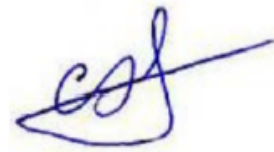


На правах рукописи



САРГСЯН Эрик Ромович

**РАЗРАБОТКА СРЕДСТВ СПЕЦИАЛЬНОГО МАТЕМАТИЧЕСКОГО И
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ПРОЦЕССОВ УПРАВЛЕНИЯ
МОНИТОРИНГОМ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ
В РАМКАХ ГИБРИДНЫХ ПЕРЕДАЮЩИХ СРЕД**

Специальность: 2.3.5. Математическое и программное обеспечение
вычислительных систем, комплексов и
компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Воронеж – 2024

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Воронежский государственный технический университет».

Научный руководитель: **Рындин Александр Алексеевич**, доктор технических наук, профессор

Официальные оппоненты: **Орлов Сергей Павлович**, доктор технических наук, профессор, ФГБОУ ВО «Самарский государственный технический университет», профессор кафедры вычислительная техника

Перепёлкин Дмитрий Александрович, доктор технических наук, профессор, ФГБОУ ВО «Рязанский государственный радиотехнический университет им. В.Ф. Уткина», профессор кафедры систем автоматизированного проектирования вычислительных средств

Ведущая организация: **ФГАОУ ВО «Сибирский федеральный университет», (г. Красноярск)**

Защита состоится «06» декабря 2024 года в 12:00 часов в конференц-зале на заседании диссертационного совета 24.2.286.04, созданного на базе ФГБОУ ВО «Воронежский государственный технический университет», по адресу: г. Воронеж, Московский просп., д. 14, ауд. 216.

С диссертацией можно ознакомиться в научно-технической библиотеке ФГБОУ ВО «Воронежский государственный технический университет» и на сайте <https://cchgeu.ru/>

Автореферат разослан «15» октября 2024 г.

Ученый секретарь
диссертационного совета
к.т.н., доцент



Гусев Константин Юрьевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время высокая степень автоматизации и информатизации привела к широкому использованию распределенных информационных систем в различных отраслях. Такие системы обрабатывают огромные объемы данных и обеспечивают передачу информации в системах различных типов и технологий. В рамках данной научной работы в качестве примера информационной системы, использующей гибридную передающую среду, представлена система управления мобильным доступом. Подобные сети часто используются в телекоммуникационной сфере для обеспечения связи и на данный момент занимают приоритетную роль в современном мире.

В связи с активным развитием области информационных систем для их контроля необходимо реализовывать мониторинговые системы. Они включают множество функционала по отслеживанию состояния элементов системы. В частности, возникает проблема своевременного оповещения пользователей о возможных аномалиях в тех или иных элементах системы, что позволяет оперативно решать возникшие проблемы в системе. Также актуален вопрос контроля передачи данных в распределенных информационных системах. Для таких систем необходимо обеспечить возможность мониторинга передачи больших объемов информации с высокой скоростью.

Значительные научные достижения в области развития математических моделей анализа больших данных и обнаружения аномалий были представлены в работах следующих отечественных ученых: Орлова С.П., Перепёлкина Д.А., Воробьева А.Е., Лаврова А.А., Щемелинина Д.А., Бычковой Е.Д., Иванова А.Б., Короткова Е.С., Суханова А.В., Чудиновой К.В., Яковлева С.В., и др. Мониторинг и прогнозирование в информационных системах раскрыты в работах Жихарева А.П., Лавлинского Н.Е., Максимова Н.В., Мещерякова С.В., Шварца Б., и др.

Таким образом, актуальность темы диссертационного исследования продиктована необходимостью дальнейшего развития средств математического и программного обеспечения мониторинга распределенных информационных систем в гибридных передающих средах. Реализация на основе поиска компонент сильной связности и методов оценки качества обслуживания информационных систем будет способствовать повышению надежности, стабильности, эффективности и безопасности функционирования распределенных информационных систем. Это позволит применять разработанные методики непосредственно в критически важных отраслях, таких как производство, транспорт, телекоммуникации, здравоохранение и других, где важную роль играет обработка и передача больших объемов данных.

Тематика диссертационной работы соответствует научному направлению ФГБОУ ВО «Воронежский государственный технический университет» «Вычислительные комплексы и проблемно-ориентированные системы управления».

Целью работы является разработка средств математического и программного обеспечения процессов управления мониторингом распределенных информационных систем в гибридных передающих средах на основе поиска компонент сильной связности и методов оценки качества обслуживания информационных систем.

Задачи исследования. Для достижения поставленной цели необходимо решить следующие задачи:

1. Разработать алгоритм мониторинга распределенной информационной системы, основанный на модификации метода поиска компонент сильной связности с использованием обхода структуры системы в глубину с временными метками.

2. Разработать алгоритм анализа качества обслуживания, основанный на учете контекстных факторов гибридных сред и групповых параметров различных сегментов системы.

3. Разработать алгоритм идентификации и коррекции аномального поведения системы, основанный на имитационном моделировании с использованием LRU-алгоритма кэширования данных.

4. Разработать структуру базы данных для хранения результатов процесса мониторинга, которая основана на колоночной модели хранения с применением индексирования данных и ограничением по времени хранения данных в базе.

5. Разработать программный прототип системы мониторинга, основанный на возможности масштабирования и прозрачной интеграции с существующими информационными системами.

Объект исследования: распределенная информационная система в гибридной передающей среде.

Предмет исследования: средства математического и программного обеспечения процесса управления мониторингом распределенных информационных систем в гибридных передающих средах на основе поиска компонент сильной связности и методов оценки качества обслуживания.

Методы исследования. При решении поставленных в диссертации задач использовались теория графов, методы кластеризации данных, теория принятия решений, имитационное моделирование, методы кэширования данных, а также методы объектно-ориентированного программирования.

Тематика работы соответствует следующим пунктам паспорта научной специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей: п. 3 «Модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем»; п. 9 «Модели, методы, алгоритмы, облачные технологии и программная инфраструктура организации глобально распределенной обработки данных».

Научная новизна

В диссертации получены следующие результаты, характеризующиеся научной новизной:

1. Алгоритм мониторинга распределенной информационной системы, отличающийся модифицированным поиском компонент сильной связности на основе обхода структуры системы в глубину с временными метками и применением многокритериального метода принятия решений, обеспечивающий более эффективный анализ состояния элементов системы и вычисления нагрузки на систему с учетом состояния ее узлов.

2. Алгоритм анализа качества обслуживания, отличающийся учетом контекстных факторов гибридных сред и групповых параметров различных сегментов системы и обеспечивающий повышение качества обслуживания с учетом переменного системного контекста.

3. Алгоритм идентификации и коррекции аномального поведения системы, отличающийся использованием имитационного моделирования и LRU-алгоритма кэширования данных при поиске уязвимых мест и обеспечивающий возможность гибкой настройки обнаружения аномального поведения и формирования мер для его коррекции.

4. Структура базы данных для хранения результатов процесса мониторинга, отличающаяся использованием колоночной модели хранения и индексирования данных, а также ограниченным временем жизни данных в базе и обеспечивающая повышенную производительность системы за счет модели агрегирования данных по типам на таблицы.

