МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра системного анализа и автоматического управления

РАЗРАБОТКА И АНАЛИЗ МАТЕМАТИЧЕСКОЙ МОДЕЛИ СИСТЕМЫ С РАСПРЕДЕЛЕННОЙ ПАМЯТЬЮ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 481 группы
направления 27.03.03 Системный анализ и управление
факультета компьютерных наук и информационных технологий
Шамхалова Данилы Евгеньевича

Научный руководитель
к. ф.-м. н., доцент ______________ И. Е. Тананко

Заведующий кафедрой
к. ф.-м. н., доцент ______________ И. Е. Тананко

Саратов 2026

ВВЕДЕНИЕ

Актуальность темы. В современной вычислительной технике системы с распределённой памятью (мультикомпьютеры, кластеры) занимают лидирующие позиции при решении ресурсоёмких научно-технических задач. Популярность таких архитектур (например, MPP или Beowulf-кластеров) обусловлена их высокой масштабируемостью, позволяющей наращивать производительность путём добавления отдельных узлов без необходимости переработки всей внутренней структуры системы.

Однако эффективное проектирование и эксплуатация таких систем невозможны без глубокого анализа их характеристик. Математическое моделирование на основе теории массового обслуживания (ТМО) и марковских процессов позволяет прогнозировать время реакции системы, загруженность процессоров и пропускную способность коммуникационных сред, не прибегая к дорогостоящим натурным экспериментам. Актуальность темы определяется необходимостью создания аналитических инструментов для оптимизации конфигурации распределённых систем и обеспечения требуемого уровня эффективности вычислений в условиях растущих информационных потоков.

Цель бакалаврской работы — разработать имитационную модель вычислительной системы с распределённой памятью и провести исследование ее характеристик.

Поставленная цель определила **следующие задачи**:

1. Изучить теоретические основы архитектур и топологий вычислительных систем с распределённой памятью.
2. Рассмотреть математический аппарат моделирования на основе цепей Маркова и теории систем массового обслуживания (СМО).
3. Построить открытую сеть СМО, адекватно описывающую структуру распределённой системы, включая распределитель запросов, процессоры и оперативную память.
4. Построить имитационную модель вычислительной системы с распределенной памятью.
5. Провести численные эксперименты для анализа влияния интенсивности входящего потока и производительности отдельных узлов на общие показатели сети.

Методологические основы разработки и анализа математической модели системы с распределенной памятью представлены в работах Ю. И. Митрофанов, В. В. Топорков, И. Е. Тананко, В. А. Романенко, Ф. М. Гафаров.

Теоретическая значимость бакалаврской работы Разработана и проанализирована модель вычислительной системы с распределенной памятью, формализован математический аппарат непрерывных цепей Маркова и построена концептуальная модель открытой СеМО из связанных приборов типа $M|M|1$. Модель адекватно описывает процессы параллельной обработки требований: от пуассоновского дробления запроса на фрагменты до их сборки в сборщике, что расширяет теоретическую базу для анализа накладных расходов на ожидание родственных подзадач.

Практическая значимость бакалаврской работы Разработанная модель вычислительной системы с распределенной памятью позволяет вычислять ключевые эксплуатационные показатели сети (время реакции, длину очередей, коэффициенты загрузки), выявлять «узкие места» и оптимизировать аппаратную конфигурацию реальных вычислительных кластеров без проведения дорогостоящих натурных испытаний.

Структура и объем работы. Бакалаврская работа состоит из введения, 4 разделов, заключения, списка использованных источников и одного приложения. Общий объем работы — 49 страниц, из них 40 страниц — основное содержание работы, включая 12 рисунков и 1 таблицу, список использованных источников — 23 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Теоретические основы построения систем с распределённой памятью» посвящен комплексному теоретическому анализу принципов организации, аппаратного построения и функционирования многопроцессорных вычислительных систем с распределенной структурой памяти.

В *подразделе 1.1* представлены архитектурные особенности, ключевые преимущества и топологические структуры систем с распределенной памятью, а также рассмотрены принципы организации независимых вычислительных узлов и механизмы межпроцессорного обмена данными через коммуникационную среду.

В *подразделе 1.2* приведено сравнительное сопоставление рассматриваемой архитектуры с альтернативными классами многопроцессорных вычислительных систем, в частности с системами с общей памятью (SMP), с подробным анализом ограничений их масштабируемости, различий в механизмах синхронизации и специфики применяемых парадигм программирования.

