

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра системного анализа и автоматического управления

**МОДЕЛЬ МАССОВОГО ОБСЛУЖИВАНИЯ ДЛЯ АНАЛИЗА
ПРОИЗВОДИТЕЛЬНОСТИ И ЭНЕРГОСБЕРЕЖЕНИЯ
ДАТА-ЦЕНТРА**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 481 группы
направления 27.03.03 — Системный анализ и управление
факультета компьютерных наук и информационных технологий
Левوشي́на Егора Романовича

Научный руководитель

доцент, к. ф.-м. н.

Е. С. Рогачко

Заведующий кафедрой

к. ф.-м. н., доцент

И. Е. Тананко

Саратов 2024

ВВЕДЕНИЕ

Актуальность темы. Современные дата-центры представляют собой ключевые инфраструктурные элементы, обеспечивающие работу множества цифровых сервисов и приложений. Они являются основой для функционирования облачных вычислений, социальных сетей, потокового видео, онлайн-игр и множества других услуг, которые ежедневно используют миллионы пользователей по всему миру. В условиях стремительного роста объемов данных и увеличения требований к вычислительным мощностям, вопрос оптимизации работы дата-центров приобретает все большую актуальность. В связи с этим, важное значение имеет изучение моделей массового обслуживания, которые могут применяться для анализа производительности и энергоэффективности таких центров.

Цель бакалаврской работы — моделирование дата-центра системой массового обслуживания с целью анализа производительности и энергосбережения.

Поставленная цель определила **следующие задачи:**

1. Изучение структуры и основных процессов функционирования дата-центра;
2. Описание математической модели дата-центра в виде системы массового обслуживания с управлением числом обслуживающих приборов;
3. Изучение метода анализа систем массового обслуживания с управлением числом обслуживающих приборов;
4. Разработка алгоритма и программы для анализа модели дата-центра;
5. Проведение исследования функционирования дата-центра, его производительности и энергосбережения при различных параметрах системы.

Методологические основы моделирования дата-центров, в том числе с использованием моделей массового обслуживания представлены в работах Hwaiyu P. E. [1], Кальфа К. В. [2], Боссерта Ш. [3], Schwartz С. [4], Romani A. [5].

Практическая значимость бакалаврской работы. В ходе выполнения выпускной квалификационной работы была разработана программа для нахождения стационарных характеристик дата-центра. Программа может использоваться для решения задачи выбора параметров дата-центра с целью сокращения энергопотребления дата-центра.

Структура и объем работы. Бакалаврская работа состоит из введения, 5 разделов, заключения, списка использованных источников и одного приложения. Общий объем работы — 50 страниц, из них 45 страниц — основное содержание, включая 16 рисунков, список использованных источников информации — 21 наименование.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Описание дата-центра» посвящен описанию структуры дата-центра и его элементов. Дата-центр (ДЦ) — централизованный объект для обработки и хранения данных. Структура ДЦ может варьироваться в зависимости от потребностей и требований организации. Обычно дата-центр состоит из здания, в котором размещается оборудование, и основных функциональных блоков [1].

В подразделе 1.1 приведено описание трёхуровневой архитектуры ДЦ, иногда ее еще называют EoR [2]. Модель коммутации end-of-row предполагает обслуживание коммутатором трафика со всех серверов из нескольких стоек, расположенных в ряд. Эта архитектура предусматривает разделение функций и компонентов дата-центра на три уровня: уровень доступа, уровень распределения и уровень ядра.

Трёхуровневая архитектура позволяет достичь высокой производительности, масштабируемости и отказоустойчивости сети дата-центра, разделяя функции и обеспечивая оптимальное управление трафиком. Этот подход также облегчает сопровождение и расширение инфраструктуры дата-центра в соответствии с растущими потребностями организации [3].

Подраздел 1.2 посвящен анализу использования энергоресурсов дата-центра. В данном подразделе рассматриваются программы и компании для определения энергоэффективности ДЦ. Можно назвать несколько примеров, среди которых выделить программу ENERGY STAR (государственная программа агентства по защите окружающей среды) и некоммерческое объединение Green Grid. В рамках этих программ используются различные показатели оценки энергоэффективности вычислительных дата-центров.

Второй раздел «Модель дата-центра в виде системы массового обслуживания с управлением числом обслуживающих приборов» посвящен рассмотрению двух моделей ДЦ: одна будет использовать обычные серверы, а вторая оптимизированные по энергопотреблению серверы.