5. Структура программного обеспечения системы мониторинга, отличающаяся прозрачной интеграцией с существующими информационными системами на гибридных передающих платформах и обеспечивающая возможность масштабирования.

Теоретическая значимость заключается в развитии ряда методов мониторинга распределенных информационных систем в гибридных передающих средах, что позволяет использовать их в проектных и научно-исследовательских организациях, занимающихся проектированием и разработкой распределенных систем, требующих мониторинга.

Практическая значимость. Реализация математического и программного обеспечения системы мониторинга распределенных информационных систем в виде специализированного программного средства с микросервисной структурой позволяет интегрировать его с существующими информационными системами для подключения функций мониторинга. На специализированное программное средство получено свидетельство о государственной регистрации.

Положения, выносимые на защиту

1. Алгоритм мониторинга распределенной информационной системы с методом поиска компонент сильной связности и применением многокритериального метода принятия решений обеспечивает более эффективный анализ состояния элементов системы и вычисления нагрузки на систему с учетом состояния ее узлов.

2. Алгоритм анализа качества обслуживания, учитывающий контекстные факторы гибридных сред и групповые параметры различных сегментов системы, повышает качество обслуживания с учетом переменного системного контекста.

3. Алгоритм идентификации и коррекции аномального поведения системы предоставляет гибкую настройку обнаружения аномального поведения и формирования мер по его коррекции.

4. Структура базы данных для хранения результатов процесса мониторинга улучшает производительность системы за счет агрегирования данных по типам на таблицы.

Результаты внедрения. Результаты исследований были успешно внедрены в учебный процесс ФГБОУ ВО «ВГТУ»: направление 09.04.02 «Информационные системы и технологии», в лабораторный практикум и лекционный курс по дисциплине «Инфраструктурные компоненты информационных систем».

Разработанные методы, математические модели и алгоритмы управления мониторингом распределенных информационных систем применяются для осуществления мониторинга работы распределенной информационной системы обеспечения услуг связи на предприятии ООО «Интерком». Интеграция направлена на улучшение состояния рассматриваемой информационной системы путем анализа мониторинговых данных для ее элементов.

Разработанный программный инструмент мониторинга распределенных информационных систем применяется в деятельности Центра прикладных исследований «Проектирование и разработка информационных систем» для обеспечения мониторинговыми функциями разрабатываемых информационных систем и продуктов.

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на следующих конференциях: Международной молодежной научной школе «Оптимизация и моделирование в автоматизированных системах», Воронеж, 2021-2023 г; Международной научно-практической конференции «Интеллектуальные информационные системы», Воронеж, 2022-2024; VIII Международной научно-практической конференции «Антропоцентрические науки: инновационный взгляд на образование и развитие личности», Воронеж, 13-14 ноября 2018 г. и XIV Международной конференции «Антропоцентрические науки: инновационный взгляд на образование и развитие личности», 19-20 апреля 2021 г.

Достоверность результатов обусловлена корректным использованием теоретических методов исследования и подтверждена результатами сравнительного анализа данных вычислительных и натуральных экспериментов.

Публикации. По результатам диссертационного исследования опубликовано 12 научных работ (5 – без соавторов), в том числе 4 – в изданиях, рекомендованных ВАК РФ и одно свидетельство о регистрации программы для ЭВМ.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы и приложений. Объем работы – 142 страницы, включая 42 рисунка, 9 таблиц. Список литературы содержит 108 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность исследования, сформулированы его цель и задачи, научная новизна и практическая значимость полученных результатов, приведены сведения об апробации и внедрении работы.

В первой главе исследуется концепция процесса управления мониторингом и его роль в обеспечении стабильной и безопасной работы распределенных информационных систем. Описывается, что это процесс систематического наблюдения, сбора, анализа и оценки данных о состоянии и работе системных элементов и систем с целью обеспечения их стабильного функционирования, а также выявления и предотвращения возможных проблем.

Дано определение, что гибридная передающая среда представляет собой комбинацию различных типов сред передачи информации, локальных и глобальных сетей, беспроводных и проводных каналов, обеспечивающих универсальность и эффективность передачи данных. Гибридная среда позволяет использовать технологии с различными характеристиками для обеспечения более надежной и гибкой работы распределенной информационной системы.

Специфика мониторинга распределенных информационных систем в гибридной передающей среде заключается в необходимости реализовать процесс, который будет отслеживать состояние рассматриваемой системы, учитывая, что гибридная среда может состоять из подсистем различных видов, использующих большой набор разнообразных технологий для передачи данных. Таким образом рассматриваемые распределенные информационные системы являются гетерогенными. Для упрощения работы с подобными системами существуют способы приведения гетерогенной информационной системы к однородной, которые заключаются в интеграции разнородных информационных систем с использованием единой модели, стандартизированных данных и многомерного подхода.

Изучены возможности и недостатки существующих способов мониторинга современных распределенных информационных систем для адаптации их в систему мониторинга реализуемую в данной научной работе. Рассмотрены особенности распределенных информационных систем в гибридных передающих средах, для понимания предметной области, а также подходов к решению задач мониторинга и коррекции аномального поведения таких систем. Результат анализа потребовал формализации данных задач, а также алгоритмизации их решения, что представлено на рисунке 1. Сформулированы цель и задачи исследования.



Рисунок 1 - Дизайн исследования

Вторая глава посвящена разработке механизмов мониторинга элементов распределенной информационной сети.

Всего предложено три механизма мониторинга распределенной информационной системы, которые будут рассмотрены далее.

1. Особенности разработки математического и программного обеспечения процесса управления мониторингом распределенных информационных систем в гибридных передающих средах на основе поиска компонент сильной связности при использовании обхода структуры в глубину с временными метками.

У распределенных информационных систем, с большим количеством элементов в структуре существует множество компонент сильной связности (КСС), которые представляются в виде циклических зависимостей между элементами системы. Структуру распределенной информационной системы можно спроецировать на модель представления графов. Наличие подобных зависимостей увеличивает время, затраченное на алгоритм обхода структуры и требует разработки более гибкого метода для решения данной проблемы. В связи с этим, рассматриваемую задачу предлагается декомпозировать на две подзадачи:

- поиск всех циклов в графе и группировка каждого из них в один узел с указанием максимальной длины пути внутри цикла;
- подсчет длины пути от исходного рассматриваемого элемента до каждого узла в системе.

Для решения подзадачи поиска максимального пути выбран алгоритм Косарайю. Он адаптирован под текущую задачу, для того чтобы рассчитывать максимальный путь от исходного элемента до каждого узла системы.

Для поиска каждой КСС выполняется множество последовательных обходов в глубину графа G . Для каждой вершины графа вызывается функция обхода в глубину и данный процесс продолжается итерационно, пока алгоритм не пройдет все вершины графа. Во время работы алгоритма для каждой вершины v запоминается время входа $Tin[v]$ и выхода $Tout[v]$.

