АННОТАЦИЯ

диссертационной работы Иманкуловой Бинары Бакытжановны “Разработка и внедрение системы IoT на основе LoRaWAN с учетом происхождения данных для управления качеством воздуха в умном городе”, представленной на соискание степени доктора философии (PhD) по образовательной программе: 8D06102 - «Компьютерная и программная инженерия»

Общая характеристика работы.

В диссертации представлены проектирование и реализация LoRaWAN-фреймворка Интернета вещей (IoT) с поддержкой происхождения данных (provenance). В качестве целевой области применения выбрано управление данными о качестве воздуха в умных городах. Фреймворк обеспечивает доверие к данным мониторинга по трём основным направлениям: происхождение, целостность и полнота. Происхождение здесь описывает, как каждое сохранённое измерение было получено, декодировано, проверено и затем сохранено. Целостность означает возможность обнаружить последующие изменения в записанной истории, а полнота показывает, все ли ожидаемые измерения были получены и, если нет, сколько из них отсутствует.

Фреймворк фиксирует происхождение данных на стороне платформы, после приёма восходящего сообщения. Он не требует изменений протокола LoRaWAN и не добавляет данные о происхождении в полезную нагрузку сенсорного узла. Записи о происхождении соответствуют семантике W3C PROV и хранятся в хеш-связанных цепочках, отдельных для каждого устройства, с контрольными точками головы цепочки. Таким образом, при заданной модели угроз архитектура обеспечивает обнаружение несанкционированных изменений в сохранённой истории. Эти же записи используются для построения графовых представлений в стиле PROV для обратной и прямой трассировки. Для каждого отсутствующего измерения фреймворк определяет этап, на котором произошла потеря: устройство, передача до платформы, приём на платформе или обработка. Причина каждой потери данных устанавливается с помощью счётчиков кадров, времени приёма, состояния записей и журнала событий сетевого сервера. Если имеющихся данных недостаточно, причина классифицируется как неустановленная.

Актуальность работы.

Технологии Интернета вещей (IoT) стали важной частью современной городской инфраструктуры. Сервисы умного города опираются на измерения, которые получают сенсорные узлы и передают через шлюзы, коммуникационные сети и системы обработки на уровне платформы. По текущим прогнозам, к 2030 году число подключённых устройств в мире достигнет 32,1 млрд. Такой рост увеличивает как объём, так и разнородность данных, используемых в мониторинге окружающей среды, управлении

дорожным движением, обеспечении общественной безопасности и других городских приложениях.

Измерения датчиков используются для мониторинга, отчётности, оперативного управления и принятия решений. Прежде чем стать частью рабочего набора данных, измерения проходят этапы считывания, беспроводной передачи, приёма на платформе, обработки, хранения и публикации. На каждом этапе могут добавляться метаданные, меняться формат данных или создаваться новая запись. Кроме того, измерения могут быть потеряны, продублированы, воспроизведены повторно, неверно декодированы или изменены при обработке на стороне платформы.

Такие проблемы не обязательно вызваны злонамеренными действиями. На итоговый набор данных также влияют потеря пакетов, повторный приём одного сообщения несколькими шлюзами, ошибки декодирования и сбои при хранении. Мониторинг качества воздуха особенно чувствителен к таким условиям, поскольку отсутствующие или неверно обработанные значения могут исказить долгосрочную статистику и экологические показатели. Если история обработки не записывается, становится трудно установить, как сохранённое значение было получено, преобразовано и записано в базу данных.

Эту историю фиксирует происхождение данных. Оно описывает источник измерения, условия приёма, операции обработки и итоговую сохранённую запись. Поэтому информация о происхождении данных может обеспечивать прослеживаемость, возможность аудита и проверку целостности. Эти функции важны, когда экологические данные используются для публичной отчётности, научного анализа и долгосрочного городского планирования.

Существующие подходы к происхождению данных обеспечивают разные формы защиты и представления. Системы на основе блокчейна позволяют хранить записи, защищённые от подделки, но часто вносят дополнительные издержки на хранение, задержку и консенсус. Криптографические механизмы позволяют обнаруживать изменение записей, однако могут требовать дополнительных вычислений, процедур управления ключами или изменений протокола. Такие требования трудно выполнить на маломощных сенсорных узлах и в сетях LPWAN.