В подразделе 2.1 описывается модель обычного дата-центра, когда каждый из серверов либо включен и обрабатывает задание, либо включен и простаивает. Если занятый сервер завершает обработку задания, а очередь пуста, сервер переходит в режим ожидания. Как только новое задание назначается простаивающему серверу, сервер становится занятым.

Подраздел 2.2 посвящен описанию второй модели ДЦ, когда количество базовых серверов, которые всегда включены и потребляют то же количество энергии, что и в модели обычного ДЦ, задается числом n , а количество зарезервированных серверов, которые будут включены по требованию, описывается числом m . В этой модели предполагается, что процесс поступления новых требований является пуассоновским с интенсивностью λ , каждый сервер принимает только одно требование на обработку с экспоненциально распределенным временем обслуживания со средним значением $\frac{1}{\mu}$. Если зарезервированные серверы активированы, энергопотребление определяется аналогично как в модели обычного ДЦ, а именно, если сервер обрабатывает задание в данный момент, он потребляет $e_{busy} = 240$ Вт, простаивающий сервер потребляет $e_{idle} = 170$ Вт. Если серверы деактивированы, каждый сервер потребляет $e_{off} = 0$ Вт. Если очередь ожидающих обработки заданий имеет длину, превышающую θ_2 ($\theta_2 \in (0, m)$), m зарезервированных серверов включаются и остаются включенными до тех пор, пока общее количество заданий в системе не упадет до θ_1 ($\theta_1 \in (0, n)$). Переход между состояниями для каждого из зарезервированных серверов показан на рисунке 1.

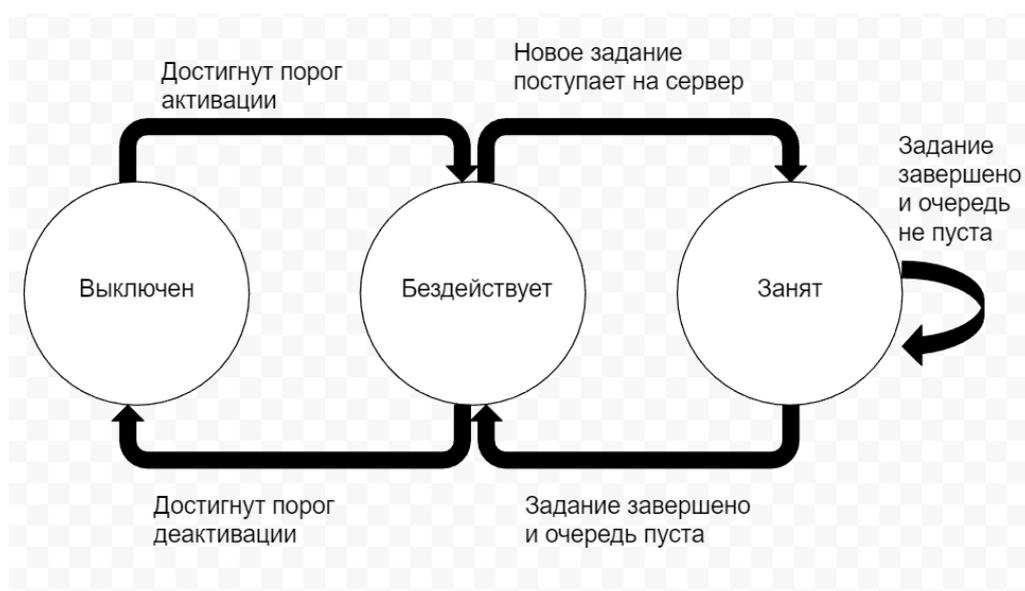


Рисунок 1 – Модель сервера с 3 состояниями.

Модель энергоэффективного дата-центра с параметрами θ_1 и θ_2 показана на рисунке 2. Состояние системы описывается парой (i, j) , где i - ко-

