

А.В. ВОРОБЬЕВ, Г.Р. ВОРОБЬЕВА
**МОДЕЛЬ ИНФОРМАЦИОННОГО ВЗАИМОДЕЙСТВИЯ
ЭЛЕМЕНТОВ МНОГОУРОВНЕВОЙ СИСТЕМЫ ЦИФРОВЫХ
ДВОЙНИКОВ**

Воробьев А.В., Воробьева Г.Р. Модель информационного взаимодействия элементов многоуровневой системы цифровых двойников.

Аннотация. Одним из решений проблемы пространственно-временной анизотропии данных является применение многоуровневой системы цифровых двойников, базирующихся на соответствующих отраслевых моделях и пополняемой базе архивных данных. Применение указанного подхода успешно зарекомендовало себя в информационных системах мониторинга параметров геомагнитного поля и его вариаций, обеспечивая пространственно-временную интерполяцию геомагнитных данных с точностью до 0.81 нТл в магнитоспокойные периоды. При этом нерешенной осталась задача информационного взаимодействия между уровнями системы цифровых двойников, что в значительной степени усугубляется постоянно растущим объемом данных и их неоднородным характером. В работе предлагается решение обозначенной задачи посредством формализованного механизма пакетирования пространственно-временной информации, при котором идентификация источников данных выполняется на базе иерархической системы бинарной токенизации. Кроме того, рассматривается предложенная программная реализация такого подхода, отличительной особенностью которой является комбинирование традиционной клиент-серверной и инновационной бессерверной архитектур для реализации высоконагруженного реактивного веб-приложения для работы с анализируемыми данными. Выделены и программно формализованы основные этапы реализации информационного взаимодействия – от получения исходной информации от ее источников до верификации данных, их анализа, обработки и формирования выходного информационного потока системы. Результаты проведенных вычислительных экспериментов на примере задачи мониторинга параметров магнитного поля Земли и его вариаций подтвердили эффективность предложенных решений, выраженную как в повышении реактивности веб-ориентированных приложений, так и в увеличении вычислительной скорости формирования и заполнения информационных хранилищ, агрегирующих информацию из распределенных гетерогенных источников.

Ключевые слова: цифровые двойники, пространственные данные, информационное взаимодействие, единое информационное пространство, обработка данных, токенизация, геоинформационные системы и технологии

1. Введение. Одним из магистральных направлений развития технологий на протяжении длительного времени является обеспечение эффективной обработки и анализа данных, описывающих сложные процессы и явления различной природы происхождения в контексте их пространственно-временного распределения. Соответствующие системы мониторинга представляют собой комплексы информационно-измерительных устройств, с различной плотностью распределенные по земной поверхности.

Так, к примеру, мониторинг геомагнитного поля и его вариаций традиционно выполняется множеством (порядка 300) магнитных обсерваторий и вариационных станций, которые в режиме реального времени регистрируют соответствующие геомагнитные данные. Данные регистрируются ежеминутно и за время существования сети магнитных обсерваторий накоплено порядка 22,8 Гб геомагнитных данных, представляющих собой временные ряды в CSV-подобном IAGA-формате. Учитывая то, что некоторые магнитные обсерватории предоставляют данные, зарегистрированные с шагом в 1 с, объем накапливаемых при этом данных пропорционально возрастает. При этом практически на каждой магнитной обсерватории в регистрируемых временных рядах ежедневно наблюдаются от 5 до 58% пропущенных значений.

Общей проблемой для подавляющего большинства систем мониторинга процессов и явлений является выраженная пространственная и временная анизотропия, неразрешимая методами интерполяции ввиду специфики анализируемых процессов (как правило, естественной природы происхождения). Не менее серьезной является неполнота соответствующих временных рядов, обусловленная перебоями в системах связи, сбоями в работе оборудования, а также человеческим фактором. Так, к примеру, анализ результатов расчета показателя относительной информационной эффективности для некоторых магнитных обсерваторий сети INTERMAGNET за 2015 г. [1] показал, что в среднем общее время работоспособного состояния для отдельных обсерваторий составляет от 64,08 до 87,95% времени наработки.

Обозначенные проблемы являются существенным препятствием на пути обработки пространственно-временных данных, что, в свою очередь, негативно сказывается как на процессе принятия решений в отраслевых областях, так и на получении новых знаний в соответствующей прикладной области.

2. Состояние вопроса. Современные подходы к решению задачи пространственно-временной анизотропии данных во многом базируются на применении интеллектуальных геоинформационных технологий. Так, к примеру, широко известны и повсеместно практикуются такие технологии пространственной интерполяции, как крикинг, сплайн, обратно-взвешенные расстояния, метод ближайшего соседа и др. [1]. Однако, к примеру, для геомагнитных данных анализ показал, что применение известных методов пространственной интерполяции сопровождается среднеквадратической ошибкой в среднем от 7,3 до 11,2 нТл при допустимой стандартами погрешности в

1 нТл. В результате интерполированные данные могут повлечь за собой серьезные ошибки в моделировании и принятии решений, что, в свою очередь, сопряжено с финансовыми и временными затратами на устранение их последствий (например, при инклинометрических исследованиях с применением магниточувствительного оборудования).

Восстановление временных рядов пространственных данных реализуется разнообразными методами как математической, так и геостатистики. Так, на ранних этапах исследований для восстановления данных преимущественно использовались методы линейной интерполяции и кубического сплайна, эффективные для устранения единичных пропусков, но непригодные для импутации более длительных фрагментов ввиду возникновения существенных частотных и амплитудных искажений [2]. Более сложные методы восстановления временных рядов базируются на аналитической обработке известных значений уровня, а также анализе периодических и сезонных составляющих ряда. Так, к примеру, известен метод гравитационного сглаживания временных рядов [3], основанный на понятии гравитационной непрерывности и связанный с квадратичной регрессией. Метод позволяет восстанавливать длительные фрагменты временных рядов, полностью исключая амплитудные и минимизируя частотные искажения. Однако необходимость решения при этом линейных систем большого порядка сопряжена с высокой трудоемкостью применения метода и, как следствие, низкой вычислительной скоростью средств их программной реализации, что недопустимо при оперативной аналитической обработке данных. Для восстановления данных со сложной нерегулярной структурой и многочисленными локальными особенностями применяется метод непрерывного вейвлет-преобразования [4]. Так, к примеру, в [5] показано, что отличие восстановленного по коэффициентам вейвлет-разложения временного ряда от оригинального не превышает 15%.

Вместе с тем ни один из известных подходов не учитывает особенности пространственной анизотропии интерполируемых данных и принимает одни и те же модели и методы для различных географических областей.

3. Постановка задачи. Одним из вариантов решения проблемы пространственной анизотропии является использование системы цифровых двойников [6], которые в совокупности с их физическими прототипами обеспечивают повышение плотности информационных измерений для соответствующих процессов или явлений различной природы происхождения.

