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

Кафедра системного анализа
и автоматического управления

ИССЛЕДОВАНИЕ МОДЕЛИ МАСШТАБИРУЕМОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 481 группы
направления 27.03.03 — Системный анализ и управление
факультета компьютерных наук и информационных технологий
Семёновой Анастасии Владимировны

Научный руководитель
к. ф.-м. н., доцент ______________ В. И. Долгов

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

Саратов 2020

ВВЕДЕНИЕ

Актуальность темы. Масштабируемость — свойство вычислительной системы, которое обеспечивает предсказуемый рост системных характеристик, например, числа поддерживаемых пользователей, быстроты реакции, общей производительности и пр., при добавлении к ней вычислительных ресурсов. Данным свойством обладают все современные системы управления базами данных, маршрутизаторы и т.п. [1].

Свойство масштабируемости актуально по двум основным причинам. Прежде всего, условия современного бизнеса меняются столь быстро, что делают невозможным долгосрочное планирование, требующее всестороннего и продолжительного анализа уже устаревших данных, даже для тех организаций, которые способны это себе позволить. Взамен приходит стратегия постепенного, шаг за шагом, наращивания мощности информационных систем. С другой стороны, изменения в технологии приводят к появлению все новых решений и снижению цен на аппаратное обеспечение, что потенциально делает архитектуру информационных систем более гибкой [2]. Одновременно расширяется межоперабельность, открытость программных и аппаратных продуктов разных производителей, хотя пока их усилия, направленные на соответствие стандартам, согласованы лишь в узких секторах рынка. Без учета этих факторов потребитель не сможет воспользоваться преимуществами новых технологий, не замораживая средств, вложенных в недостаточно открытые или оказавшиеся бесперспективными технологии. В области хранения и обработки данных требуется обязательное масштабирование систем [3].

Цель бакалаврской работы — изучение масштабируемых вычислительных систем и построение их математических моделей в виде систем массового обслуживания.

Поставленная цель определила **следующие задачи**:

1) изучение основных понятий теории массового обслуживания;

2) построение математической модели масштабируемой вычислительной системы в виде системы массового обслуживания;

3) разработка и программная реализация алгоритма для анализа модели масштабируемой вычислительной системы;

4) описание интерфейса, структуры разработанного приложения;

5) детальное описание входных данных, идентификаторов, методов вы-

числения;

6) проведение численных экспериментов;

7) исследование зависимостей изменения характеристик системы.

Методологические основы исследования модели масштабируемой вычислительной системы представлены в работах В. В. Воеводина, В. М. Вишневского, В. В. Топоркова, Ю. И. Митрофанова, И. Е. Тананко, В. И. Долгова, Е. С. Рогачко, А. А. Назарова, Е. С. Вентцеля, Г.П. Климова, Л. Клейнрока, А. Кофмана.

Практическая значимость бакалаврской работы. Разработанные алгоритм и программный комплекс для анализа масштабируемых вычислительных систем позволяют исследовать системы такого типа, достигать необходимые характеристики системы при различных входных параметрах.

Структура и объём работы. Бакалаврская работа состоит из введения, 6 разделов, заключения, списка использованных источников одного приложения. Общий объем работы – 66 страниц, из них 40 страниц – основное содержание, включая 16 рисунков и 2 таблицы, 19 страниц приложения, список использованных источников информации – 21 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

Первый раздел "Основные понятия теории массового обслуживания".

В подразделе 1.1 говорится о широком использовании систем массового обслуживания для моделирования стохастических систем [4, 5]. Рассказывается о целях использования, основных требованиях, выборе подходящей системы, сборе статистических данных.

В подразделе 1.2 описываются случайные последовательности требований. Внимание уделяется дисциплинам обслуживания, дается определение потоку требований, рассказывается о пуассоновском потоке требований, о его свойствах, об источниках требований и о видах систем массового обслуживания [6].

В подразделе 1.3 рассказывается о параметрах и характеристиках систем массового обслуживания. Вводятся следующие обозначения: λ — интенсивность входящего потока требований (интенсивность поступления требований), μ — интенсивность обслуживания требований одним прибором. Важным параметром СМО является коэффициент использования СМО, который подсчитывается следующим образом [4]:

$$\psi = \lambda\overline{\nu}/\kappa,$$

где $\overline{\nu}$ – математическое ожидание длительности обслуживания требований прибором, $\overline{\nu} = 1/\mu$.