Второй раздел «Математический аппарат моделирования вычислительных систем» посвящен описанию теоретического фундамента, необходимого для формализации и анализа процессов в распределенных вычислительных системах. Раздел разделен на два ключевых направления: марковские процессы и теорию массового обслуживания (ТМО).

Рассматриваются цепи Маркова как основа моделирования, для которых основополагающим свойством является отсутствие последействия: Будущее состояние системы зависит исключительно от ее текущего состояния и никак не связано с предысторией процесса.

Для расчета сетей фундаментальную роль играют уравнения Колмогорова-Чепмена, которые связывают матрицы переходных вероятностей и позволяют прогнозировать поведение системы. На их основе выводятся дифференциальные уравнения Колмогорова, описывающие динамику вероятностей состояний системы во времени. При стремлении времени к бесконечности ($t \to \infty$) эти уравнения позволяют находить стационарные распределения вероятностей.

В качестве математического аппарата декомпозиции рассматривается теория массового обслуживания. Аппарат ТМО позволяет представить сложную вычислительную систему в виде взаимосвязанных компонентов: очередей, каналов передачи и обслуживающих приборов. Для краткого описания таких систем используется стандартная классификация Кендалла формата $A/S/c/B/Z$. Важнейшим инструментом анализа выступает формула Литтла:

$$n = \lambda u.$$

Которая связывает среднее число заявок в системе с интенсивностью потока и временем их пребывания.

В *подразделе 2.1* описано применение цепей Маркова для моделирования динамики систем. Рассматривается классификация по времени (дискретные и непрерывные цепи), а также особый случай — процесс размножения и гибели. В данной модели переходы возможны только между соседними состояниями, где «размножение» означает поступление новой заявки ($n \to n+1$), а «гибель» — окончание обслуживания текущей ($n \to n-1$). Это позволяет строго аналитическим путем определять средние характеристики сети, такие как длина очередей и время ожидания.

В *подразделе 2.2* описана структура систем массового обслуживания. Модель включает в себя входящий поток требований, очередь для ожидания, обслуживающие приборы (процессоры и модули памяти) и дисциплину обслуживания. В качестве простейших базовых моделей рассматриваются системы типов $M/M/1$, $M/M/c$ и $M/GI/1$, где выход одного прибора становится входом для другого, что позволяет объединить отдельные компоненты в единую сеть массового обслуживания (СеМО).

Третий раздел «Разработка математической модели вычислительной системы» посвящен детальной формализации, математическому описанию и алгоритмической реализации расчетных инструментов, предназначенных для исследования и комплексной оценки функционирования многопроцессорных вычислительных комплексов с распределенной структурой памяти.

В *подразделе 3.1* разработана аналитическая модель исследуемой вычислительной сети на основе теории систем массового обслуживания. В качестве базового математического элемента микропроцессора, модулей оперативной памяти и распределителя запросов обосновано применение однoприборных СМО типа $M|M|1$ с неограниченной очередью и пуассоновским входящим потоком. Описана структурная логика функционирования парной связки «микропроцессор — оперативная память» и выведены ключевые аналитические формулы для определения фундаментальных вероятностно-временных характеристик сети, включая математическое ожидание числа требований и среднее время реакции системы.

В *подразделе 3.2* сформирована концептуальная модель вычислительной системы, отражающая специфику параллельной обработки данных в мультикомпьютерах. Формализован процесс децентрализации вычислений, при котором неделимый внешний запрос от источника мгновенно дробится на фиксированное число уникальных фрагментов, распределяемых по узлам сети. Описаны принципы моделирования задержек обращения к локальной памяти, а также функционирование барьерной синхронизации в блоке сборки, где завершенные фрагменты ожидают объединения в исходное требование.

В *подразделе 3.3* приведено подробное описание алгоритмов функционирования разработанной имитационной модели, базирующейся на событийно-ориентированном принципе продвижения модельного времени. Автором последовательно формализована и математически описана логика сквозного жизненного цикла заявок в системе. Алгоритм детально раскрывает процессы генерации требований источником и их последующего расщепления на субальгоритмы, правила распределения фрагментов по локальным очередям приборов и условия начала их непосредственной обработки. Дополнительно структурированы механизмы высвобождения вычислительных ресурсов при наступлении моментов окончания обслуживания, а также алгоритмически описана работа барьерного сборщика, осуществляющего учет, финальное воссоединение родственных фрагментов в единый запрос и расчет результирующего времени реакции сети. Завершает описание подраздела математическая схема дискретно-событийного продвижения текущего времени к моменту совершения ближайшего запланированного изменения в системе.