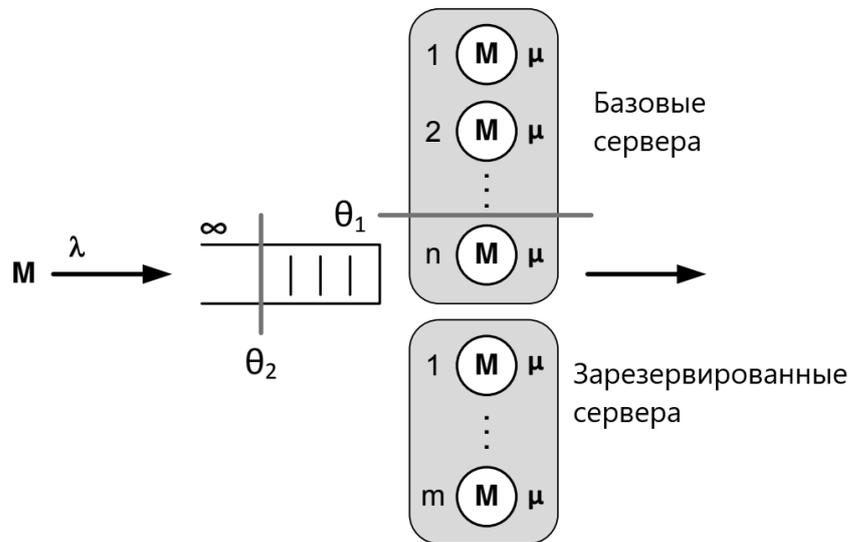


Рисунок 2 – Модель энергоэффективного дата-центра.

личество заданий в системе, а j равно 1, если зарезервированные серверы активированы, и 0, если они не активированы [4].

Обозначим $x(i, j)$ стационарную вероятность того, что i заданий находятся в системе, и зарезервированные серверы активированы ($j = 1$) или деактивированы ($j = 0$). Основываясь на диаграмме интенсивностей переходов, можно сформулировать уравнения макросостояний, определяемые как сумма всех уравнений локального баланса для состояний, относящихся к соответствующим макросостояниям.

Используя найденные в результате решения уравнений макросостояний стационарные вероятности состояний, можно вычислить другие характеристики системы. Пусть нагрузка системы

$$a = \frac{\lambda}{\mu},$$

коэффициент использования

$$p = \frac{\lambda}{\mu(n + m)}.$$

Средняя длина очереди ω вычисляется по формуле

$$\omega = \sum_{i=n}^{n+\theta_2} (i-n)x(i,0) + \sum_{i=n+m}^{+\infty} (i-(n+m))x(i,1). \quad (1)$$

Тогда среднее время ожидания заданий в системе

$$E[W] = \frac{\omega}{\lambda}. \quad (2)$$

Формула для вычисления среднего энергопотребления имеет вид

$$\begin{aligned} E = & \sum_{i=0}^n x(i,0)(ie_{busy} + (n-i)e_{idle} + me_{off}) + \\ & + \sum_{i=n+1}^{n+\theta_2} x(i,0)(ne_{busy} + me_{off}) + \\ & + \sum_{i=\theta_1}^{n+m} (x(i,1)(ie_{busy} + (n+m-i)e_{idle}) + x(i > n+m)(n+m)e_{busy}). \end{aligned} \quad (3)$$

Подраздел 2.3 посвящен вычислению стационарных вероятностей состояний системы, удовлетворяющих уравнениям макросостояний. Все вероятности содержат коэффициент $x(0,0)$, который может быть вычислен с использованием условия нормировки вероятностей.

В подразделе 2.4 описывается рекурсивный алгоритм вычисления стационарных вероятностей состояний системы. Сначала определим $x(0,0)$ как константу K_0 , а затем вычислим вероятности состояний. Для начала вычислим $x(i,0)$ для $0 < i < \theta_1$, зависящие от $x(0,0)$. Чтобы получить вероятность $x(\theta_1,0)$, требуется не только $x(\theta_1-1,0)$, но и $x(\theta_1,1)$, которая пока не вычислена. Поскольку это имеет место для всех $x(i,0)$ для $\theta_1 \leq i \leq n+\theta_2$, можно ввести константу K_1 для $x(\theta_1,1)$ и вычислить все $x(i,0)$ для $\theta_1 \leq i \leq n+\theta_2$ в виде линейной комбинации $x(\theta_1-1,0)$ и K_1 следующим образом:

$$x(i,0) = x(\theta_1-1,0)u_i + K_1v_i. \quad (4)$$

Для $i = \theta_1$ имеем $u_{\theta_1} = \frac{a}{\theta_1}$ и $v_{\theta_1} = 1$.