Также при обходе каждой вершины она помечается цветом. При первом посещении вершины - серым цветом и после обработки всех ее смежных вершин она помечается черным цветом. Из этого следует, что:

- $Tin[v]$ – время, за которое вершина обнаружена и закрашена в серый цвет;
- $Tout[v]$ – время, за которое вершина обработана и закрашена в черный цвет.

Каждая вершина v соответствует неравенству: $Tin[v] < Tout[v]$.

В результате полного обхода формируется список всех найденных компонент сильной связности.

Далее применяется многокритериальный метод принятия решений разработанный на основе подхода TOPSIS, отличающийся тем, что предпочтительное решение должно иметь не только наибольшую близость к идеальному значению, но также быть самым дальним от неприемлемого решения.

Формируется матрица решений R .

$$R = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \quad (1)$$

Для нормализации матрицы решений используется метод нормализации вектора.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}, i = 1, 2, \dots, m; j = 1, 2, \dots, n. \quad (2)$$

Вычисляется нормализованная матрица решений. Вес каждого критерия определяется экспертом. Взвешенная нормализованная матрица получается, как произведение нормализованной матрицы на взвешенную матрицу критерия.

$$V = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \dots & w_n r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \dots & w_n r_{2n} \\ \dots & \dots & \dots & \dots \\ w_1 r_{m1} & w_2 r_{m1} & \dots & w_n r_{mn} \end{bmatrix} \quad (3)$$

где $W = (w_1, w_2, \dots, w_n)$ представляет собой набор весов, удовлетворяющих выражению $\sum_{i=1}^n w_i = 1$.

Если критерий оказывает положительное влияние, то положительно идеальное решение имеет наибольшее значение по каждому критерию, а отрицательное идеальное решение имеет наименьшее значение по каждому критерию.

$$A^* = \{(j \in J_+), (j \in J_-), i = 1, 2, \dots, m\} = \{v_1^*, v_2^*, \dots, v_n^*\} \quad (4)$$

$$A^- = \{(j \in J_+), (j \in J_-), i = 1, 2, \dots, m\} = \{v_{1-}, v_{2-}, \dots, v_{n-}\} \quad (5)$$

где $J_+ = \{j = 1, 2, \dots, n\}$ – критерии, оказывающие положительное влияние; $J_- = \{j = 1, 2, \dots, n\}$ – критерии, оказывающие негативное влияние.

Вычисляется мера разделения (евклидово расстояние) для каждой альтернативы. Расстояние между целевой альтернативой i и отрицательным идеальным решением равно:

$$S_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_{j-})^2}, i = 1, 2, \dots, n \quad (6)$$

Расстояние между альтернативой i и положительным идеальным решением:

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j)^2}, i = 1, 2, \dots, n \quad (7)$$

Вычисляется относительная близость к положительному идеальному решению:

$$C_i = \frac{S_i^-}{S_i^- + S_i^+}, 1 \geq C_i \geq 0; i = 1, 2, \dots, m \quad (8)$$

Ранжируется порядок предпочтений:

$$C_i = 1 \rightarrow A_i = A^*$$

$$C_i = 0 \rightarrow A_i = A^-$$

В результате получаем способ принятия решений для вычисления оптимальных параметров для элементов в системе.

Таким образом, разработана математическая модель механизма мониторинга элементов распределенной информационной системы, отличающаяся модифицированным поиском компонент сильной связности на основе обхода структуры системы в глубину с временными метками, обеспечивающая оптимизацию количества требуемых на обработку запросов ресурсов и более эффективный анализ состояния элементов информационной системы.

2. Для расчета качества обслуживания системы предлагается учитывать контекстные факторы гибридных сред и групповых параметров различных сегментов системы. Расчет качества обслуживания позволяет обеспечивать доступность элементов информационной системы, безопасную передачу данных и эффективное использование существующей пропускной способности системы, вместо постоянной модификации инфраструктуры.

Для вычисления параметров качества обслуживания системы применяются SLA (Service Level Agreement) соглашения. SLA соглашение отображает цель, которую необходимо достигнуть, по какому-либо заданному направлению работы системы. Дочерней сущностью SLA является показатель SLO (Service Level Objective) представляющий цель уровня обслуживания и определяющий желаемые показатели качества конкретного аспекта системы. Например, SLO может задавать время безотказной работы системы и по этому показателю система обязана иметь значение не меньше, чем было установлено. В основе структуры соглашений находится SLI (Service Level Indicator) – индикатор уровня обслуживания, который измеряет соответствует ли уровень обслуживания SLO заданным требованиям.

В общем виде расчет показателя SLA можно произвести по формуле 9.

$$SLA = \sum_{i=1}^N SLO_i \quad (9)$$

Значение SLA вычисляется как сумма вычисленных значений SLO, которые включены в данное SLA. Значение конкретного SLO можно вычислить по формуле 10.

$$SLO = aggFunc(SLI_0, SLI_1, \dots, SLI_N) \quad (10)$$

SLI вычисляется на основе конкретных параметров элементов системы и может быть задан набором формул. Формула 11 описывает расчет показателя доступности элемента системы с учетом контекстных параметров.

$$SLI_{\text{доступность}} = \frac{T_w - (\sum_{i=1}^N P_i)}{T_w} * 100, \quad (11)$$

где T_w – заданное необходимое время бесперебойной работы, с; P_i – контекстный параметр, влияющий в негативную сторону на показатель доступности, с.

Формула 12 описывает расчет показателя джиттера конкретного элемента системы.

$$SLI_{\text{джиттер}} = \frac{\sum_1^N T_j}{N_t}, \quad (12)$$

где T_j – время между двумя последовательными пакетами данных, мс; $\sum_o^N T_j$ – сумма значений задержки между пакетами из последовательности N пакетов, идущих друг за другом, мс; N_t – количество последовательных значений задержек между пакетами, учитываемых в расчетах.

Формула 13 описывает расчет показателя джиттера для сегмента, состоящего из множества элементов системы.

$$SLI_{\text{сег}_{\text{джиттер}}} = \sum_1^N SLI_{\text{джиттер}}, \quad (13)$$

где N – количество элементов сегмента системы.

Формула 14 позволяет рассчитать успешность времени отклика элемента системы.

$$SLI_{\text{отклик}} = \frac{T_x}{\text{avg}(T_e) - \Delta t}, \quad (14)$$

где $\text{avg}(T_e)$ – среднее время отклика элемента системы, мс; Δt – допустимая погрешность времени отклика, задается вручную экспертом, мс; T_x – целевое время отклика элемента системы, мс. Если значение $SLI_{\text{отклик}} \geq 1$, то среднее время отклика элемента можно считать успешным, с учетом заданной погрешности.

Проанализировать качество обслуживания распределенной системы, основываясь на контекстных факторах гибридных сред, позволят аналитические и эмпирические модели потери сигнала при передаче его от одного элемента системы к другому. Для этого были выбраны следующие модели:

- Окамура-Хата;
- COST231-Хата;
- Кся-Бертони.

Модель Окамура-Хата применяется в диапазоне частот сигнала от 150 до 1500 МГц, а модель COST231 – Хата в диапазоне от 1500 до 2000 МГц.