Основанием для выделения уровней системы цифровых двойников является пространственная кластеризация непосредственно двойников и их физических прототипов. На основании заданного параметра разбиение исходного множества источников данных на подмножества осуществляется известными геоинформационными методами, в частности, по результатам расчета индекса Морана-I [7], традиционно являющегося мерой пространственной автокорреляции пространственных объектов. Верификация цифровых двойников осуществляется методами машинного обучения с учетом данных, регистрируемых физическими прототипами.

При объединении цифровых двойников и их физических прототипов в многоуровневую структуру их информационное взаимодействие приобретает особую специфику, обусловленную необходимостью сохранения данных при их интеграции в единое информационное пространство в составе соответствующих систем поддержки принятия решений.

Вместе с тем необходимо формализовать процедуру информационного взаимодействия элементов многоуровневой системы цифровых двойников таким образом, чтобы обеспечить интеграцию, где каждый из элементов систем в отдельности и их агрегации выступает самостоятельным источником данных для систем поддержки принятия решений.

Для решения поставленной задачи необходимо выполнение следующих этапов исследования:

1. Разработка концепции информационного взаимодействия в системе цифровых двойников, учитывающей ее многоуровневую структуру, пространственную кластеризацию источников данных и формирование исходящего информационного потока во внешнюю информационную систему (или систему поддержки принятия решений).
2. Формализация обозначенной концепции в теоретико-множественном базисе для систематизации требований к построению системы информационного межуровневого взаимодействия цифровых двойников и их физических прототипов.
3. Разработка системы самоидентификации источников данных в многоуровневой системе цифровых двойников.
4. Определение и формализация моделей активного поведения программных модулей, реализующих информационное взаимодействие в системе цифровых двойников и

обеспечивающих оперативную передачу информации между уровнями системы в соответствии с заданными критериями.

5. Разработка программной архитектуры системы информационного взаимодействия цифровых двойников, обеспечивающей эргономичный открытый доступ к данным в различных интерфейсах – как на уровне конечных пользователей, так и для применения внешними программными модулями и системами.

В качестве сквозного примера в работе рассматривается система магнитных обсерваторий и вариационных станций, которые в режиме реального времени регистрируют параметры геомагнитного поля и его вариаций – геомагнитные данные. На данном примере был проведен ряд вычислительных экспериментов по оценке целесообразности объединения цифровых двойников и их физических прототипов в многоуровневую систему. В ходе исследований были проанализированы годовые минутные данные более чем 300 магнитных станций, неравномерно распределенных по трем пространственным областям – высокоширотные, среднеширотные и приэкваториальные станции. Анализ результатов экспериментов показал, что дополнение физических прототипов соответствующими цифровыми двойниками позволит снизить количество пропусков в среднем на 54,3 % в среднеширотных и приэкваториальных областях и на 27,4% в высокоширотных областях.

4. Концепция и модель информационного взаимодействия в многоуровневой системе цифровых двойников. Ключевой идеей информационного взаимодействия в многоуровневой системе цифровых двойников является создание информационных узлов на каждом из ее уровней (рис. 1). При этом на нижнем уровне абстракции имеют место информационные потоки, с заданной периодичностью направляемые от цифровых двойников и их физических прототипов в выделенные информационные хранилища. Далее указанные информационные хранилища объединяются внутри соответствующего локального уровня. На более высоком уровне абстракции информационные потоки с локальных уровней поступают в информационные хранилища отдельных кластеров, которые, в свою очередь, передают свои данные в централизованное хранилище данных, например, в выделенное единое информационное пространство в составе системы поддержки принятия решения в отраслевой области.

В общем виде указанную концепцию можно формализовать в теоретико-множественном базисе следующим образом.

Пусть задана пара вида «физический прототип – его цифровой двойник»:

$$\{P, T\} : P \cup T \neq \emptyset; P \neq T, P \not\subset T;$$

$$P = \{p_1, p_2, \dots, p_K\}; T = \{t_1, t_2, \dots, t_K\},$$

где P – физический прототип, T – цифровой двойник, p_1, p_2, \dots, p_K – атомарные данные, зарегистрированные физическим прототипом в моменты времени $1 \dots K$, t_1, t_2, \dots, t_K – атомарные данные, полученные в результате моделирования цифровым двойником за те же моменты времени $1 \dots K$.

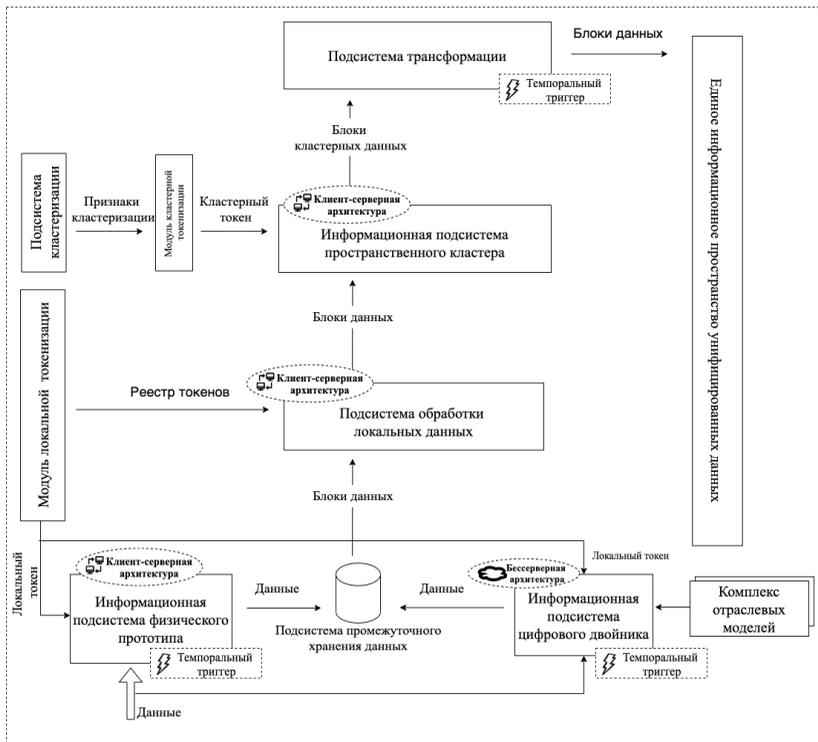


Рис. 1. Обобщенная схема информационного взаимодействия в системе цифровых двойников

В процессе обработки информационных потоков, синхронно поступающих от цифрового двойника и его физического прототипа с

заданным временным интервалом, данные агрегируются в специализированном промежуточном хранилище, выделенном для каждой пары вида «физический прототип – его цифровой двойник». Далее полученный массив анализируется программным модулем, в основе построения которого лежат модели и методы машинного обучения. По результатам анализа из двух входных информационных потоков формируется один набор данных, содержащий результаты верификации наблюдений и выборку наиболее достоверных (с точки зрения используемого модулем математического алгоритма и специфики прикладной предметной области) значений.