Для $\theta_1 < i \leq n$

$$u_i = \frac{a}{i} u_{i-1}, \quad (5)$$

$$v_i = \frac{a}{i} v_{i-1} + \frac{\theta_1}{i}. \quad (6)$$

Для $n < i \leq n + \theta_2$:

$$u_i = \frac{a}{n} u_{i-1}, \quad (7)$$

$$v_i = \frac{a}{n} v_{i-1} + \frac{\theta_1}{n}. \quad (8)$$

Таким образом, получаем вероятность $x(n + \theta_2, 0)$, зависящую от $x(\theta_1 - 1, 0)$ и $K_1 = x(\theta_1, 1)$:

$$x(n + \theta_2, 0) = x(\theta_1 - 1, 0) u_{n+\theta_2} + K_1 v_{n+\theta_2}. \quad (9)$$

Из (9) находим

$$K_1 = \frac{\frac{a}{\theta_1} u_{n+\theta_2}}{\frac{a}{\theta_1} v_{n+\theta_2} + \theta_1} x(\theta_1 - 1, 0), \quad (10)$$

что позволяет рассчитать вероятности $x(i, 0)$ для $\theta_1 \leq i \leq n + \theta_2$ используя формулу (4). Теперь можно вычислить вероятности состояний, в которых были активированы зарезервированные серверы.

Для $\theta_1 < i \leq n + \theta_2 + 1$:

$$x(i, 1) = \frac{a}{i} (x(i - 1, 1) + x(n + \theta_2, 0)). \quad (11)$$

Для $n + \theta_2 + 1 < i \leq n + m$:

$$x(i, 1) = \frac{a}{i} x(i - 1, 1). \quad (12)$$

Теперь, когда известны все вероятности относительно K_0 , применяем условие нормировки, чтобы найти K_0 и нормировать значения всех вероятностей.

Третий раздел «Метод анализа систем массового обслуживания с управлением числом приборов» содержит краткий обзор метода анализа систем массового обслуживания с управлением числом приборов. Особенностью рассматриваемых работ является наличие одного основного прибора при нескольких резервных приборах. Эта постановка задачи явля-

ется частным случаем предлагаемой в разделе 2.

В подразделе 3.1 описывается система с простейшим входящим потоком с параметром λ и экспоненциальным распределением времени обслуживания с параметром μ . Система имеет один основной прибор и $m - 1$ резервных приборов. Определяются состояния системы и строится диаграмма интенсивностей переходов между состояниями системы.

В подразделе 3.2 описывается нахождение стационарных вероятностей состояний системы. При выводе уравнений для определения вероятностей состояний используются результаты для сечений стохастических графов.

В подразделе 3.3 рассматриваются частные случаи систем массового обслуживания с управлением числом приборов [5]. Первый случай соответствует бесконечному числу подключаемых приборов обслуживания, причем приборы подключаются при одной и той же длине очереди n , когда в систему поступает $(n + 1)$ -е требование и выключаются только тогда, когда «разгрузят» очередь до нуля. Второй случай представляет собой модель, где число резервных обслуживающих приборов ограничено.

Четвертый раздел «Описание алгоритма и программы для анализа производительности и энергосбережения дата-центра» посвящен описанию алгоритма метода анализа модели дата-центра и программы для анализа производительности и энергосбережения дата-центра.

Подраздел 4.1 содержит описание алгоритма метода нахождения характеристик системы на основе рекурсивного алгоритма вычисления стационарных вероятностей состояний системы. Алгоритм состоит из четырех блоков, которые выполняются последовательно.

Блок 1. Ввод параметров дата-центра. Вводятся начальные данные, необходимые для анализа дата-центра.

Блок 2. Вычисляются стационарные вероятности состояний системы.

1. Полагаем, что $\bar{x}(0, 0) = 1$ и вычисляем $\bar{x}(i, 0)$ для $0 < i < \theta_1$.
2. Вычисляем вспомогательные величины $u_i, v_i, i = \theta_1, \dots, n + \theta_2, K_1$, используя (5)-(8), (10).
3. Для $\theta_1 \leq i \leq n + \theta_2$ находим $\bar{x}(i, 0)$, используя (4).
4. Полагаем $\bar{x}(\theta_1, 1) = K_1$ и для $\theta_1 < i \leq n + m$ вычисляем $\bar{x}(i, 1)$, используя (11), (12).

5. Вычисляем K_0 по формуле

$$x(0, 0) = \left(1 + \sum_{k=1}^{n+\theta_2} \bar{x}(k, 0) + \sum_{k=\theta_1}^{i_{max}} \bar{x}(k, 1)\right)^{-1},$$

где i_{max} — максимальное количество требований в системе, затем для $0 < i \leq n + \theta_2$

$$x(i, 0) = x(0, 0) \cdot \bar{x}(i, 0),$$

и для $\theta_1 \leq i \leq i_{max}$

$$x(i, 1) = x(0, 0) \cdot \bar{x}(i, 1).$$

Блок 3. Стационарные характеристики. Используя формулы (1)-(3), вычисляем стационарные характеристики.