Модель для расчета в городских условиях будет выглядеть следующим образом:

$$L_{\text{ОХ(город)}} = 69,55 + 26,16 * \log f - 13,83 * \log h_{\text{ИС}} + (44,9 - 6,55 * \log h_{\text{ИС}}) * \log r - a(h_{\text{ПС}}) \quad (15)$$

где f – частота сигнала (МГц); $h_{\text{ИС}}$ – высота элемента источника сигнала, м; r – расстояние зоны передачи, м; $a(h_{\text{ПС}})$ – поправочный коэффициент высоты приемника сигнала.

$$a(h_{\text{ПС}}) = (1,1 * \log f - 0,7) * h_{\text{ПС}} - (1,56 * \log -0,8) \quad (16)$$

где $h_{\text{МА}}$ – высота элемента приемника сигнала, м.

Для расчета потери сигнала в сельской или открытой местности необходимо использовать следующую формулу:

$$L_{\text{ОХ(село)}} = L_{\text{ОХ(город)}} - 4,78 * (\log f)^2 + 18,33 * \log f - 40,94 \quad (17)$$

Далее рассматривается модель COST231-Хата в рамках городских условий. Формула выглядит следующим образом:

$$L_{\text{ОХ(город)}} = 46,3 + 33,9 * \log f - 13,82 * \log h_{\text{ИС}} - a(h_{\text{ПС}}) + (44,9 - 6,55 * \log h_{\text{ИС}}) * \log r + C \quad (18)$$

где f – частота сигнала (МГц); $h_{\text{ИС}}$ – высота элемента источника сигнала, м; r – расстояние зоны передачи, м; $a(h_{\text{ПС}})$ – поправочный коэффициент высоты элемента приемника сигнала, рассчитывается по формуле 10; C – коэффициент, со значениями $C = 0$ для открытой местности и небольших городов, $C = 3$ для крупных городов.

Следовательно, также рассмотрим модель COST231-Хата для сельской и открытой местности:

$$L_{OX(\text{село})} = L_{OX(\text{город})} - 4,78 * (\log f)^2 + 18,33 * \log f - 40,94 \quad (19)$$

В сравнении с двумя приведенными выше моделями, также будет рассмотрена модель Кся-Бертони, которая позволяет учесть множество дополнительных параметров внешней среды.

Модель основана на уравнении волновой оптики и отлично подходит для ситуаций с множеством факторов внешней среды.

$$L = -10 * \log \left(\frac{\lambda}{4\pi * R} \right)^2 - 10 * \log \left(\frac{\lambda}{2\pi^2 * r} * \left[\frac{1}{\theta} - \frac{1}{2\pi + \theta} \right]^2 \right) - \\ - 10 * \log \left(2,35^2 * \left[\frac{\Delta h_b}{R} * \sqrt{\frac{d}{\lambda}} \right]^{1,8} \right) \quad (20)$$

где λ – длина волны, м. ($\lambda = \frac{c}{f}$, где $c = 3 * 10^8$ м/с – скорость света); R – расстояние между элементом источником сигнала и элементом приемником сигнала, м; $\theta = \arctan \frac{\Delta h_M}{x} * \frac{180}{\pi}$, рад; $\Delta h_M = h_0 - h_{ПС}$ – разность высот среднего уровня препятствий и элемента приемника сигнала соответственно, м; $x = \frac{\omega}{2}$ – расстояние по горизонтали между элементом приемника сигнала и краем препятствия, на которое попадает волна, м; ω – средняя ширина улиц, м; $\Delta h_b = h_{БС} - h_0$ – разность высот элемента источника сигнала и среднего уровня препятствий соответственно, м; d – средний интервал между кварталами, м.

Расстояние от точки преломления луча до элемента приемника сигнала представлено формулой:

$$r = \sqrt{\Delta h_M^2 + x^2} \quad (21)$$

$\Delta h_M = h_0 - h_{ПС}$ – разность высот среднего уровня препятствий и элемента приемника сигнала, м; $x = \frac{\omega}{2}$ – расстояние по горизонтали между элементом приемника сигнала и краем препятствия, на которое попадает волна, м.

Угол падения преломленного луча на элемент приемника сигнала представлен формулой:

$$\theta = \arctan \left(\frac{\Delta h_M}{x} \right) * \frac{180}{\pi} \quad (22)$$

где: $\Delta h_M = h_0 - h_{ПС}$ – разность высот среднего уровня препятствий и высоты приемника сигнала, м; $x = \frac{\omega}{2}$ – расстояние по горизонтали между элементом приемника сигнала и краем препятствия, на которое попадает волна, м.

Стоит отметить, что достоверные результаты получаются только при вычислении на расстоянии не более 1,5-2 км.

Таким образом, разработана математическая модель метода анализа качества обслуживания, отличающаяся учетом контекстных факторов гибридных сред и групповых параметров различных сегментов системы и обеспечивающая повышение качества обслуживания с учетом переменного системного контекста.

3. Для анализа состояния элементов распределенной информационной системы разрабатывается метод идентификации и коррекции аномального поведения. Данный метод основывается на имитационном моделировании с использованием LRU кэша, что позволяет повысить производительность итогового алгоритма и проверять состояние элементов системы при различных заданных входных параметрах, такие как внутренние параметры элементов сети и внешние параметры гибридной среды.

Идентификация аномального поведения элемента системы проводится несколькими способами:

- анализ возможности передачи информации другим элементам системы путем расчета потери данных в процессе передачи;
- сравнение основных показателей элемента системы с эталонными значениями, учитывая погрешности;
- анализ трафика, с помощью параметров пропускной способности, скорости передачи и задержки данных, позволяет выявить аномальное поведение в различных элементах системы. Например, обнаружение большого объема трафика с необычных источников.

Так как передача данных является одним из наиболее значимых и нагруженных частей работы системы, то рассмотрим подробнее способы обеспечения безотказной работы этого аспекта.

В данной научной работе для экспериментальных расчетов используется распределенная информационная система управления мобильным доступом, в которой одним из способов передачи данных является беспроводная среда. Следовательно, данные будут передаваться сигналом, на который влияет окружающая среда с множеством различных факторов. Одним из таковых является физическое явление дифракция.

Дифракция сама по себе рассматривается как изменение направления сигнала с учетом его потери и измеряется в децибелах (дБ). Дифракционные потери рассчитываются по модели, представленной в формулах [23-28]:

$$v = h \sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}} \quad (23)$$

$$v \leq -1 \rightarrow DL = 0 \quad (24)$$

$$-1 < v \leq 0 \rightarrow L = 20 * \log_{10}(0.5 - 0.62 * v) \quad (25)$$

$$0 < v \leq 1 \rightarrow DL = 20 * \log_{10}(0.5 * e^{-0.95*v}) \quad (26)$$

$$1 < v \leq 2.4 \rightarrow DL = 20 * \log_{10} \left(0.4 - \sqrt{0.1184 - (-0.1 * v + -0.38)^2} \right) \quad (27)$$

$$2.4 < v \rightarrow DL = 20 * \log_{10} \left(\frac{0.255}{v} \right) \quad (28)$$

Рассмотрим детальнее, что означают представленные параметры:

- v – дифракционный параметр;
- d_1 – дистанция между элементом источника сигнала и препятствием, м;
- d_2 – дистанция между препятствием и элементом приемника сигнала, м;
- λ – длина волны передаваемого сигнала, м;
- h – высота препятствия, м.