LoRaWAN широко применяется для распределённого мониторинга окружающей среды, поскольку обеспечивает связь на большие расстояния при низком энергопотреблении. Протокол также обеспечивает конфиденциальность полезной нагрузки и проверку целостности сообщений при передаче. Однако он не сохраняет историю обработки измерения после приёма восходящего сообщения. Поэтому сохранённые и экспортированные записи могут быть изменены без соответствующей записи о происхождении. Недавние исследования показывают, что практических подходов, которые сочетают совместимый с LoRaWAN сбор данных о происхождении, хранение с обнаружением подделки и прослеживаемость, по-прежнему немного.

Ещё одна проблема, которую существующие подходы редко решают, это пропуски данных. В долгосрочных развёртываниях LoRaWAN значительная часть ожидаемых измерений так и не попадает в хранилище: устройства перестают передавать данные, кадры теряются на радиотракте, сообщения теряются между сетевым сервером и платформой или не проходят декодирование. Полноту обычно указывают одним процентом, который показывает, сколько данных отсутствует, но не где они были потеряны. Без этой информации операторы не могут решить, нужно ли ремонтировать устройства, расширять покрытие шлюзов или исправлять платформу, а аналитики не могут оценить, случайны пропуски или систематичны.

Систематический обзор литературы в главе 1 выявил четыре пробела в существующих решениях по происхождению данных для IoT и систем умного города: сбор данных о происхождении, который требует изменений на оконечных устройствах или охватывает лишь часть конвейера обработки; отсутствие сведений о радиоприёме в стандартных моделях происхождения; заявления о целостности, не привязанные к явной модели угроз; отсутствие привязки пропущенных данных к этапу конвейера. Эти пробелы определяют пять требований к LoRaWAN-платформе с поддержкой происхождения данных: сбор данных о происхождении на всех этапах без изменений на оконечных устройствах, стандартная модель происхождения, включающая сведения о радиоприёме, обнаружение подделки сохранённых данных о происхождении при заданной модели угроз, ограниченные накладные расходы и привязка пропущенных данных к этапу конвейера. В диссертации разработан фреймворк, удовлетворяющий этим требованиям.

Цель и задачи исследования. Целью диссертации является разработка фреймворка с поддержкой происхождения данных на стороне платформы для прослеживаемого управления данными о качестве воздуха с обнаружением несанкционированных изменений в системах мониторинга умного города на основе LoRaWAN.

Для достижения цели в диссертации решаются четыре **задачи**:

1. проанализировать существующие подходы к происхождению данных в IoT и выявить их ограничения в системах умного города и сетях LPWAN;
2. разработать и формализовать облегчённую модель происхождения данных, целостности и привязки пропущенных данных к этапу конвейера для мониторинга на основе LoRaWAN;
3. разработать и реализовать прототип на стороне платформы, который собирает, хранит и проверяет записи о происхождении данных, а также восстанавливает их графовые представления;
4. оценить прототип в реальном развёртывании мониторинга на LoRaWAN и с помощью моделирования по полноте данных и привязке пропущенных данных, накладным расходам платформы, производительности проверки целостности и прослеживаемости происхождения данных.

Объектом исследования является жизненный цикл измерительных данных IoT в системах мониторинга умного города на основе LoRaWAN.

Предметом исследования являются методы и механизмы сбора данных о происхождении, защиты их целостности и обеспечения прослеживаемости на протяжении жизненного цикла измерений на стороне платформы.

Методы исследования. В исследовании сочетались четыре типа методов: аналитические, проектные, эмпирические и имитационные. Аналитический этап включал систематический обзор литературы, проведённый в соответствии с PRISMA 2020. В нём рассматривались происхождение данных в IoT, защита целостности, полнота данных и связь в сетях LPWAN, а также были определены требования к предлагаемому фреймворку. Затем на этапе проектирования эти требования были воплощены в архитектуру на стороне платформы и формальную модель формирования данных о происхождении, проверки целостности и привязки пропущенных данных к этапу конвейера. Кроме того, был реализован функциональный прототип, который фиксирует происхождение данных после приёма каждого восходящего сообщения (uplink), хранит его в хеш-связанных цепочках для каждого устройства, проверяет целостность каждой цепочки и восстанавливает графовые представления для запросов прослеживаемости.