В общем виде данный процесс можно описать в виде следующих соотношений. Пусть в результате промежуточной интеграции информационных потоков за период времени K формируется массив данных вида:

$$D = P \times T = \{ \{ p_1, t_1 \}, \{ p_2, t_2 \}, \dots, \{ p_K, t_K \} \},$$

где D – промежуточный массив как результат попарной интеграции элементов данных, получаемых от физического прототипа P и его цифрового двойника T .

В подсистеме промежуточного хранения выполняется анализ представленных в D пар значений. Применяется процедура трансформации F , которая на основании регрессионного анализа накопленных ретроспективных данных задает однозначное отображение пар значений физического прототипа и двойника в элемент нового множества D' :

$$D' = \{ d_1, d_2, \dots, d_K \} \mid \forall i = 1, \dots, K \{ p_i, t_i \} \xrightarrow{F} d_i,$$

где d_i – элемент нового множества, полученный в результате применения процедуры трансформации к каждой из K пар данных, полученных от физического прототипа P и его цифрового двойника T .

Все множество пар вида «физический прототип – его цифровой двойник» является информационной базой для хранилища локальных данных, куда направляются все прошедшие аналитическую трансформацию и обработку данные из множества D' .

Пусть для каждого источника данных, например, магнитной обсерватории или вариационной станции, имеется пара вида «физический прототип – цифровой двойник». Таким образом, имеет место соотношение вида:

$$\exists N \in R : \forall i \in N \exists \{P, T\}^i : P^i \cup T^i \neq \emptyset; P^i \neq T^i; P^i \not\subset T^i;$$

$$P^i = \{p_1^i, p_2^i, \dots, p_K^i\}; T^i = \{t_1^i, t_2^i, \dots, t_K^i\},$$

где N – количество доступных источников данных, каждый из которых представлен парами вида «физический прототип – его цифровой двойник».

Тогда в результате трансформации для каждого источника данных из N формируется промежуточное хранилище для K моментов времени следующего вида:

$$D^{ij} = \{d_1^j, d_2^j, \dots, d_K^j\} | \forall i = 1, \dots, K \{p_i^j, t_i^j\} \xrightarrow{F} d_i^j, \forall j = 1, \dots, N,$$

где D^{ij} – массив данных промежуточного хранилища j -го источника данных, для которого существует пара «физический прототип P^j – его цифровой двойник T^j », которые в каждый i -й из K моментов времени предоставляют данные вида p_i^j и t_i^j соответственно.

В результате формирования хранилища локальных данных за период времени K может быть представлено в виде объединения соответствующих множеств D^{ij} . Иными словами, имеет место соотношение вида:

$$Z = D^{11} \cup D^{12} \cup \dots \cup D^{1N},$$

где Z – множество данных, поступающих за период K от промежуточных хранилищ N источников данных.

При этом

$$Z = \{d_1^1, d_2^1, \dots, d_K^1, \dots, d_1^j, d_2^j, \dots, d_K^j\} |$$

$$d_i^j \in D^{ij}; i = 1, \dots, K; j = 1, \dots, N,$$

где d_i^j – элемент множества Z , соответствующий данным, полученным от j -го источника данных за i -й момент времени.

В соответствии с требованиями заинтересованных лиц, определяемыми в процессе проектирования информационных систем (в том числе систем поддержки принятия решений), выделяется перечень признаков, на основании которых источники данных разделяются на кластеры. Например, применительно к решению задач геомагнетизма

такovým выступает разделение земной поверхности на высокоширотные, приэкваториальные и среднеширотные области. В соответствии с этим критерием магнитные обсерватории и вариационные станции могут быть в соответствии с их пространственной привязкой отнесены к высокоширотному, приэкваториальному и среднеширотному кластерам.

В этой связи представляется целесообразным ввести следующие соотношения.

Пусть заданы M пространственных кластеров, в каждом из которых представлены от 1 до L источников данных в соответствии с заданным признаком (признаками) кластеризации:

$$\exists G = \{G_1, G_2, \dots, G_M\} : \{S_1, \dots, S_L\} \in G_i,$$

где G – множество пространственных кластеров мощностью M , i -й элемент которого представляет собой отдельный пространственный кластер G_i , каждый из которых, в свою очередь, содержит от 1 до L источников данных S .

Для каждого пространственного кластера G_l должна быть сформирована собственная база данных (база данных l -го пространственного кластера), куда поступают срезы данных из локального хранилища данных Z :

$$Z_{G_l} = Z' \subset Z:$$

$$Z' = \{d_1^1, d_2^1, \dots, d_K^1, \dots, d_1^j, d_2^j, \dots, d_K^j\} |:$$

$$i = 1, \dots, K; j = 1, \dots, R; R \leq N,$$

где Z' – подмножество множества данных из локального хранилища, которое содержит данные вида d_K^j для K моментов времени R источников данных в составе l -го пространственного кластера, при этом число источников данных в источнике, очевидно, не должно превышать числа доступных источников данных.

Важно отметить, что в базе данных пространственного кластера представлены данные не непосредственно физических прототипов источников информации, а те значения, которые были помещены в соответствующие промежуточные хранилища после аналитической трансформации данных пары «физический прототип – его цифровой двойник».

При этом одни и те же источники данных могут быть представлены более, чем в одном кластере. При этом для заданного

набора кластеров не исключается существование таких источников данных, которые не принадлежат ни одному из возможных пространственных кластеров. Такие источники данных будут рекластеризованы в дальнейшем при введении дополнительных признаков.

На завершающем этапе выполняется интеграция баз данных пространственных кластеров в единое информационное пространство в составе информационной системы (системы поддержки принятия решений). Данные подвергаются дополнительной обработке (трансформации) для приведения в унифицированную форму.

5. Система иерархической токенизации элементов системы цифровых двойников. Для повышения эффективности процессов обмена и обработки данных между уровнями системы цифровых двойников предлагается подход, основанный на принципе токенизации информационных ресурсов.

Технология токенизации предполагает замену некоторого реального значения суррогатным, активно используется для защиты и шифрования в современных системах информационной безопасности, а также для реализации систем блокчейн [8, 9].

В настоящей работе предлагается использовать токенизацию для разметки источников данных и их физических прототипов, а также результатов их агрегирования на основании пространственной кластеризации по заданным признакам.

В соответствии с предлагаемой концепцией информационного взаимодействия элементов системы цифровых двойников предполагается маркировка соответствующих наборов данных для идентификации их источника в процессе обработки и анализа информации в составе единого информационного пространства (информационной системы, системы поддержки принятия решений и пр.).