Методы расчета дифракции позволяют идентифицировать anomальное поведение системы, путем экспериментальных расчетов с использованием имитационного моделирования, и оптимизировать распределение ресурсов распределенной информационной системы.

Коррекция anomального поведения элемента системы проводится несколькими способами:

- перераспределение нагрузки на другие узлы системы;
- изоляция - индивидуализация узлов, которые проявляют anomальное поведение, путем изоляции их от продуктивных узлов или сети;
- восстановление - восстановление элементов системы, которые проявили anomальное поведение, в исходное состояние или состояние, близкое к исходному;
- мониторинг и анализ – непрерывный мониторинг и анализ поведения системы для выявления возможных anomалий.

Таким образом, разработана математическая модель метода идентификации и коррекции anomального поведения информационной системы, обеспечивающая возможность гибкой настройки обнаружения anomального поведения и формирования мер для его коррекции.

Третья глава посвящена алгоритмизации методов мониторинга работы распределенной информационной системы, а также методов анализа качества обслуживания, на основе контекстных факторов гибридных средств и способов идентификации и коррекции anomального поведения системы.

На рисунке 2 представлен алгоритм мониторинга работы распределённой информационной системы.

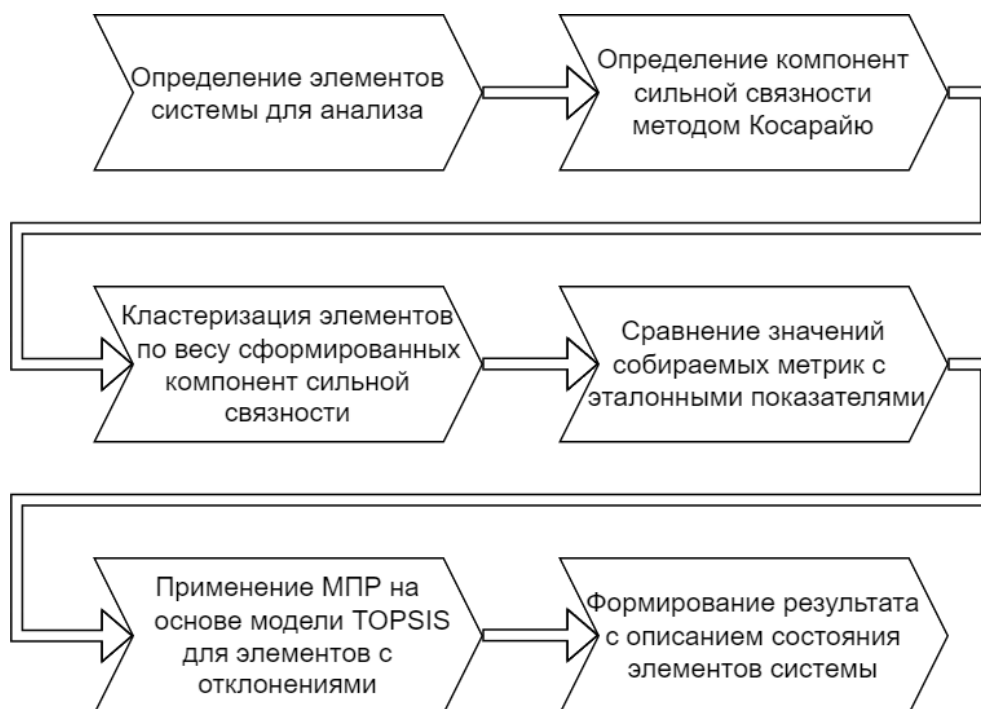


Рисунок 2 – Алгоритм мониторинга распределенной информационной системы, основанный на обходе структуры системы с постановкой временных меток

При решении задачи мониторинга состояния элементов распределенной информационной системы создан алгоритм обхода структуры системы, основанный

на поиске компонент сильной связности с временными метками. В данном алгоритме реализована кластеризация элементов системы на компоненты сильной связности различного размера для последующего применения многокритериального метода принятия решения, основанного на модели TOPSIS.

Таким образом, разработан алгоритм механизма мониторинга элементов распределенной информационной системы, отличающийся модифицированным поиском компонент сильной связности на основе обхода структуры системы в глубину с временными метками, обеспечивающий оптимизацию количества требуемых на обработку запросов ресурсов, и более эффективное перераспределение нагрузки с учетом состояний узлов системы.

Для анализа качества обслуживания элементов системы разработан алгоритм, представленный на рисунке 3.



Рисунок 3 - Алгоритм анализа качества обслуживания элементов системы

В данном алгоритме используется метод выделения компонент сильной связности, на основе результатов которого будет проводиться тестирование элементов системы. Выделяется ряд показателей характеризующих состояние элементов системы и их возможность передавать данные для последующих вычислений по моделям Окамура-Хата, COST 231-Хата и Кся-Бертони. Формируются

результаты вычисления SLA соглашений и в результате работы алгоритма выводятся пользователю для возможности дальнейшего анализа.

Таким образом, разработан алгоритм анализа качества обслуживания, отличающийся учетом контекстных факторов гибридных сред и групповых параметров различных сегментов системы и обеспечивающий повышение качества обслуживания с учетом переменного системного контекста.

Для идентификации и коррекции аномального поведения элементов системы разработан алгоритм, представленный на рисунке 4. В рамках данного алгоритма происходит обход указанных элементов системы с использованием LRU кэширования, для оптимизации процесса доступа к данным. Параметры каждого элемента сравниваются с эталонными значениями и при несоответствии применяются меры коррекции, которые определены для данного элемента и параметра. Также реализован процесс проверки интеграции между смежными элементами системы с помощью алгоритма имитационного моделирования процесса передачи данных. Данный подход позволяет оценить, есть ли необходимость изменения параметров передачи данных или переноса элементов системы для обеспечения безопасности передачи данных в системе.