Стационарные характеристики систем массового обслуживания [7, 8]:

p_n – стационарная вероятность пребывания в СМО точно n требований;

$\overline{n}$ — математическое ожидание (м. о.) числа требований в СМО;

$\overline{b}$ — м. о. числа требований в очереди СМО;

$\overline{g}$ — м. о. числа свободных приборов в СМО;

$\overline{h}$ — м. о. числа занятых приборов в СМО;

$\overline{u}$ — м. о. длительности пребывания требований в СМО;

$\overline{w}$ — м. о. длительности пребывания требований в очереди СМО (м. о. времени ожидания).

По определению [9]:

$$\overline{n} = \sum_{n=0}^{Z} np_n,$$

$$\overline{b} = \sum_{n=\kappa+1}^{Z} (n-\kappa)p_n,$$

$$\overline{g} = \sum_{n=0}^{\kappa} (\kappa - n)p_n,$$

$$\overline{h} = \sum_{n=0}^{\kappa} np_n + \kappa \sum_{n=\kappa+1}^{Z} p_n.$$

Подраздел 1.4 посвящен процессу размножения и гибели, его значимости при описании эволюции систем массового обслуживания. Приводятся основополагающие формулы для нахождения вероятностей состояний систем массового осблуживания [10]:

$$p_n = \frac{\lambda_0\lambda_1...\lambda_{n-1}}{\mu_1\mu_2...\mu_n}p_0 = p_0\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}},\ n = 1, 2, ...$$

Используя условие нормировки $\sum\limits_{i=1}^{n} p_n = 1$, получаем [11]:

$$p_o = \frac{1}{1 + \sum\limits_{n=1}^{\infty}\prod\limits_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}}.$$

Второй раздел "Элементарные системы массового обслуживания" посвящен основным простейшим системам , таким как: система без потерь $M/M/1$, система с ожиданием $M/M/\kappa$, система с одним обслуживающим прибором и ограниченной очередью $M/M/1/B$, система с потерями $M/M/\kappa$. Для каждой из них приводятся формулы для нахождения стационарных характеристик.

Третий раздел "Масштабируемые вычислительные системы".

В подразделе 3.1 дается определение масштабируемым вычислительным системам. Общим свойством таких систем, обеспечивающее возможность повышения производительности, является распределение имеющихся ресурсов в зависимости от загруженности системы. Различают вертикальное и горизонтальное масштабирование.

В подразделе 3.2 описывается математическая модель масштабируемой вычислительной системы. В качестве модели рассматривается система массового обслуживания $M/M/\kappa/B$ с переменным числом обслуживающих приборов κ и мест в очереди B. В данной системе ограниченная общая очередь и κ одинаковых обслуживающих приборов, работающих параллельно. Причем, обслуживающие приборы (процессоры) занимают объем памяти, равный e, а места для ожидания – d. В результате, общий объем памяти системы равен:

$$D = e\kappa + dB.$$

Будем считать, что $e >> d$.

Для рассматриваемой системы выведены следующие формулы.

Коэффициент использования СМО определяется следующим образом:

$$\psi = \frac{\lambda\overline{\nu}}{\kappa}.$$

В данном случае коэффициент использования системы может быть интерпретирован как м. о. доли занятых приборов, так как приборы (процессоры) имеют одинаковое распределение длительности обслуживания.

Вероятности пребывания в системе в стационарном режиме n требований определяются [7]:

$$p_n = \begin{cases} p_0 \frac{(\kappa\psi)^n}{n!}, & n \le \kappa, \\ p_0 \frac{\psi^n \kappa^\kappa}{\kappa!}, & \kappa \le n \le \kappa + B. \end{cases}$$

где

$$p_0 = \begin{cases} \left[\sum_{n=0}^{\kappa-1} \frac{(\kappa\psi)^n}{n!} + \frac{(\kappa\psi)^\kappa (1-\psi^{B+1})}{\kappa!(1-\psi)} \right]^{-1}, & \psi \ne 1, \\ \left[\sum_{n=0}^{\kappa-1} \frac{(\kappa\psi)^n}{n!} + \frac{(\kappa\psi)^\kappa}{\kappa!}(B+1) \right]^{-1}, & \psi = 1. \end{cases}$$

Математическое ожидание числа требований в системе вычисляется по формуле:

$$\overline{n} = \sum_{i=0}^{\kappa+B} i p_i = p_1 + p_2 + ... p_{\kappa+B} = \sum_{i=1}^{\kappa} p_0 \frac{(\kappa\psi)^i}{i!} + \sum_{i=\kappa+1}^{\kappa+B} p_0 \frac{\psi^i \kappa^\kappa}{\kappa!}.$$