Регистрируемые физическим прототипом или моделируемые его цифровым двойником данные проходят многоэтапную обработку. Так, для физического прототипа на начальном этапе данные попадают в модуль регистрации, где фиксируются принятым в источнике данных способом, например, в виде простых электронных таблиц или файлов csv-подобного формата. Далее выполняется их предварительная обработка – устранение артефактов, обусловленных несовершенством информационно-измерительного оборудования, а также человеческим фактором. На следующем этапе к данным применяются алгоритмы сжатия для экономии дискового пространства, с одной стороны, и упрощения передачи данных на следующий уровень, с другой. На

последующем этапе пакетирования данные упаковываются в объект, по структуре подобный HTTP-сообщению (заголовок и тело сообщения). В заголовке проставляются идентификационные данные источника, направляющего информационный поток. Данные в заголовке являются, в том числе, и авторизационными, поскольку позволяют принимающему их модулю оценить, что является источником информации.

Например, для геомагнитных данных одним из вариантов такой авторизационной информации является заголовок, предусмотренный форматом IAGA-2002 [10]. Заголовок содержит основную информацию о магнитной обсерватории / вариационной станции: название, уникальный код, ведомственная принадлежность, пространственные координаты, характеристики оборудования, регистрируемые параметры и пр. С каждый новым блоком передаваемых данных метадаанные, которые и представлены в заголовке, передаются в точку отправления, нагружая как каналы передачи данных (что существенно при насыщенном информационном потоке), так и средства их обработки, которые вынуждены «очищать» полученный набор данных от служебной информации.

Для совершенствования процедуры обработки и передачи данных между уровнями системы цифровых двойников предлагается использовать систему токенизации. При этом каждый токен представляет собой «свертку» метадаанных о соответствующем источнике информационного потока, который при необходимости можно декомпозировать и выделить необходимую для работы служебную информацию.

Предлагается подход, основанный на последовательном расширении токена источника данных по мере перемещения с уровня на уровень в системе цифровых двойников. Исходный код токена формируется на начальном уровне – для отдельных физических прототипов и цифровых двойников. Код является глобально уникальным для всех возможных источников данных, задействованных в системе цифровых двойников. Код токена является суррогатным и представляет собой бинарный код, сформированный посредством перевода в двоичный формат уникального десятичного номера.

При этом предлагается следующий формат представления токена на локальном уровне: $L: Type - XXXXX - (ParentID)$, где L – метка локального уровня (неизменный параметр); $Type$ – тип источника данных (неизменный параметр), домен допустимых значений которого представлен в виде (P, T) , где P – соответствует физическому прототипу, T – цифровому двойнику; $XXXXX$ – уникальный двоичный

код источника данных, сформированный из его порядкового номера, представленного в десятичном формате (неизменный параметр); *ParentID* – ссылка на уникальный двоичный код физического прототипа (параметр, обязательный только для тех меток, которые используются для токенизации источников данных на локальном уровне).

Каждый новый источник данных, который будет использован в системе цифровых двойников, проходит обязательную процедуру токенизации, программно реализуемой посредством инструментария подсистемы токенизации, предусмотренной в предложенной архитектуре системы. Каждому источнику данных присваивается порядковый регистрационный номер (чем обеспечивается его уникальность в системе цифровых двойников), значение которого переводится в двоичный формат.

Далее в случае, если выполняется токенизация цифрового двойника в реестре источников данных, осуществляется поиск его физического прототипа, определяется его уникальный двоичный идентификатор, который поставляется в позицию «*ParentID*» формулы токенизации источника данных на локальном уровне. В результате обеспечивается однозначное соответствие между физическими прототипами и их цифровыми двойниками.

Вопрос ссылочной целостности при кодировании токена цифрового двойника и его прототипа предлагается решать следующим образом. По аналогии с реляционными базами данных предполагается применение строгого правила ссылочной целостности, согласно которому невозможны манипуляции с изменением или удалением токена физического прототипа в случае наличия соответствующего ему цифрового двойника. При этом изменение или удаление токена цифрового двойника никак не сказывается на токене соответствующего ему физического прототипа.

Токен источника данных является его уникальным идентификатором в реестре источников данных системы цифровых двойников, на основании которого можно получить информацию о служебных параметрах (метаданных). В частности, метаданные относительно географического расположения источников данных применяются при формировании пространственных кластеров с соответствующей токенизацией составляющих их источников данных.

В соответствии с заданными признаками выделяются пространственные кластеры, каждый из которых (по аналогии с источниками данных) токенизируется уникальным образом: выполняется преобразование в двоичный формат на основании

уникального порядкового номера пространственного кластера, представленного изначально в десятичном формате.

В результате предлагается следующая форма представления токена пространственного кластера: *Cluster: XXXXX*, где *Cluster* – флаг, указывающий на применение токена к идентификации пространственного кластера; *XXXXX* – уникальный двоичный код пространственного кластера, сформированный из его порядкового номера, представленного в десятичном формате (неизменный параметр).

По аналогии с именованием атрибутов в реляционной и свойств в объектной моделях данных полный токен каждого задействованного в системе цифровых двойников источника данных будет складываться из двух токенов: локального и кластерного, разделенных символами «← →».

Иными словами, в общем виде полный токен источника данных можно представить следующим образом: *Cluster: XXXXX – L: Type – XXXXX – (ParentID)*.

Уникальность полного токена источника данных обеспечивается при этом отсутствием пересечений множеств источников данных, образующих пространственный кластер. Иными словами, должно быть выполнено условие вида:

$$\neg \exists G_i \in \{S_1, \dots, S_{L1}\} \ \& \ G_j \in \{S_1, \dots, S_{L2}\} : G_i \cap G_j \neq \emptyset; \\ \forall G_i, G_j \in G = \{G_1, G_2, \dots, G_M\},$$

где G – множество пространственных кластеров мощностью M , i -й элемент, который представляет собой отдельный пространственный кластер G_i , каждый из которых, в свою очередь, содержит от 1 до L источников данных S .

При этом предполагается, что пара вида «физический прототип – его цифровой двойник» должна быть представлена в составе одного и только одного пространственного кластера:

$$S_i = (P_i, T_i), \forall S_i \in G_j, \forall G_j \in G = \{G_1, G_2, \dots, G_M\}.$$

Недопустимы ситуации, при которых физический прототип источника данных (например, информационно-измерительного устройства или иного технически сложного объекта/системы) и его цифровой двойник отнесены к разным пространственным кластерам,

поскольку сочетание двойника и его прототипа определяет единый информационный поток, исходящий из источника данных.

В результате поэтапного изменения токена имеет место его многоуровневая структура, согласно которой каждый новый уровень системы цифровых двойников добавляет к токену свою метку (так же токен), который позволяет отследить иерархию родительских по отношению к источнику данных (любого типа – физического прототипа или его цифрового двойника) элементов. Например, таким образом можно определить, для какого пространственного кластера используется источник данных, что позволит дополнить анализируемую информацию в том числе и набором метаданных об ее источнике.

Кроме того, отличительной особенностью предлагаемого подхода является двойственное применение токена, что зависит от технических возможностей соответствующей информационной системы, реализующей описанную выше концепцию многоуровневой организации цифровых двойников.