В *подразделе 3.4* определена общая структура имитационной модели, а также зафиксирован исчерпывающий перечень ее входных и выходных параметров. Математические связи и вероятностные маршруты перемещения фрагментов между узлами жестко детерминированы представленной матрицей маршрутизации Θ. Сформированы перечни выходных статистических характеристик, регистрируемых в файловой структуре программы для последующей верификации и численного анализа.

Четвертый раздел «Анализ характеристик и эксперименты» посвящен практической верификации разработанной имитационной модели и комплексному исследованию вероятностно-временных характеристик вычислительной системы с распределенной памятью на основе серии численных экспериментов.

В *подразделе 4.1* приводится детальное описание методологии и условий проведения пяти контролируемых экспериментов, выполненных на различных структурных конфигурациях сети систем массового обслуживания (СеМО) при варьировании интенсивностей входящего потока и производительности узлов.

Пример 1 Рассмотрим открытую сеть систем массового обслуживания с $L = 10$ параллельными системами обслуживания $S_i, i = 1, ..., 10$. Вектор интенсивностей обслуживания требований в системах имеет вид:

$$\mu = (4.0 \ \ 4.0 \ \ 4.0 \ \ 4.0 \ \ 4.0 \ \ 2.0 \ \ 2.0 \ \ 2.0 \ \ 2.0 \ \ 2.0),$$

где интенсивность обслуживания для микропроцессоров равна 4.0, а для оперативной памяти — 2.0. Каждое требование, поступающее в сеть, делится на фрагменты, при этом максимальное количество фрагментов $b_{max} = 3$, а минимальное $b_{min} = 2$. Интенсивность входящего потока в сеть составляет $\lambda_0 = 1.0$.

Маршрутная матрица Θ имеет следующий вид:

$$\Theta = \begin{pmatrix}
0.0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.7 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.7 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 & 0.0 & 0.0 \\
0.7 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 & 0.0 \\
0.7 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 \\
0.7 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3 \\
0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0
\end{pmatrix}$$

Применяя разработанный алгоритм имитационного моделирования, получаем среднее число фрагментов, на которое делится требование, равное 2.962. Математическое ожидание длительности пребывания фрагментов в сети (время реакции СеМО) составляет 0.224, а математическое ожидание длительности времени сборки требований из фрагментов имеет значение 0.346 при интенсивности сборки 2.891. Математическое ожидание длительности пребывания требований в сети обслуживания и в системе сборки равно 0.396. Среднее число неполных требований в системе сборки составляет 0.345.

Низкая загруженность приборов позволяет входящим требованиям практически всегда делиться на максимальное число фрагментов ($2.962 \approx 3$). Невысокое время реакции СеМО указывает на избыточность ресурсов сети для входящего потока данной интенсивности.

Пример 2 Если запустить систему с параметрами: интенсивность входящего потока $\lambda_0 = 7.0$ при $L = 10$ параллельных системах обслуживания, то время реакции сети составит 80.625, из-за чего система практически перестает справляться со столь высокой нагрузкой. Для оптимизации функционирования СеМО и снижения задержек необходимо масштабировать вычислительные мощности, и поэтому рассмотрим открытую сеть систем массового обслуживания с увеличенным числом параллельных систем до $L = 14$. Вектор интенсивностей обслуживания требований в системах имеет вид:

$$\mu = (2.0 \;\; 2.0 \;\; 2.0 \;\; 2.0 \;\; 2.0 \;\; 1.0 \;\; 1.0 \;\; 1.0 \;\; 1.0 \;\; 1.0),$$

где интенсивность обслуживания микропроцессоров составляет 2.0, а оперативной памяти - 1.0. Коэффициент разветвления составляет от $b_{max} = 5$ до $b_{min} = 1$. Интенсивность входящего потока в сеть зафиксирована на высоком уровне и составляет $\lambda_0 = 7.0$.

Применяя метод имитационного моделирования, получаем, что среднее число фрагментов, на которое делится требование, равно 2.157. Математическое ожидание длительности пребывания фрагментов в сети (время реакции СеМО) составляет 1.646, а математическое ожидание длительности времени сборки требований из фрагментов имеет значение 0.749 при интенсивности сборки 1.335. Математическое ожидание длительности пребывания требований в сети обслуживания и в системе сборки равно 2.021. Математическое ожидание числа неполных требований в системе сборки составляет 5.228.

Повышение числа вычислительных узлов с 10 до 14 приводит к резкому снижению времени реакции сети (до 1.646). Данный результат экспериментально доказывает высокую эффективность масштабирования вычислительных систем с распределенной памятью при обработке интенсивных потоков информации.