Используя формулу Литтла, находим м. о. длительности реакции системы [12]:

$$\overline{\tau} : \overline{u} = \overline{\tau} = \frac{\overline{n}}{\lambda} = \frac{1}{\lambda} \Big(\sum_{i=1}^{\kappa} p_0 \frac{(\kappa\psi)^i}{i!} + \frac{(\kappa\psi)^i}{i!} + \sum_{i=\kappa+1}^{\kappa+B} p_0 \frac{\psi^i \kappa^\kappa}{\kappa!} \Big).$$

Математическое ожидание числа требований в очереди:

$$\overline{b} = \sum_{n=\kappa+1}^{B} (n-\kappa) p_n = \sum_{n=\kappa+1}^{B} (n-\kappa) p_0 \frac{\psi^n \kappa^\kappa}{\kappa!}.$$

Математическое ожидание длительности пребывания требований в очереди находится по формуле Литтла:

$$\overline{w} = \frac{\overline{b}}{\lambda} = \frac{1}{\lambda} \sum_{n=\kappa+1}^{B} (n-\kappa) p_0 \frac{\psi^n \kappa^\kappa}{\kappa!}.$$

Для оптимизации данной системы необходимо найти такое число обслуживающих приборов и мест в очереди, которое приводило бы к желаемому математическому ожиданию длительности реакции, не превышающее заданное значение $\overline{\tau} = \overline{u}$ (математическое ожидание длительности пребывания требований в СМО). Этого можно достигнуть путем изменения количества обслуживающих приборов и тем самым мест в очереди. Также отмечаем, что в таких системах при непрерывном увеличении интенсивности входящего потока требований (изменении нагрузки системы), можно найти оптимальное распределение ресурсов, необходимое для ее успешного функционирования.

В подразделе 3.3 рассматривается другой вид масштабируемой вычислительной системы, в которой в качестве математической модели берется система массового обслуживания типа $M/M/\kappa$ с ожиданием, и приводятся результаты исследования. Пусть l — стоимость пребывания одного требования в системе в единицу времени, а m — стоимость использования одного обслуживающего прибора в единицу времени. Тогда общая стоимостная функция будет равна:

$$S = \overline{u}l + \kappa m.$$

Под масштабируемостью рассматриваемой системы понимается изменение количества обслуживающих приборов в зависимости от требуемых значений стоимостной функции и средней длительности пребывания требований в СМО. В результате проведенного исследования данного типа масштабируемых систем делаем вывод о том, что полученная функция зависимости стоимостной функции от количества приборов практически линейная. Так как математическое ожидание длительности пребывания требований мало, стоимостная функция фактически зависит от количества приборов в системе. При увеличении числа обслуживающих приборов стоимостная функция достаточно быстро возрастает, поэтому для достижения оптимальной стратегии при использовании данных систем требуются дополнительные исследования.

Четвертый раздел "Алгоритм анализа масштабируемой вычислительной системы".

В подразделе 4.1 представлена структурная схема разработанного алгоритма анализа масштабируемой вычислительной системы.

В подразделе 4.2 описываются блоки алгоритма.

В первом блоке происходит обработка входных данных. Во втором задаются начальные условия для системы ($\kappa = 1, B = \frac{D - \kappa e}{d}$). Затем вычисляются стационарные характеристики системы. После их вычисления проверяется условие достижения заданного математического ожидания длительности реакции системы $\overline{\tau_0}$ ($\overline{\tau_0} \geq \overline{\tau}$). Если оно не выполнено, число обслуживающих приборов увеличивается на единицу, пересчитывается число мест в очереди и алгоритм переходит ко 2 шагу. Если же выполнено, выводится результат вычисления.

Пятый раздел "Описание программы для анализа масштабируемой вычислительной системы" посвящен описанию программы для оптимизации разработанного алгоритма.

В подразделе 5.1 рассказывается о назначении и структуре программного комплекса. Программный комплекс по данному алгоритму, разработанный на языке программирования $C\#$, позволяет при заданных параметрах системы вычислять ее стационарные вероятности и характеристики, такие как: м. о. числа требований в системе, м. о. числа требований в очереди, м. о. длительности пребывания требований в очереди. Также определяет оптимальное число приборов и мест в очереди, обеспечивающее необходимое м. о. длительности реакции СМО. Выводится значение максимально близкое к желаемому при заданных параметрах системы. Разработанная программа, кроме того, обеспечивает наглядное представление результатов и графически демонстрирует зависимость характеристик от изменения параметров.