С одной стороны, уникальный токен может быть использован как идентификатор источника данных в соответствующем реестре (базе данных). Передача токена в запросе как один из вариантов фильтрации данных позволит обратиться непосредственно к кортежу, который содержит искомую служебную информацию об источнике данных. Преимущества, связанные с простотой реализации и возможностью интеграции в существующую информационную архитектуру, несколько нивелируются тем недостатком, что помимо метаданных необходимо хранить еще и токен.

Другой вариант применения токена источника данных – «свертка» метаданных непосредственно в идентификатор. Применение алгоритма сжатия данных Deflate [11] позволяет выполнить преобразование заданной блочной структуры в компактную «свертку», представленную в шестнадцатеричном формате. Недостаток такого подхода, связанный с необходимостью введения дополнительного программного модуля для реализации процедуры кодирования/декодирования соответствующего токена, компенсируется тем, что метаданные непосредственно встроены в сообщение, передаваемое в модуль обработки или централизованное хранилище от источника данных. Задачей указанного программного модуля, по сути, является простое считывание из токена цепочек бит, упакованных в байты, формирование сжатого блока при кодировании и извлечение оригинального значения при его декодировании.

Еще одним применением токена является решение задачи авторизации источника данных, направляющего информационный поток в модуль обработки или выделенное хранилище данных. При поступлении данных в программный модуль выполняется процедура детокенизации – выделения токена из поступившего сообщения, его декодирования и выделения метаданных. Далее результаты анализа передаются в тот компонент хранилища, который соответствует заданному токеном пространственному кластеру (или отдельному источнику данных любого типа).

6. Темпоральные триггеры в системе цифровых двойников.

Для реализации процедуры информационного взаимодействия компонентов системы цифровых двойников предложен механизм, получивший название темпорального триггера.

В основе данного подхода лежит фундаментальное понятие триггеров, применимое к транзакционной системе баз данных. Согласно известному термину [12], триггер представляет собой программную процедуру, запускаемую на выполнение в случае наступления в базе данных одного из трех событий – создания, удаления или модификации кортежа (или группы кортежей) данных.

Применительно к системе цифровых двойников событием для привязки к триггеру является достижение определенной временной метки. Предполагается, что темпоральный триггер по принципу своего функционирования подобен программному таймеру и на физическом уровне он реализуется с помощью соответствующих программных процессов-демонов, которые выполняются по принципу Cron [13].

В архитектуре системы цифровых двойников предполагается применение трех темпоральных триггеров. Два из них должны быть задействованы на уровне атомарных источников данных – физических прототипов и их цифровых двойников. По истечению заданного периода времени данные, накопленные в виртуальном буфере, выгружаются в подсистему промежуточного хранения, где выполняется их анализ, обработка и формирование единого информационного блока для передачи на следующий уровень системы.

Еще один темпоральный триггер функционирует на уровне пространственного кластера. По сути, он выполняется по тому же принципу, что и два предшествующих: по истечению заданного временного интервала выполняется выгрузка данных из базы данных пространственного кластера в единое информационное пространство.

Таким образом, темпоральный триггер запускает на выполнение процедуру выгрузки наборов данных из хранилища одного уровня в систему хранения более высокого уровня.

Настройка темпоральных триггеров осуществляется при проектировании системы цифровых двойников и должна учитывать специфику соответствующей прикладной (отраслевой) области. Так, к примеру, для геомагнитных данных предполагается установка темпорального триггера в 1-минутный интервал относительно заданной временной точки. Другой пример – данные с прогнозом вероятности полярных сияний в области аврорального овала. Здесь темпоральный триггер целесообразно установить во временной интервал, не менее 5-минутного.

7. Программная архитектура информационного взаимодействия системы цифровых двойников. В общем виде процесс информационного взаимодействия в системе цифровых двойников представляет собой сбор данных, их краткосрочное размещение в промежуточных хранилищах с последующей передачей через уровни системы в единое информационное пространство. Каждый подпроцесс информационного взаимодействия программно оформлен в виде подсистемы (совокупности модулей) или отдельных модулей.

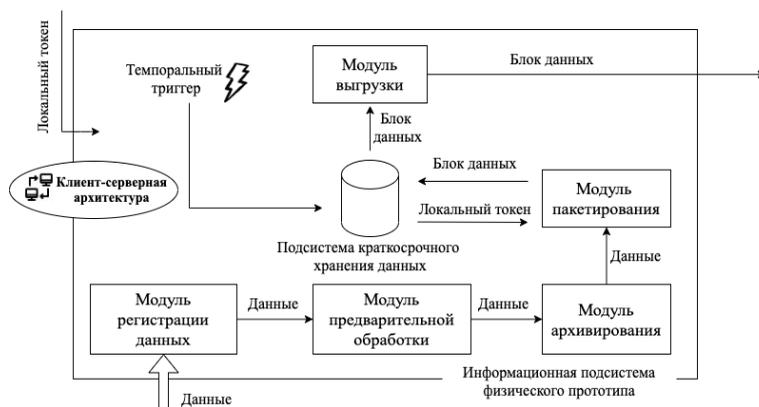
Схема системы информационного взаимодействия представлена на рисунке 1. Выделены четыре программных уровня, каждый из которых обеспечивает сбор, обработку и хранение информации с учетом уровня иерархии токена источника данных (рис. 2).

На начальном этапе осуществляется процедура получения информации физическими прототипами на основании результатов мониторинга некоторых процессов / явлений / объектов, полученных как непосредственно с измерительных устройств, так и с иных источников информации. Данные поступают в модуль регистрации, результатом работы которого является массив данных унифицированного формата (указываемого при проектировании информационной системы). Далее данные направляются в модуль предварительной обработки, где выполняется устранение ошибок, допущенных несовершенством технического обеспечения, а также имеющих место ввиду человеческого фактора. Далее к данным применяется механизм сжатия (модуль архивирования), в основу работы которого положен алгоритм по типу gzip [14].

Прошедший архивирование набор данных передается на вход модулю пакетирования, который присоединяет к архивированным данным локальный токен и передает полученный массив данных для последующей отправки в подсистему промежуточного хранения данных. При этом локальный токен физического прототипа размещается в той же подсистеме, которая формирует исходящий набор

данных, а поступает туда изначально из внешнего модуля локальной токенизации. Последний, в свою очередь, формирует идентификатор локального токена применительно к источнику данных на основании его уникального номера и метаданных, полученных в процессе регистрации.

Аналогичным образом функционирует информационная подсистема цифровых двойников в рамках описываемой системы. Поскольку цифровой двойник по определению [15] представляет собой математическую модель соответствующего физического прототипа, основным компонентом его информационной подсистемы является модуль моделирования. Результатом его выполнения является формирование набора данных (в заданном на этапе проектирования формате), который соответствует синхронно зарегистрированным значениям физического прототипа. Данные, как и в случае с физическим прототипом, проходят процедуру архивирования и пакетирования. При этом локальный токен аналогичным образом передается в подсистему из модуля локальной токенизации, поскольку цифровой двойник, также как и его физический прототип, регистрируется в реестре источников данных в составе информационной подсистемы.