Далее на эмпирическом этапе прототип оценивался в ходе 12-месячного развёртывания LoRaWAN для мониторинга качества воздуха. Оценивались полнота данных, производительность приёма данных на серверной части, накладные расходы на хранение, время проверки цепочек происхождения, обнаружение подделки и прослеживаемость. Привязка пропущенных данных к этапам оценивалась с помощью компьютерного моделирования. Радиоуровень моделировался в сетевом симуляторе ns-3 с модулем LoRaWAN, к нему добавлялись сбои устройств, шлюзов и платформы с известными причинами, чтобы сравнить установленные причины потерь с истинными.

Научная новизна

Предложена модель происхождения данных на стороне платформы для мониторинга на основе LoRaWAN, в которой сведения о приёме сообщения шлюзом представлены как явная сущность W3C PROV и не требуется изменений устройств или протокола. Это подтверждено обратной и прямой трассировкой сохранённых измерений до сведений об их приёме.

Разработана схема хранения с обнаружением несанкционированных изменений, адаптированная к жизненному циклу восходящих сообщений LoRaWAN. Схема сочетает хеш-цепочки для каждого устройства и контрольные точки головы цепочки. Она обнаружила все пять проверенных типов изменения записей и проверила 403 900 записей за 5,87 с.

Предложен метод привязки пропущенных измерений к этапам конвейера, который, в отличие от агрегированных показателей полноты, позволяет установить, где были потеряны данные: на устройстве, при передаче до платформы, при приёме на платформе или при обработке. В моделировании на ns-3 метод допустил 3 ошибки на 270 107 атрибутированных измерений.

Основные положения, выносимые на защиту:

1. Архитектура LoRaWAN-фреймворка с поддержкой происхождения данных на стороне платформы для мониторинга качества воздуха, которая фиксирует происхождение данных, проверяет его целостность и привязывает пропущенные данные к этапам конвейера без какого-либо изменения протокола LoRaWAN или полезной нагрузки сенсорных узлов.

2. Модель происхождения данных для жизненного цикла восходящих сообщений LoRaWAN, которая представляет этапы приёма, загрузки и хранения в виде сущностей, действий и агентов W3C PROV, включая сведения о приёме сообщений шлюзами, и обеспечивает как обратную, так и прямую прослеживаемость сохранённых измерений.

3. Метод хранения данных о происхождении с обнаружением несанкционированных изменений на основе хеш-связанных цепочек для каждого устройства и контрольных точек головы цепочки, который при заданных допущениях о доверии обнаруживает изменение, удаление, вставку и перестановку записей, а также усечение хвоста цепочки.

4. Метод привязки пропущенных измерений к этапам конвейера, который использует сведения, зафиксированные платформой и сетевым сервером, чтобы отнести каждое пропущенное измерение к одному из этапов (устройство, передача до платформы, приём на платформе или обработка) либо классифицировать его как неустановленное.

Научные результаты

1. Проведён систематический обзор литературы в соответствии с PRISMA 2020. Из 1 201 записи, найденной в Scopus, Web of Science и дополнительных источниках, в обзор включено 90 исследований. По результатам обзора, существующие подходы редко сочетают сбор данных о происхождении на стороне платформы, применимость к LoRaWAN, хранение с обнаружением несанкционированных изменений, прослеживаемость и привязку пропущенных данных к этапам.

2. Спроектирован и реализован LoRaWAN-фреймворк с поддержкой происхождения данных для мониторинга качества воздуха. Фреймворк фиксирует сведения о приёме, обработке и хранении после получения каждого восходящего сообщения без изменения протокола или полезной нагрузки сенсорных узлов.

3. Разработана формальная модель происхождения данных и целостности. Модель задаёт стабильные идентификаторы для устройств, восходящих сообщений, приёмов шлюзами, измерений и записей о происхождении и представляет жизненный цикл на стороне платформы тремя действиями: RX, INGEST и STORE. В ходе развёртывания получено 135 000 записей RX, 135 000 записей INGEST и 133 900 записей STORE.