В *подразделе 4.2* приведен анализ результатов моделирования, математически обосновывающий эффективность фрагментации задач. Здесь выявляются точки насыщения ресурсов и доказывается преимущество локализации памяти для предотвращения глобальных заторов при обработке интенсивных информационных потоков.

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы была разработана имитационная модель вычислительной системы с распределенной памятью и проведен комплексный анализ ее рабочих характеристик.

Для реализации поставленной цели были успешно решены следующие задачи:

Изучены теоретические основы построения вычислительных систем с распределенной памятью. Проведен детальный обзор архитектурных особенностей мультикомпьютеров и систем массово-параллельной обработки (MPP). В рамках сравнительного анализа с системами с общей памятью (SMP) были выделены их ключевые ограничения, связанные с пропускной способностью общей системной шины, что обосновало безальтернативность распределенных структур для сверхмощных вычислений.

Рассмотрен математический аппарат моделирования на основе однородных цепей Маркова с непрерывным временем и теории систем массового обслуживания. Изучены уравнения Колмогорова-Чепмена и процессы размножения и гибели, составляющие фундамент для вывода стационарных вероятностей состояний, средних длин очередей и времени ожидания требований в СеМО.

Построена концептуальная и математическая модель открытой сети СМО, адекватно описывающая структуру распределенной системы. Модель представляет собой сеть связанных приборов типа $M|M|1$ с неограниченными очередями и дисциплиной обслуживания FCFS. В разработанной модели формализован процесс параллельной обработки задач, включающий мгновенное дробление исходного требования на фиксированное число фрагментов b_w, их случайное распределение по микропроцессорам, обмен данными с локальной оперативной памятью через маршрутную матрицу Θ, а также последующую барьерную синхронизацию в блоке сборки.

Программно реализован алгоритм имитационного моделирования, базирующийся на событийно-ориентированном принципе продвижения модельного времени. Созданный программный комплекс (на языке Python) позволяет рассчитывать ключевые вероятностно-временные показатели сети: математическое ожидание числа требований в системе, коэффициенты использования приборов ψ и время реакции сети τ.

Проведены численные эксперименты, в ходе которых исследовалось влияние интенсивности входящего пуассоновского потока задач λ_0, производительности узлов μ и степени фрагментации на общие показатели эффективности.

На основе анализа результатов проведенных экспериментов были сформулированы следующие научно-практические выводы:

Стратегия деления поступающих требований на фрагменты показывает высокие результаты в повышении общей скорости обработки информации, обеспечивая существенное снижение среднего времени пребывания задачи в системе за счет распараллеливания вычислений.

Исследуемая распределенная архитектура обладает свойством высокой масштабируемости: увеличение количества доступных вычислительных узлов позволяет эффективно компенсировать рост интенсивности входящего потока требований, сохраняя стабильные показатели времени реакции сети.

Барьерная синхронизация в точке сборки выступает основным фактором, ограничивающим производительность, однако временной выигрыш от параллельной обработки фрагментов на независимых процессорах значительно превосходит задержки, вызванные ожиданием последнего «родственного» фрагмента.

Таким образом, разработанная математическая модель и программное обеспечение подтвердили свою теоретическую и практическую адекватность. Они могут быть использованы в качестве эффективного инструментария для оптимизации конфигураций реальных кластерных систем и прогнозирования их производительности в условиях растущих информационных потоков.

Основные источники информации:

1. Митрофанов, Ю. И. Анализ систем массового обслуживания: учебно-методическое пособие / Ю. И. Митрофанов, Е. С. Рогачко, Н. П. Фокина. — Саратов: Изд-во «Научная книга», 2009. — 55 с.
2. Топорков, В. В. Модели распределенных вычислений / В. В. Топорков. : Издат. Физмалит, 2004. — 320 с.
3. Романенко, В. А. Системы и сети массового обслуживания / В. А. Романенко. : Издат. Физмалит, 2021. — 68 с.
4. Гафаров, Ф. М. Параллельные вычисления / Ф. М. Гафаров, А. Ф. Галимянов. — Казань : Издат. Казан. ун-та, 2018. — 149 с.

5. Митрофанов, Ю. И. Анализ сетей массового обслуживания / Ю. И. Митрофано
— Саратов : Издат. Научная книга, 2005. — 175 с.
6. Строгалев, В. П. Имитационное моделирование / В. П. Строгалев, И. О. Толкаче
: Издат. МГТУ им. Н.Э.Баумана, 2015. — 295 с.