а)

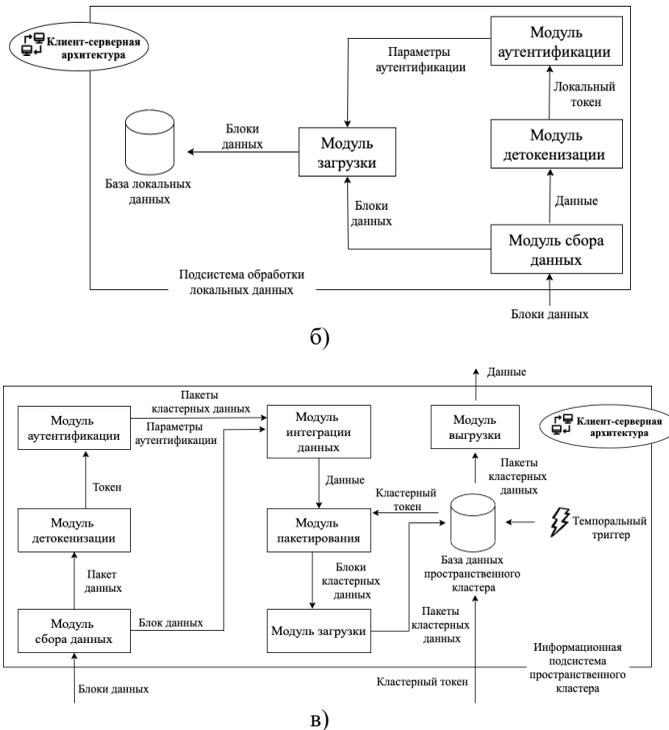


Рис. 2. Структурные схемы подсистем информационного взаимодействия системы цифровых двойников: а) информационная подсистема физического прототипа; б) подсистема обработки локальных данных; в) информационная подсистема пространственного кластера

Целесообразно отметить, что модуль выгрузки данных в информационных подсистемах как физического прототипа, так и его цифрового двойника запускается на выполнение в фоновом режиме соответствующим процессом-демоном, имитирующим заданный темпоральный триггер с применением принципов Stop.

В подсистеме промежуточного хранения выполняется верификация синхронных данных, полученных от физического прототипа и его цифрового двойника. На основании уже накопленной информации, выступающей в качестве прецедентов, формируется исходящий из подсистемы информационный поток, в котором представлены данные, наиболее вероятные и достоверные с точки зрения соответствующей регрессионной модели. Следующим уровнем архитектуры выступает подсистема обработки локальных данных.

Точкой входа подсистемы выступает модуль сбора данных, сохраняющий в виртуальный кэш сервера поступающие из распределенных источников данных и прошедшие предварительную обработку информационные потоки.

Далее управление передается модулю детокенизации, основным назначением которого является выделение локального токена из поступившего в составе информационного потока блока данных. Токен считывается из сформированного на предыдущем этапе набора данных (заголовка сформированного пакета).

На следующем этапе выделенный токен передается модулю аутентификации, который сопоставляет его значение с идентификаторами в реестре источников данных. В случае соответствия блок передается в данных окружающих соответствующую секцию базы локальных данных через программный модуль загрузки. В противном случае источник данных классифицируется как неопознанный, в соответствующую подсистему сообщается о возникшей ошибке, транзакция по записи информации в базу локальных данных отменяется.

Для подсистемы обработки локальных данных выгрузка информационных блоков из базы данных осуществляется в соответствии с заданным на этапе проектирования темпоральным триггером, выполнение которого реализуется соответствующими программными процессами-демонами, функционирующими по принципу Cron.

На следующем этапе управление передается информационной подсистеме пространственного кластера. По аналогии с описанной выше информационной подсистемой поток данных поступает с предшествующего уровня на вход модулю сбора данных. Выделенный в результате работы модуля детокенизации токен представлен в своем полном формате, включая значение токена соответствующего пространственного кластера. На уровне модуля аутентификации выполняется оценка принадлежности источника данных, от которого поступил информационный блок, соответствующему пространственному кластеру. В случае обнаружения несоответствия источник данных классифицируется как неопознанный, в соответствующую подсистему направляется сообщение о возникшей ошибке, а транзакция по записи информации в базу данных пространственного кластера отменяется. При успешном завершении процедуры аутентификации источника данных соответствующий блок данных передается в модуль интеграции, основным назначением которого является агрегирование всех блоков данных, поступающих от

источников данных пространственного кластера в течение заданного временного интервала.

На следующем шаге данные из модуля интеграции поступают для обработки в модуль пакетирования, где к агрегированному набору данных добавляется токен соответствующего пространственного кластера. Последний, в свою очередь, поступает в подсистему в результате работы модуля кластерной токенизации программной подсистемы токенизации синхронно с регистрацией нового пространственного кластера в реестре. Примечательно, что при необходимости присутствующий в блоках данных токен кластера может быть обновлен в соответствии с тем значением, которое присваивается заголовку блока в модуле пакетирования.

Модуль загрузки помещает данные для краткосрочного хранения в базу данных пространственного кластера. По аналогии с предшествующими уровнями системы информационного взаимодействия, для подсистемы обработки пространственного кластера выгрузка информационных блоков из базы данных осуществляется в соответствии с заданным на этапе проектирования темпоральным триггером, выполнение которого реализуется соответствующими программными процессами-демонами, функционирующими по принципу Cron.

В результате выполнения модулей подсистемы пакеты кластерных данных консолидируются в выделенном хранилище в составе единого информационного пространства для последующего применения в информационных системах (системах поддержки принятия решений).

8. Особенности программной реализации информационного взаимодействия системы цифровых двойников. Одним из технических требований к системе цифровых двойников является обеспечение оперативного доступа к данным, территориально распределенным на физически удаленных вычислительных машинах. Эффективным способом организации решения, удовлетворяющего указанному критерию, является применение клиент-серверной архитектуры, которая обеспечивает возможность единого доступа к данным большому количеству пользователей (и программ) одновременно.

Результатом программной реализации предложенного в работе решения по информационному взаимодействию системы цифровых двойников является веб-ориентированное клиент-серверное приложение, бизнес-логика которого построена по принципу бессерверных вычислений [16–18]. Преимуществом выбранного

подхода является то, что он позволяет избежать необходимости выделения и настройки специализированных серверов.

На физическом уровне разработанная программная система представлена совокупностью информационных подсистем физических прототипов, пространственного кластера и обработки локальных данных. Все компоненты системы реализованы на базе фреймворка Django [19–21], одной из отличительных особенностей которого является поддержка архитектурного паттерна «Модель – Представление – Контроллер». Указанная особенность Django в рамках решения задачи программной реализации системы цифровых двойников позволила в явном виде отделить задачи хранения, обработки и визуализации соответствующих данных.