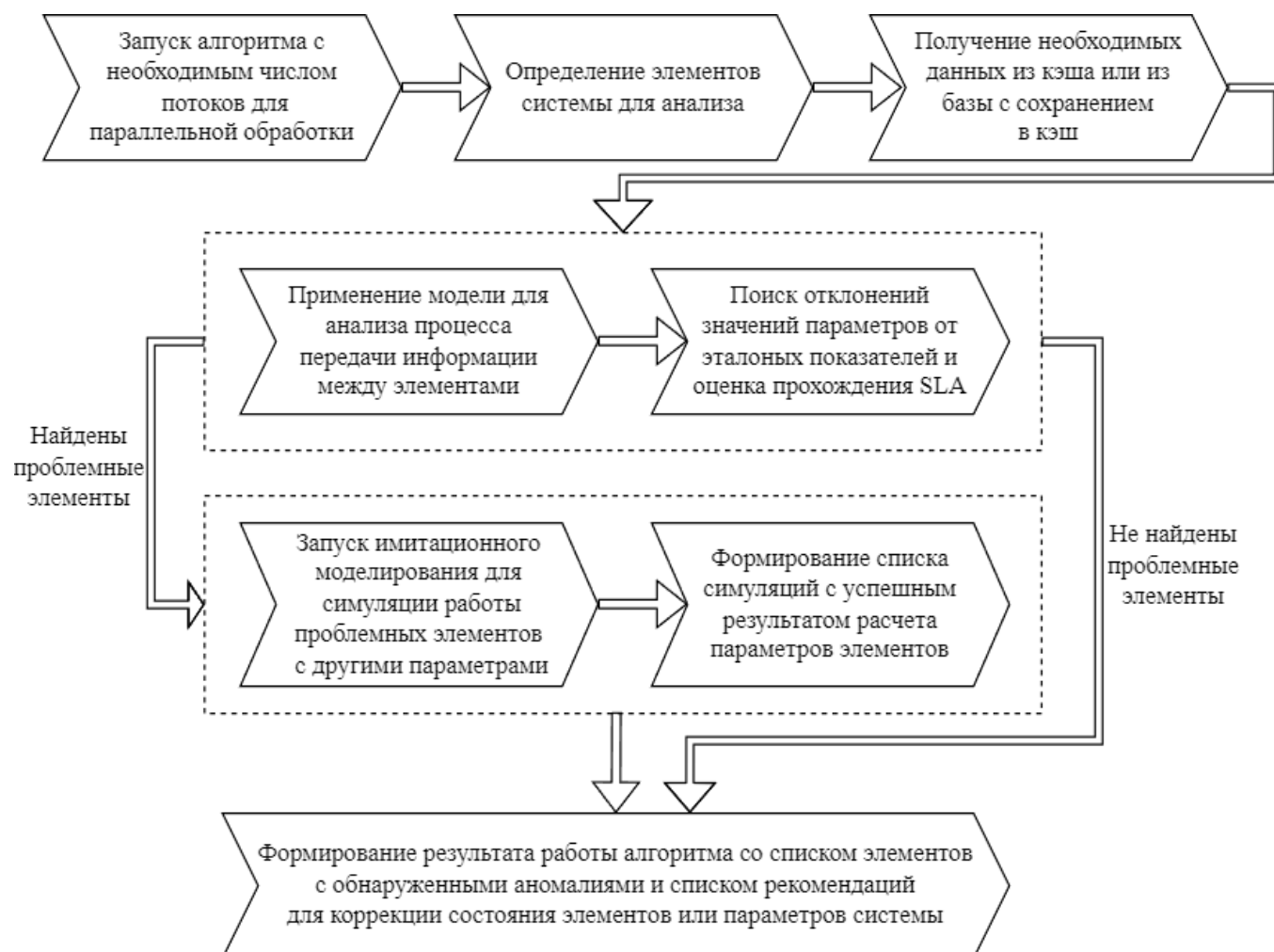


Рисунок 4 – Алгоритм идентификации и коррекции аномального поведения системы

Таким образом, разработан алгоритм идентификации и коррекции аномального поведения системы, отличающийся использованием имитационного моделирования и LRU-алгоритма кэширования данных.

Разработана модель хранения собираемых мониторинговых метрик в аналитической базе данных ClickHouse, которая отличается применением колоночной структуры для хранения данных. Так как формат и предназначение данных предполагают большое количество операций записи и не требуют частого обновления данных, то данная база данных подходит для решения поставленной задачи хранения мониторинговых данных.

На рисунке 5 представлена часть разработанной структуры базы данных, которая состоит из множества таблиц со смежной структурой. Это реализовано для более оптимизированного хранения данных в различных агрегационных формах. Помимо представленных таблиц, также существуют таблицы для хранения данных агрегированных за 3 часа, 1 день, 1 неделю и 1 месяц. Это позволяет хранить данные за долгое время, но при этом не затрачивать большие объемы пространства на внутреннем хранилище и оптимизировать запросы на получение данных. Следовательно, приводит к экономии финансовых ресурсов, затрачиваемых на расширение хранилища данных и временных ресурсов ожидания результата.

monitoring_data	
PK	<u>timestamp</u> [DateTime]
PK	<u>metric_name</u> [String]
PK	<u>object_id</u> [String]
	value [String]

monitoring_data_1minute	
PK	<u>timestamp</u> [DateTime]
PK	<u>metric_name</u> [String]
PK	<u>object_id</u> [String]
	value [String]

monitoring_data_15minute	
PK	<u>timestamp</u> [DateTime]
PK	<u>metric_name</u> [String]
PK	<u>object_id</u> [String]
	value [String]

monitoring_data_1hour	
PK	<u>timestamp</u> [DateTime]
PK	<u>metric_name</u> [String]
PK	<u>object_id</u> [String]
	value [String]

Рисунок 5 – Структура базы данных хранения метрик мониторинговой системы

Также, для экономии используемого физического пространства, база данных поддерживает механизм удаления данных по истечении настраиваемого временного периода. Данный механизм называется Time to live (TTL) и позволяет для каждой таблицы выставить свое значение TTL для своевременной очистки устаревших исторических данных.

При этом, по разработанной структуре базы данных, можно сделать вывод, что отсутствуют связи между таблицами. Это позволяет избавиться от запросов, которые требуют агрегации данных между таблицами и зачастую имеют продолжительное время выполнения запросов. Таблицы в данной базе ориентированы на то, чтобы быть

самодостаточными и хранить в себе всю необходимую информацию для конкретного запроса.

Индексирование данных в таблицах происходит по первичному ключу. Например, для таблиц хранения метрик, первичный ключом является уникальный идентификатор метрики, объекта системы и время, за которое данная метрика хранит значение. Это позволяет реализовать быстрый поиск данных по таблицам.

Таким образом, разработана структура базы данных с использованием колоночной модели представления данных для хранения собираемых метрик мониторинговой системой с применением агрегационного подхода к сохранению данных.

Четвертая глава посвящена разработке программного обеспечения на основе ранее разработанных алгоритмов и моделей. Программное обеспечение включает возможность интеграции с существующими информационными системами так как реализована поддержка REST архитектуры и HTTP протокола передачи данных, а также обмен сообщениями с помощью брокера Apache Kafka.

Архитектура состоит из следующих компонентов:

1. Микросервис MonAuth. Необходим для обеспечения взаимодействия пользователя с мониторинговой системой. Любой запрос в систему должен пройти авторизацию через данный сервис.

2. Микросервис Monitor UI. Обеспечивает удобный для пользователя графический интерфейс для управления функционалом системы мониторинга и визуализации собираемых данных.

3. Микросервис API Gateway. Является узловым сервисом, обеспечивающим безопасность для внутренних сервисов системы. Все запросы, которые поступают в систему через API Gateway узел будут проходить проверку на безопасность и корректность.

4. Микросервис Notificator. Позволяет формировать и отправлять оповещения для пользователей на различные мониторинговые события, например, начало и окончание процесса мониторинга элементов или выход из строя элементов системы.