В подразделе 5.2 приводится список идентификаторов и методов программного комплекса

В подразделе 5.3 описывается интерфейс и правила использования программы. Программный комплекс работает в диалоговом режиме. Пользователь вводит необходимые параметры для решения поставленной задачи. Данные проверяются на корректность и выполнение требуемых ограничений. Приводятся изображения интерфейса программы.

В подразделе 5.4 описывается подробная инструкция для пользователя по работе с программой. Приводятся изображения вывода сообщения о некорректности вводимых данных. Также представлен пример использования программы для гипотетической масштабируемой вычислительной системы.

Шестой раздел "Анализ зависимостей характеристик модели масштабируемой вычислительной системы" посвящен анализу зависимостей характеристик системы от входных параметров. Приводятся графики зависимостей изменения стационарных характеристик от изменения числа обслуживающих приборов и мест в очереди. В результате исследования делается вывод о том, что при увеличении числа обслуживающих приборов характеристики системы улучшаются. А при увеличении числа мест в очереди ухудшаются. Это происходит за счет того, что при данном изменении, увеличении мест в очереди, уменьшается память системы, тем самым и число обслуживающих приборов (прямо пропорциональная зависимость). Все стационарные характеристики при малом количестве приборов хуже, чем если бы приборов было больше. Поэтому для того, чтобы достигнуть требуемых характеристик и максимально уменьшить потерю требований (вероятность отказа), необходимо анализировать данные системы и разрабатывать алгоритмы для достижения поставленных целей.

ЗАКЛЮЧЕНИЕ

Система массового обслуживания с переменным числом обслуживающих приборов и мест в очереди используется в качестве математической модели масштабируемой вычислительной системы.

В данной бакалаврской работе были рассмотрены основные понятия теории массового обслуживания, изучены элементарные системы массового обслуживания, которые являются основой для изучения более сложных. Была построена и исследована математическая модель масштабируемой вычислительной системы, для этого были разработаны алгоритм и соответствующий программный комплекс. Также были построены графики зависимостей изменения стационарных характеристик системы от входных данных. Данные изменения характеристик изучены, по ним сделаны соответствующие выводы.

Результаты исследования данной системы могут быть применены для анализа масштабируемых вычислительных систем, для грамотного распределения имеющихся ресурсов с целью повышения производительности, минимизируя при этом потери.

Отдельные части бакалаврской работы были опубликованы и представлены на конференциях:

— XI Всероссийская конференция ИТО-Саратов 2019. Тема работы: "Использование программного комплекса для анализа масштабируемых вычислительных систем в процессе обучения бакалавров по направлению "Системный анализ и управление;

— Студенческая научная конференция факультета компьютерных наук и информационных технологий 2020 года. Тема работы: "Исследование математической модели масштабируемой вычислительной системы".

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Воеводин, В. В. Параллельное вычисление / В. И. Воеводин, В. Х. Кириллов. - С. : 2001. - 68 с.

2 Вишневский, В.М. Теоретические основы построения компьютерных сетей. / Вишневский, В.М. - М. : Техносфера, 2003. - 512 с.

3 Топорков, В.В. Модели распределенных вычислений. - М. 2004. - 320 с.

4 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.

5 Тананко, И. Е. Основы моделирования систем: Учебное пособие. - Саратов; ООО Издат. центр «Наука», 2018. - 116 с.

6 Долгов, В. И. Методы анализа сетей массового обслуживания: Учебно-методическое пособие / В. И. Долгов. - Саратов: Научная книга, 2009. - 34 с.

7 Вентцель, Е. С. Теория вероятностей и инженерные приложения/ Е. С. Вентцель. - М.: Наука, 1988, - 480 с.

8 Климов, Г. П. Теория вероятностей и математическая статистика. / Г. П. Климов. - М. : Изд-во Моск. ун-та, 1983. - 328 с.

9 Назаров, А. А. Теория массового обслуживания: учеб. пособие. / А. А. Назаров, А. Ф. Терпугов. - Т. : Изд-во НТЛ, 2004. - 228 с.

10 Рогачко, Е. С. Управляемые марковские процессы и их приложения в системном анализе: Учеб. пособие / Под ред. Ю. И. Митрофанова. - Саратов: Издательство "Научная книга 2008. - 108 с.

11 Клейнрок, Л. Теория массового обслуживания. / Л. Клейнрок. Пер. с англ. - М. : Машиностроение, 1979. - 432 с.

12 Кофман, А. Массовое обслуживание. Теория и приложения. / А. Кофман, Р. Крюон. Пер. с фран. - М. : Мир, 1965. - 303 с.