При этом доступ к отдельным компонентам приложения реализован по принципу инфраструктуры CDN (Content Delivery Network – сеть доставки содержимого) [22], что позволило обеспечить, соответственно, более высокую скорость доступа для клиентского рендеринга по сравнению с другими способами организации веб-инфраструкты.

Важно отметить, что в рамках предложенной реализации для каждого источника данных (независимо от того, является ли он физическим прототипом или его цифровым двойником) выделяется собственная информационная подсистема, построенная по принципу организации клиент-серверной архитектуры и представляющая собой, по сути, компоненту единой информационной многоуровневой системы цифровых двойников. Каждая подсистема оформлена в виде облачной функции, реализованной на основе платформы Yandex.Cloud [24]. Возможность динамического масштабирования выделенной инфраструктуры, поддержка языка программирования Python (используемого и в рамках остальных компонент системы информационного взаимодействия), контейнеруемость – стали определяющими параметрами для выбора данного способа реализации информационной подсистемы. В результате применения архитектуры бессерверных вычислений для решения поставленной задачи была достигнута высокая реактивность подсистемы, составившая порядка 0,28 с времени отклика на атомарный запрос в противовес 78,41 с для «классической» клиент-серверной архитектуры.

Для конечных пользователей (например, операторов системы со стороны источников данных или соответствующих программных сервисов) в рамках указанной клиент-серверной архитектуры в соответствии с предложенной программной реализацией предусмотрен единый веб-ориентированный интерфейс доступа, посредством

которого осуществляется управление загрузкой регистрируемых или моделируемых наборов данных с заданным временным интервалом и в заданном формате взаимодействия. Дальнейшая обработка данных выполняется исключительно на серверной стороне, в то время как на уровне клиентского приложения (интерфейса) возможно одностороннее наблюдение за процессом смены состояний данных.

Так, согласно предложенной программной реализации, данные, поступившие через клиентский веб-ориентированный интерфейс от физического прототипа (как вариант – магнитной обсерватории или вариационной станции) последовательно передаются от одного серверного модуля другому, проходя соответственно основные этапы преобразования – от предварительной обработки до пакетирования и передачи на следующий уровень (в подсистему промежуточного хранения данных).

Предложен подход, согласно которому передача данных между программными модулями на уровне веб-сервера осуществляется посредством перенаправления результирующих данных с выхода одного модуля на вход другому в виде виртуальных информационных потоков (stream) без дополнительного физического сохранения на дисковом буфере. Результатом выполнения всех операций преобразования на этом уровне в конечном итоге является так называемый пакет или блок данных, сформированный путем добавления уникального токена к архивированному и прошедшему обработке набору исходных данных, полученных непосредственно от технической системы (например, информационно-измерительного устройства магнитной обсерватории или вариационной станции, непрерывно регистрирующего данные о параметрах магнитного поля Земли и его вариациях).

Поскольку одним из определяющих факторов информационного взаимодействия системы цифровых двойников является скорость передачи соответствующих информационных потоков, то в процессе реализации предложенных решений возникает задача представления данных, передаваемых между уровнями системы. В этой связи в качестве основного был принят формат колоночного представления данных [23], который являет собой бинарное представление агрегированных в столбцы записей и предполагает обязательное архивирование (например, на основе популярного алгоритма сжатия gzip). К основным преимуществам такого подхода относятся высокая реактивность программных средств для обработки данных такого типа, а также относительно низкие вычислительные затраты, связанные с физическим хранением больших наборов данных. Указанные

возможности позволят успешно решить ряд проблем обработки данных, связанных с их большим и непрерывно растущим объемом (кроме того, возможно сокращение затрат на хранение и представление сопровождающей атрибутивные данные пространственной информации).

Также важно отметить, что в соответствии с предложенной программной реализацией в подсистеме промежуточного хранения данных, куда поступают потоки данных от информационных подсистем физических прототипов и их цифровых двойников, инкапсулирован модуль аналитической обработки, реализующий модели и методы статистического анализа, в частности, регрессионного анализа. В результате выполнения операций, предусмотренных данным модулем, осуществляется верификация двух наборов данных – физического прототипа и цифрового двойника, и на основании накапливаемой базы прецедентов определяется тот набор данных, который содержит наиболее вероятные для данных пар значения.

9. Анализ эффективности предложенной инфраструктуры.

Для оценки эффективности предложенных моделей и методов информационного взаимодействия компонентов многоуровневой системы цифровых двойников была разработана информационная веб-ориентированная система, обеспечивающая визуализацию геомагнитных данных, регистрируемых магнитными обсерваториями и вариационными станциями, а также моделируемая соответствующими цифровыми двойниками.

Проект «Geomagnetic Plots» (режим доступа: <https://geomagnetic.ru>) реализован на основе фреймворка Django и предназначен для оперативной веб-ориентированной визуализации пространственных данных, описывающих динамику изменения во времени параметров магнитного поля Земли и его вариаций [25]. В качестве источников данных выступают магнитные обсерватории и вариационные станции, а также их цифровые модели, верификация которых выполняется программным модулем, реализующим процедуру регрессионного анализа (рис. 3).

Работа веб-проекта «Geomagnetic Plots» иницируется конечным пользователем: вводятся (выбираются посредством элемента управления) дата и время, а также визуализируемый параметр магнитного поля. На основании передаваемых на сервер параметров выполняется поиск соответствующих геомагнитных данных, представленных в едином информационном пространстве. Последние, в свою очередь, передаются в систему из многоуровневой системы цифровых двойников, информационное взаимодействие которых

реализовано по предложенному в работе принципу и на основании описанной выше смешанной клиент-серверной и бессерверной архитектур.

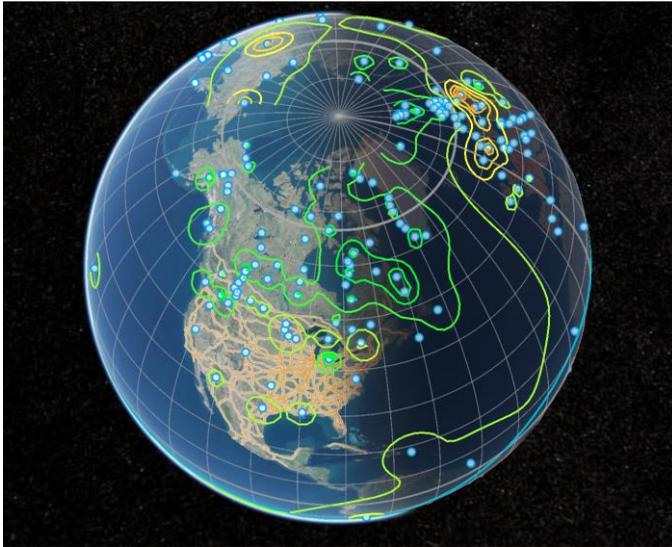


Рис. 3. Экранная форма веб-проекта «Geomagnetic Plots»

Для сравнения был разработан исследовательский прототип веб-ориентированной системы, реализующей функциональность, аналогичную проекту «Geomagnetic Plots» [26]. Однако формирование и заполнение единого информационного пространства здесь реализовано посредством системы прямых запросов к источникам данных (физическим прототипам и их цифровым двойникам) в строгом соответствии с архитектурой «клиент-сервер».