4. Разработан метод хранения с обнаружением несанкционированных изменений на основе хеш-связанных цепочек для каждого устройства и контрольных точек головы цепочки. Проверка 403 900 записей заняла 5,87 с (0,0145 мс на запись), обнаружены все пять проверенных типов изменений. В частности, усечение хвоста цепочки обнаружено сравнением с контрольной точкой за 0,94 мс.

5. Разработан метод восстановления графовых представлений в стиле PROV по запросу. Метод поддерживает обратную трассировку от сохранённого измерения до исходного восходящего сообщения и прямую трассировку от восходящего сообщения до сохранённых записей без хранения постоянного глобального графа.

6. Фреймворк оценён в ходе 12-месячного развёртывания с тремя сенсорными узлами и двумя шлюзами, в результате которого получено 135 000 логических восходящих сообщений, 133 900 сохранённых измерений и 403 900 записей о происхождении. Эффективная полнота составила 84,9 %, база данных заняла 164,08 МБ, из них 114,14 МБ приходится на записи о происхождении.

7. Проведено парное сравнительное тестирование на 5 000 логических восходящих сообщений в 10 прогонах. Сбор данных о происхождении увеличил медианную задержку обработки с 1,565 до 2,189 мс на сообщение, то есть на 0,62 мс.

8. Метод привязки пропущенных измерений к этапам оценён с помощью моделирования с радиоуровнем на основе ns-3, откалиброванным по данным развёртывания. В 20 прогонах продолжительностью один год каждый только 3 из 270 107 атрибутированных измерений отнесены к неверному этапу. Покрытие составило 66,4 % (93,7 % без повторных подключений устройств), а при отсутствии журнала событий сетевого сервера точность снизилась до 96,5 %.

Теоретическая и практическая значимость работы.

Теоретическая значимость состоит в предложенной модели происхождения данных для систем мониторинга на основе технологий LPWAN. Модель описывает жизненный цикл измерений IoT на стороне платформы, от приёма восходящего сообщения до сохранения в базе данных, и объединяет сбор данных о происхождении, проверку их целостности и привязку пропущенных данных к этапам. Кроме того, предложенная модель может служить основой для дальнейших исследований проверяемой прослеживаемости и полноты данных в IoT-средах с ограниченными ресурсами.

Практическая значимость работы заключается в реализованном прототипе LoRaWAN с поддержкой происхождения данных для мониторинга качества воздуха в умных городах. Прототип работает на стороне платформы без изменения протокола LoRaWAN или полезной нагрузки сенсорных узлов и позволяет анализировать пропущенные измерения, ошибки декодирования и нарушения целостности цепочек происхождения. В частности, результаты показывают, что журнал событий сетевого сервера необходимо хранить независимо от сборщика платформы, иначе сбои платформы невозможно отличить от потерь на радиотракте.

Связь данной работы с другими исследовательскими проектами.

Исследование, представленное в диссертации, было начато в рамках проекта «*Программно-аппаратный комплекс для анализа и мониторинга климатических и экологических изменений окружающей среды*» (грант №

AP05134597-ОТ-20 Министерства образования и науки Республики Казахстан, 2018-2020 гг.). Продолжение исследования будет осуществляться в рамках проекта для молодых учёных «Жас ғалым» (грант № AP32523239, 2026-2028 гг.).

Апробация работы. Основные положения и научные результаты диссертации были представлены и обсуждены на следующих научных мероприятиях:

1. научные семинары кафедры компьютерной инженерии Международного университета информационных технологий, Алматы, Казахстан;
2. 5-я Международная конференция по интеллектуальным информационным системам и технологиям (SIST 2025), Астана, Казахстан;
3. X Международная научная конференция «Математика, механика и информационные технологии» (ICMMIT 2025), Казахский национальный университет имени аль-Фараби, Алматы, Казахстан;
4. научный семинар в Astana IT University, Астана, Казахстан (28 августа 2026 г.);
5. международная конференция по цифровым технологиям и их применению (ICDTA 2026), Vaughn College of Aeronautics and Technology, Нью-Йорк, США (9–10 октября 2026 г.).

Результаты диссертации внедрены в SDU University (Каскелен, Казахстан), что подтверждается актом внедрения, приведённым в приложении Е.

Научные публикации. Результаты диссертации опубликованы в 6 научных работах: 3 статьи в журналах, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования Министерства науки и высшего образования Республики Казахстан (МНВО РК), 2 статьи в журналах, индексируемых в Scopus и Web of Science (WoS), и 1 статья в сборнике трудов международной конференции.

