекса, реализованного на основе разработанной модели балансировки нагрузки. В главе представлены результаты статистических исследований, полученные при участии уникальных DNS-запросов и DNS-запросов, принятых от реальных пользователей, проведен анализ полученных результатов. С помощью критерия согласия Пирсона была проверена гипотеза о соответствии распределения входящих DNS-запросов среди узлов кластерной системы равномерному закону распределения.

Проведенное исследование позволило автору вполне обоснованно сделать вывод о том, что реализованный на основе предложенной модели балансировки нагрузки программный комплекс в полной мере решает задачу равномерного распределения входящих запросов среди узлов кластерной системы.

*В заключении* подводятся итоги проведенного диссертационного исследования.

### **Оценка новизны полученных результатов**

*Научной новизной*, на мой взгляд, обладают все полученные в диссертационной работе результаты, а именно:

1. Модель балансировки нагрузки в кластерной системе, отличительными чертами которой является оригинальная схема разбиения области значений базовой хеш-функции, позволяющая избежать необходимости хранения значительного количества служебных данных на постоянном запоминающем устройстве для обеспечения функционирования алгоритмов. Также разработанная автором модель балансировки позволяет гарантированно определять узел, ответственный за обработку элемента данных, за один цикл поиска.

2. Алгоритм поиска узла, ответственного за обработку элемента данных, отличительной чертой которого является константная временная сложность, что позволяет исключить отрицательное влияние на масштабируемость кластерной системы.

3. Алгоритмы репликации, обеспечивающие наличие заданного числа резервных копий в зависимости от конкретных требований к надежности и отказоустойчивости системы, имеющие константную временную сложность и не ограничивающие масштабируемость системы.

4. Новые алгоритмы, построенные на основе разработанной автором модели распределения нагрузки, отличающиеся высоким

быстродействием за счет специальным образом подготовленной инфраструктуры программного комплекса.

### **Обоснованность и достоверность результатов**

Полученные в диссертации результаты и выводы обоснованы с позиций методологии исследования, основанной на корректном использовании математических и статистических методов обработки данных, теории вероятностей. Диссертация Шилова С.Н. четко структурирована, характеризуется научной строгостью и последовательностью изложения материала. В диссертации соискатель указал необходимые ссылки на авторов и источники заимствования отдельных материалов, использованных в работе.

*Достоверность результатов* подтверждается приведенными результатами вычислительного эксперимента, апробацией основных результатов на международных и всероссийских конференциях и семинарах. Основные научные результаты достаточно полно отражены в опубликованных автором 11 работах, в том числе в 4 статьях в изданиях, рекомендованных ВАК РФ. Получено 2 свидетельства о государственной регистрации программ для ЭВМ.

### **Теоретическая и практическая значимость работы**

*Теоретическая значимость* диссертации обусловлена ее новизной и заключается в развитии методов построения моделей и алгоритмов балансировки нагрузки в кластерных системах.

*Практическая значимость* работы определяется возможностью использования предложенных моделей и алгоритмов для построения программных комплексов балансировки нагрузки в различных кластерных системах с возможностью повышения отказоустойчивости за счет наличия алгоритмов репликации. Полученные в работе результаты использованы в сторонней организации (имеется акт об использовании результатов

диссертационного исследования), что также подтверждает их практическую значимость.

### **Замечания к работе**

В качестве замечаний к диссертационной работе необходимо отметить следующее:

1. В разделе 2.4 (стр. 61) автором недостаточно четко обоснован выбор 64-битной хеш-функции в качестве базовой хеш-функции распределенной хеш-таблицы.

2. В диссертационной работе под наиболее оптимальной балансировкой нагрузки понимается ее равномерное распределение среди узлов кластерной системы, однако данный подход не учитывает потенциальной разницы в уровнях производительности участвующих узлов, зависящих от аппаратного обеспечения.

3. Было бы желательно рассмотреть в диссертации системы массового обслуживания, динамика изменения состояний которых не обязательно описывается системой уравнений Колмогорова, так как информационные потоки не всегда являются простейшими.

4. В главе 4 явно не хватает графических иллюстраций (графиков теоретической функции распределения и совмещенной с ней гистограммы) для более наглядного представления результатов проведенного диссертантом исследования.

5. В диссертации встречаются неудачные литературные обороты, например, «Маловероятным по удовлетворительности, однако потенциально возможным...» на стр. 84.

Однако отмеченные недостатки не снижают достаточно высокого качества работы и не оказывают влияния на результаты диссертации.

## Заключение

Диссертация Шилова С.Н. является завершенной научно-исследовательской работой, выполненной на высоком научном уровне. Научные положения и выводы, сформулированные в диссертации, являются обоснованными. Основные результаты диссертации полно и своевременно опубликованы. Автореферат выполнен с соблюдением установленных требований, полностью и правильно отражает содержание диссертации.

Диссертация соответствует паспорту специальности 05.13.17 – «Теоретические основы информатики» по следующим областям исследований:

- п. 1 – Исследование, в том числе с помощью средств вычислительной техники, информационных процессов, информационных потребностей коллективных и индивидуальных пользователей;

- п. 2 – Исследование информационных структур, разработка и анализ моделей информационных процессов и структур;

- п. 14 – Разработка теоретических основ создания программных систем для новых информационных технологий.


По актуальности избранной темы, глубине проработки поставленных задач, научной и практической значимости полученных результатов и обоснованности выводов можно сделать заключение, что представленная диссертационная работа соответствует требованиям «Положения о присуждении ученых степеней», предъявляемым к кандидатским диссертациям, а ее автор, Шилов Сергей Николаевич, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 – «Теоретические основы информатики».

Официальный оппонент, заведующий кафедрой «Математического моделирования и информационных технологий» ФГБОУ ВПО «Тамбовский государственный университет имени Г.Р. Державина», доктор технических наук, профессор Арзамасцев Александр Анатольевич

Адрес: 392000, г. Тамбов, ул. Интернациональная, 33

Телефон: 8 (4752) 72-34-34

E-mail: arz\_sci@mail.ru

ФГБОУ ВПО «Тамбовский государственный университет имени Г.Р. Державина»  
Подпись:   
ЗАВЕРЯЮ  
Начальник управления кадров  
" 23 " 11 20 15 г.