5. Микросервис Data Manager. Сервис реализует функционал по оптимизации процесса сохранения большого объема данных в потоковом режиме и предоставляет возможность получения данных из необходимой базы с высокой производительностью запросов.

6. Базы данных для хранения информации:

– Clickhouse: содержит данные различных вычислений метрик, SLA, SLO и SLI соглашений.

– PostgreSQL: необходима для сохранения структурных данных, содержащих информацию о таких сущностях как метрики, SLA, SLO и SLI.

7. Микросервис Monitoring Manager. Включает в себя основной функционал по обеспечению мониторинга распределенных информационных систем, который основан на разработанных алгоритмах мониторинга, и обеспечивает возможность настройки основных параметров процесса мониторинга.

8. Брокер сообщений Kafka. Используется для работы микросервиса Monitoring Manager. Является источником данных, которые необходимо обработать. Данные в

Кafka попадают из внешних систем. Подход, использующий брокер сообщений, обеспечивает более высокую надежность обработки всей полученной информации.

Рассмотренные выше элементы архитектуры системы мониторинга позволяют в совокупности обеспечивать надежную и непрерывную работу по обеспечению мониторинга распределенных информационных систем. Разработанная микросервисная архитектура представлена на рисунке 6.

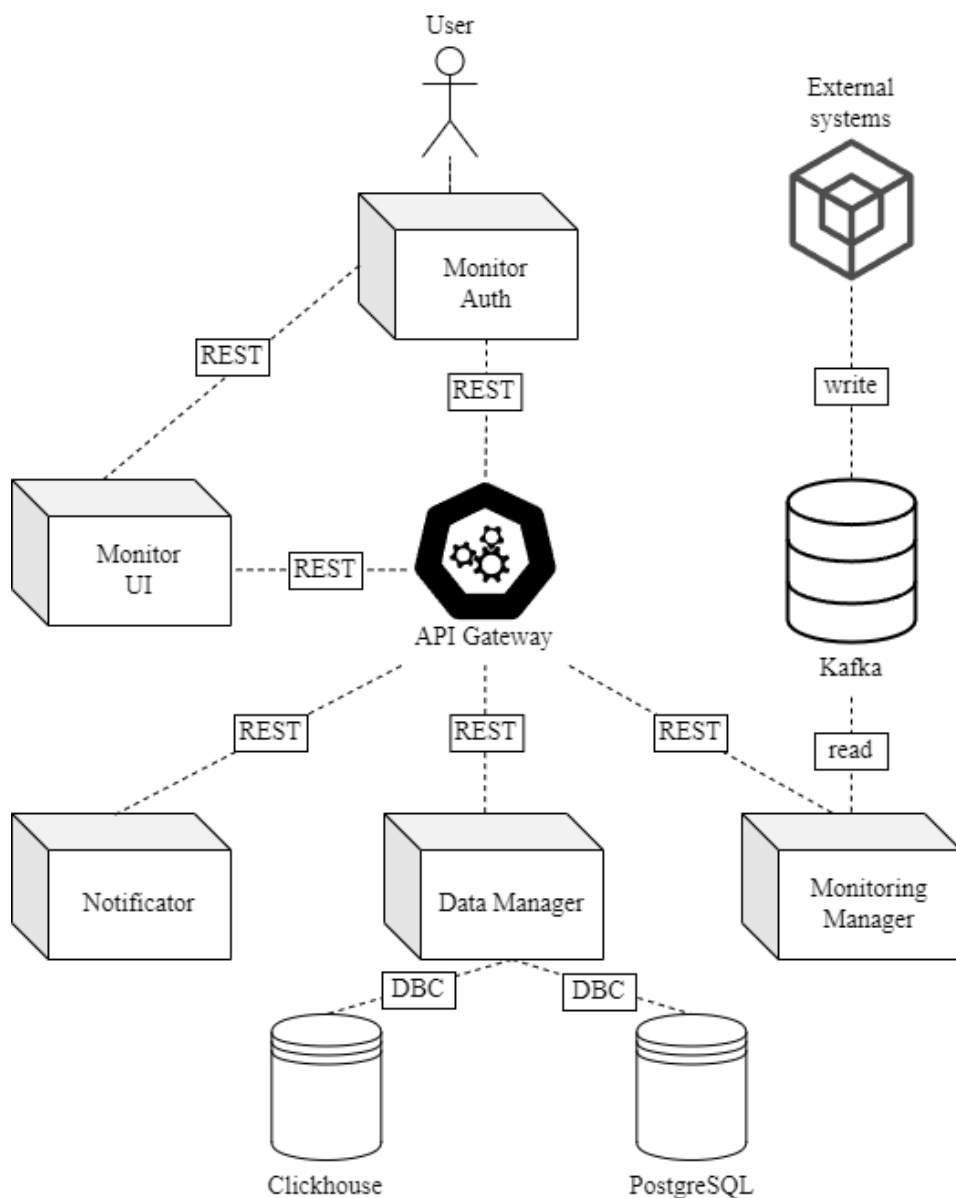


Рисунок 6 – Микросервисная архитектура системы мониторинга

Также в текущей главе рассмотрены используемые инструменты разработки, такие как языки программирования, среды разработки, платформы для установки программного обеспечения и способы взаимодействия разработанного приложения с другими информационными системами.

Приведена оценка результативности применения разработанного программного обеспечения на примере мониторинга распределенной

информационной системы по управлению мобильным доступом. Сделаны выводы о возможностях разработанной мониторинговой программной среды по анализу состояния элементов рассматриваемой распределенной информационной системы.

Заключение

Разработанный мониторинговый инструмент для распределенных информационных систем в гибридной передающей среде позволяет управлять процессом мониторинга максимально автоматизируя данный процесс. Это значительно уменьшает затрачиваемые ресурсы на контроль за состоянием элементов, в особенности для сложных многоуровневых систем. Требуется лишь начальная настройка и дальнейшая поддержка при возникновении каких-либо нештатных ситуаций. При этом вся основная работа по сбору информации о состоянии элементов, анализу их работоспособности и формированию отчетов, будет выполняться разработанным инструментом мониторинга.

Существуют ограничения для применения данного инструмента, так как распределенные информационные системы — это многослойные сложные объекты с разноплановой архитектурой. Следовательно, невозможно одним инструментом реализовать решения для всех существующих систем. Однако текущий мониторинговый инструмент позволяет применить его в большом количестве информационных систем, в особенности, если они работают в сфере предоставления услуг по передаче данных.

На основе проведенных исследований в работе получены следующие основные результаты:

1. Проведен анализ особенностей процесса управления мониторингом распределенных информационных систем в гибридных передающих средах и способов анализа состояния элементов системы.

2. Разработан алгоритм мониторинга распределенной информационной системы на основе поиска компонент сильной связности с обходом структуры системы в глубину с временными метками, позволяющий анализировать состояние элементов системы с помощью многокритериального метода принятия решений, основанного на модели TOPSIS.

3. Разработана модель для расчета качества обслуживания параметров распределенной информационной системы, на основе которого реализован алгоритм, учитывающий контекстные факторы и групповые параметры различных сегментов системы, обеспечивающих повышение качества обслуживания с учетом переменного системного контекста.