1. **B. Imankulova**, W. M. S. Yafooz, K. Alshammari, B. Alshemali, A. E. Yahya and A. Abd Almisreb. Tracing Data Origins in Smart Cities: An IoT Perspective. IEEE Access. 2025. Vol. 13, pp. 130201-130214. DOI: 10.1109/ACCESS.2025.3591277 [WoS Q2, Scopus percentile: 91, Cite Score(2025): 9.3, available: https://ieeexplore.ieee.org/document/11087496]
2. Mukash Zh., Bakranova D., Serikkanov A., Bakranov N., **Imankulova B.** Energy-Aware AI-Enabled Intelligent Sensing Systems for IoT and Edge–Cloud Architectures. Sensors 2026, Volume 26, Issue 21, 6927, 10.3390/s26216927 [WoS Q2, Scopus percentile: 83, Cite Score (2025): 9.4]
3. **B. Imankulova**, A. Sagymbekova. Privacy Challenges in Environmental IoT Systems: A Case Study on Air Quality Monitoring, 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2025, pp. 1-6, doi: 10.1109/SIST61657.2025.11139338 [Scopus]
4. **Imankulova B.B**. Privacy challenges and framework design for IoT devices: A comprehensive approach. Bulletin of AUPET. 2025. Vol. 68, Issue 1. p. 256–267. DOI: https://doi.org/10.51775/2790-0886_2025_68_1_256 [КОКНВО]

5. **Imankulova, B. B**., Kozina, L. A., Duzbayev, N. T., Amanzholova, S. T. Transmission, storage and preprocessing of data received from sensors using LoRa technology. Bulletin of AUPET. 2024. Vol. 64, Issue 1. p.16-24. DOI: https://doi.org/10.51775/2790-0886_2024_64_1_16 [КОКНВО]

6. **Imankulova B.B.**, Amanzholova S.T., Duisebekova K.S. Software and technical aspects of data collection from end nodes: problems and prospects. Bulletin of AUPET. 2020. Vol. 51, Issue 4. p.164-172. DOI: https://doi.org/10.51775/1999-9801_2020_51_4_164 [КОКНВО]

Основное содержание диссертации

В главе 1 рассматривается современное состояние исследований в области происхождения данных IoT для систем умного города. В главе анализируются понятия происхождения данных, требования к данным умного города, модели представления, механизмы обеспечения целостности и полнота данных, а также приводится систематический обзор литературы, проведённый в соответствии с PRISMA 2020. По результатам обзора выявлены четыре пробела в исследованиях и сформулированы пять требований к LoRaWAN-платформе с поддержкой происхождения данных.

В главе 2 представлено проектирование предлагаемого фреймворка. В главе описаны принципы проектирования и многоуровневая архитектура, интеграция со стеком связи LoRaWAN, а также модель угроз и точный смысл заявленной защищённости. Далее представлена формальная модель фреймворка: компоненты системы и идентификаторы, записи и граф происхождения, а также модели формирования данных о происхождении, проверки целостности и привязки пропущенных данных к этапам. В конце главы описаны хранение и проверка данных о происхождении.

В главе 3 описана реализация фреймворка: архитектура платформы, полевое развёртывание, приём данных через MQTT и фиксация происхождения, схема хранения, процедура проверки целостности, запросы к данным о происхождении и процедура привязки пропущенных данных к этапам.

В главе 4 фреймворк оценивается по результатам 12-месячного развёртывания: полнота данных, качество канала и работа адаптивной скорости передачи данных (ADR), накладные расходы на хранение и обработку, проверка целостности и прослеживаемость. Кроме того, привязка пропущенных данных к этапам оценивается относительно известных причин с помощью моделирования с радиоуровнем LoRaWAN на основе ns-3. В конце главы приведены сравнение с наиболее близкими исследованиями и обсуждение ограничений.

В заключении подведены основные результаты и вклад диссертации, описаны ограничения предлагаемого фреймворка и определены направления дальнейших исследований.

Структура и объём диссертации

Диссертация изложена на 155 страницах, содержит 34 рисунка, 22 таблицы, список литературы из 171 источника и 4 приложения.