Блок 4. Вывод результатов. Полученные значения характеристик системы выводятся на экран.

В подразделе 4.2 описывается программа для анализа производительности и энергосбережения дата-центра. Программная реализация алгоритма была написана на языке программирования C# с помощью среды разработки Visual Studio на основе Windows Forms, платформы пользовательского интерфейса для создания разнообразных клиентских приложений рабочего стола в Windows.

Программа состоит из двух модулей. Program.cs - модуль, отвечающий за создание приложения и отображение главного окна программы. Form1.cs - модуль, отвечающий за расчет основных характеристик дата-центра, а также нахождение вероятностей состояний системы.

Разработанная программа имеет оконный интерфейс. При запуске программы открывается окно, где пользователю предлагается ввести параметры дата-центра. Когда все значения параметров дата-центра введены в соответствующие поля, необходимо нажать на кнопку «Вычислить». На этом этапе программа проверит корректность введенных параметров и в случае возникновения ошибок ввода, откроет окно, в котором будет выведено сообщение об ошибке. Если введенные данные прошли проверку на корректность, то в окно будут выведены значения основных характеристик функционирования дата-центра.

Пятый раздел «Результаты анализа производительности и энергосбережения дата-центра» посвящен исследованию зависимостей различных характеристик функционирования дата-центра от изменения исходных параметров.

Описываются результаты вычисления основных функциональных характеристик дата-центра в зависимости от количества зарезервированных серверов. Рассматриваются примеры с различным числом зарезервированных серверов. По результатам исследования был сделан следующий вывод: при увеличении числа базовых серверов среднее число требований, ожидающих обслуживания в очереди, уменьшается; также уменьшается среднее время ожидания заданий в системе, а энергопотребление ДЦ, напротив, увеличивается.

Далее исследуется зависимость функциональных характеристик ДЦ с зарезервированными серверами от параметра θ_2 . По результатам исследования был сделан следующий основной вывод: при увеличении θ_2 среднее число требований, ожидающих обслуживания в очереди, увеличивается, а среднее энергопотребление дата-центра уменьшается.

Далее рассматривается зависимость функциональных характеристик ДЦ с зарезервированными серверами от интенсивности обслуживания сервера. По результатам исследования был сделан следующий вывод: при увеличении интенсивности обслуживания сервера, среднее число требований, ожидающих обслуживания в очереди, среднее время ожидания заданий в системе и энергопотребление дата-центра уменьшаются.

ЗАКЛЮЧЕНИЕ

В бакалаврской работе была изучена математическая модель дата-центра с зарезервированными серверами.

Был разработан алгоритм метода анализа модели ДЦ с зарезервированными серверами. Данный алгоритм при заданных параметрах дата-центра позволяет находить стационарные вероятности состояний системы и другие ее стационарные характеристики.

По разработанному алгоритму была написана программа на языке C# с помощью среды разработки Visual Studio на основе Windows Forms для анализа дата-центра. Данная программа позволяет задать параметры ДЦ и вычислить его характеристики.

С помощью программы были исследованы зависимости характеристик от различных параметров дата-центра и рассмотрены способы повышения энергоэффективности дата-центра при изменении его начальных параметров.

Основные источники информации:

- 1 Hwaiyu, P. E. Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center/ P. E. Hwaiyu — Wiley. — 2001. — 717 p.
- 2 Кальфа , К. В. Архитектура центров обработки данных / К. В. Кальфа, В. А. Докучаев. — Москва : Издательство «Горячая линия - Телеком», 2022. — 240 с.
- 3 Боссерт, Ш. ЦОД как на ладони [Электронный ресурс] // Журнал сетевых решений/LAN. — 2012. — № 2. — URL:<http://www.osp.ru/lan/2012/02/13012848/>(дата обращения: 15.05.2024).
- 4 Schwartz, C. A queuing analysis of an energy-saving mechanism in data centers/ C. Schwartz, R. Pries, P. Tran-Gia // The International Conference on Information Network. — 2012. — P. 70-75.
- 5 Romani, A. A queueing model with a variable number of channels / A. Romani // Trabajos de estadística. — 1957. — Vol. 3. — P. 175-189.