4. Реализован алгоритм идентификации и коррекции аномального поведения системы на основе модели расчета потери данных при передаче между элементами системы с использованием процесса имитационного моделирования для анализа состояния системы. Алгоритм использует данные из LRU кэша для оптимизации процесса имитационного моделирования.

5. Спроектирована структура базы данных для хранения результатов процесса мониторинга, отличающаяся использованием колоночной модели хранения и ограниченным временем жизни данных в базе. Данная структура обеспечивает

повышенную производительность для сохранения и получения данных за счет разработанной модели агрегирования данных по типам на таблицы.

6. Разработана структура программного обеспечения системы мониторинга, которая отличается прозрачной и современной интеграцией с существующими информационными системами и обеспечивает гибкую структуру масштабирования ресурсов.

7. Выполнена оценка результатов применения разработанного программного обеспечения на распределенной информационной системе управления мобильным доступом. На основе полученных результатов можно сделать вывод, что работа мониторинговой среды положительно сказывается на работе распределенной информационной системы, при этом не требует значительных ресурсов на ее поддержку. За время тестирования и работы были выявлены следующие результаты:

- разработанная мониторинговая среда успешно прошла нагрузочное тестирование с обработкой 100000 активных элементов;
- системой мониторинга выявлено около 100 проблемных элементов в день;
- собрана статистика примерно из 300 элементов в день, у которых выявляется тенденция к ухудшению состояния;
- уменьшение затрат на ресурсы для мониторингового процесса на 27%;
- уменьшение времени, затрачиваемого на взаимодействие с базой данных хранения мониторинговой информации, на 75%.

Использование разработанных моделей и алгоритмов позволило разработать мониторинговый инструмент, который позволяет применить его для оценки состояния распределенной информационной системы в гибридной передающей среде.

Рекомендации, перспективы дальнейшей разработки темы

Рекомендации при использовании математического аппарата для решения задач мониторинга распределенных информационных систем: использование методов и моделей, которые учитывают гетерогенность рассматриваемых систем и обеспечивают корректные расчеты с учетом предполагаемых масштабов систем.

Рекомендации при выполнении расчетов и экспериментов: использовать большие объемы данных, которые максимально приближены к реальным. Это позволит более качественно анализировать результаты рассматриваемых моделей.

Рекомендации при использовании языка для реализации программной части: программное обеспечение необходимо разрабатывать на языке, поддерживающем современные инструменты разработки, для обеспечения работы с актуальными протоколами передачи данных и интеграции в существующие информационные системы. К таким языкам можно отнести, например, Golang, Java, C# и C++.

Перспективы дальнейшей разработки темы:

1. Методы для анализа возможности и влияния использования различных протоколов передачи данных между частями системы.

2. Добавление новых методов и моделей для идентификации и коррекции аномального поведения системы.

3. Развитие программного аппарата с добавлением более гибкого управления процессом мониторинга системы и улучшением пользовательского интерфейса.

Основные результаты диссертации опубликованы в следующих работах:
Публикации в изданиях списка ВАК

1. Саргсян Э.Р. Прогнозирование поведения сети передачи данных в системе мониторинга телекоммуникационных сетей на основе модифицированного алгоритма Косарайю / Э.Р. Саргсян, А.А Рындин // Вестник Воронежского государственного технического университета. 2020. Т. 16. №3. С. 20 – 26. – Текст: непосредственный.

2. Саргсян Э.Р. Исследование моделей расчета потерь радиосигналов в сетях радиодоступа / Э.Р. Саргсян, А.А Рындин // Моделирование, оптимизация и информационные технологии. – Воронеж: ВИВТ, 2021. Т. 9. №4. 14 с. [Электронный ресурс]. - Режим доступа: <https://moitvivr.ru/ru/journal/pdf?id=1078>.

3. Саргсян Э.Р. Алгоритм проектирования телекоммуникационной сети, основанный на расчетах дифракционных потерь сигнала / Э.Р. Саргсян, А.А Рындин // Системы управления и информационные технологии. – Воронеж: ВГТУ, 2022. №4. С. 61 – 66. – Текст: непосредственный.

4. Саргсян Э.Р. Моделирование и оптимизация процесса передачи данных между базовыми станциями в телекоммуникационной сети / Э.Р. Саргсян, А.А Рындин // Информационные технологии. 2023. №10. С. 522-528. – Текст: непосредственный.

Свидетельство о государственной регистрации программы для ЭВМ

5. Саргсян Э.Р. Автоматизированная система проектирования телекоммуникационных сетей на основе имитационного моделирования. Свидетельство о регистрации программы для ЭВМ № 2022614792 от 25.03.2022.

Статьи и материалы конференций

6. Саргсян Э.Р. Алгоритм обнаружения и коррекции аномального поведения системы / Э.Р. Саргсян, А.А Рындин // Интеллектуальные информационные системы. Труды Международной научно-практической конференции. Воронеж, 2024. С. 205-208.

7. Саргсян Э.Р. Методы оптимизации параметров работы телекоммуникационной радиосети передачи данных / Э.Р. Саргсян // Оптимизация и моделирование в автоматизированных системах. Труды Международной молодежной научной школы. – Воронеж: ВГТУ, 2022. – С. 51–55. – Текст: непосредственный.

8. Sargsyan E.R. Telecommunication data transmission network optimization methods based on the open ran technology / E.R. Sargsyan // Антропоцентрические науки: инновационный взгляд на образование и развитие личности. Материалы XIV Международной научно-практической конференции. – 2021. – Р. 350-352. – Текст: непосредственный.

9. Саргсян Э.Р. Проектирование структуры телекоммуникационной сети с использованием имитационного моделирования // Оптимизация и моделирование в автоматизированных системах. Труды Международной молодежной научной школы. – Воронеж: ВГТУ, 2023. – С. 69–72. – Текст: непосредственный.

10. Саргсян Э.Р. Мониторинг телекоммуникационных сетей передачи данных на основе нейронных сетей / Э.Р. Саргсян, А.А Рындин // Интеллектуальные информационные системы. Труды Международной научно-практической конференции. – Воронеж: ВГТУ, 2019. №2. С. 8–10. – Текст: непосредственный.

11. Саргсян Э.Р. Возможности оптимизации RAN сетей с помощью сетей нового поколения / Э.Р. Саргсян // Оптимизация и моделирование в автоматизированных системах. Труды Международной молодежной научной школы. – Воронеж: ВГТУ, 2021. – С. 153–156. – Текст: непосредственный.

12. Саргсян Э.Р. Решение задачи поиска максимального пути в графе с областями сильной связности / Э.Р. Саргсян // Интеллектуальные информационные системы. Труды Международной научно-практической конференции. – Воронеж: ВГТУ, 2022. – С. 86–90. – Текст: непосредственный.

Подписано в печать 04.10.2024

Формат 60x84/16. Бумага для множительных аппаратов.

Усл. печ. л. 1,0. Тираж 100 экз. Заказ № _____

ФГБОУ ВО «Воронежский государственный технический университет»

394006 Воронеж, 20-летия Октября, 84