Оценка качества обеих программных систем была проведена на основании методики, представленной в ГОСТ 28195-89 «Оценка качества программных средств». В соответствии с этой методикой были оценены показатели надежности, сопровождения, удобства применения, эффективности, универсальности и корректности программной системы. Тестирование было проведено в нормальных и экстремальных условиях, а также исключительных ситуациях. Согласно результатам вычислительных экспериментов, обе программные системы корректно выполняют свои функции. В экстремальных и исключительных ситуациях программы выдают

сообщения о соответствующей ошибке и продолжают работу в штатном режиме.

Кроме того, в ходе проведения вычислительных экспериментов была выполнена сравнительная оценка времени отклика соответствующих программных систем. Анализ результатов эксперимента показал, что выполнение запроса к данным и рендеринга их в условиях применения персонального компьютера со средней производительностью (процессор с частотой 1.6 ГГц, 2 ядра, оперативная память 4 Гб, скорость интернет-соединения 342.7 Мбит/с) в обоих случаях занимает в среднем около 6 с, что меньше времени отклика при использовании практикуемого способа, составляющего ~49 с (~12.2 % от исходного времени).

Значительные различия в производительности были обнаружены при исследовании выполнения фоновых серверных процессов каждой из программных систем, связанных с передачей данных между уровнями системы цифровых двойников и их выгрузкой в единое информационное пространство.

Предложенный подход, основанный на применении промежуточных виртуальных хранилищ и комбинировании традиционной клиент-серверной и бессерверной программных архитектур, сравнивался с практикуемым подходом последовательного клиент-серверного опроса каждого источника данных (как физического прототипа, так и его цифрового двойника).

Вычислительные эксперименты были проведены для годовых геомагнитных данных 147 магнитных обсерваторий и цифровых двойников на веб-сервере с характеристиками следующего порядка: 72 * Intel(R) Xeon(R) Gold 6140 CPU @ 2.30GHz.

Результаты проведенных экспериментов показали, что применение предложенного подхода к решению задачи информационного взаимодействия компонентов системы цифровых двойников позволит повысить время выполнения запроса в среднем в 71,3 раза по сравнению с монолитной клиент-серверной архитектурой доступа к физическим прототипам и цифровым двойникам.

10. Заключение. В настоящее время проблема пространственно-временной анизотропии особенно остро проявляется при решении задач изучения объектов и процессов окружающей среды. Методы математической статистики, геостатистики, геоинформационного моделирования и других областей знаний частично решают задачи пространственной и временной интерполяции, но не одинаково эффективно для различных пространственных областей и прикладных задач. К примеру, пространственная интерполяции геомагнитных

данных такими методами обеспечивает значение среднеквадратической ошибки, существенно выше допустимой геофизическими стандартами. Похожие результаты наблюдаются и в других областях.

Одним из вариантов решения этой проблемы является применение системы цифровых двойников, которые на основе регрессионных моделей и данных физических прототипов позволяют агрегировать наиболее вероятные значения данных за искомые пространственные и временные интервалы. При этом информационное взаимодействие между уровнями системы значимо усугубляется постоянно растущим объемом и сложностью регистрируемых и обрабатываемых данных.

В работе для решения поставленной задачи была предложена организация информационного взаимодействия в системе цифровых двойников на основе формализованного механизма пакетирования пространственно-временной информации, при котором идентификация источников данных выполняется посредством иерархической системы бинарной токенизации. Отличительной особенностью предложенного подхода является самоидентифицируемость источников данных, обеспечиваемая включением в состав информационного пакета соответствующего токена, который содержит необходимые для анализа метаданные источника данных. Еще одним отличием является предложенная архитектура информационного взаимодействия, предполагающая буферное хранение данных для организации их предварительной обработки перед передачей информационного пакета либо последующему уровню, либо во внешний программный модуль.

Проведенные вычислительные эксперименты (на примере геомагнитных данных) показали, что применение предложенных моделей и методов информационного взаимодействия для реализации системы цифровых двойников позволит:

- 1) повысить скорость сбора, обработки и агрегирования информации с физических прототипов и их цифровых двойников за счет минимизации долговременного хранения промежуточных данных посредством применения темпоральных триггеров;
- 2) увеличить реактивность сценариев сбора, обработки и агрегирования информации с физических прототипов и их цифровых двойников посредством комбинирования клиент-серверной и бессерверной программных архитектур;
- 3) сократить расходы на физическое хранение данных, агрегируемых с физических прототипов и их цифровых двойников посредством формирования специализированных

- пакетов данных, содержащих идентификационный токен в своем заголовке и архивированные бинарные данные в теле;
- 4) упростить процедуру идентификации источников данных в процессе сбора, обработки и агрегирования информации с физических прототипов и их цифровых двойников посредством введения многоуровневой системы токенизации.

Ряд сервисов информационной системы размещен на внешнем веб-сервере с открытым физическим доступом, а остальные находятся стадии интеграционного тестирования и отладки на локальном виртуальном сервере. Дальнейшие исследования предполагают полное размещение сервисов на внешнем ресурсе и последующее развитие заявленной функциональности.

Литература

1. *Воробьев А.В., Воробьева Г.П.* Подход к оценке относительной информационной эффективности магнитных обсерваторий сети INTERMAGNET // Геомагнетизм и аэронавигация. 2018. Т. 58. № 5. С. 648–652.
2. *Гвишиани А.Д., Лукьянова Р.Ю., Соловьёв А.А.* Геомагнетизм: от ядра Земли до Солнца. М.: РАН, 2019. 186 с.
3. *Гвишиани А.Д., Агаян С.М., Богоутдинов Ш.Р., Каган А.И.* Гравитационное сглаживание временных рядов // Труды Института математики и механики УрО РАН. 2011. Т.17. № 2. С. 62–70.
4. *Рябова С.А.* Геомагнитные вариации и синхронные с ними вариации уровня подземных вод и микросейсмического фона в условиях средних широт: дисс. на соискание ученой степени кандидата физико-математических наук. М., 2018.
5. *Kondrashov D., Shprits, Y., Ghil, M.* Gap filling of solar wind data by singular spectrum analysis // Geophys. Res. Lett. 2010. no. 37. P. L15101 DOI:10.1029/2010GL044138
6. *Аббасова Т.С.* Развитие виртуальных инструментов для создания цифровых двойников // Информационно-технологический вестник. 2019. № 2(20). С. 79–88.
7. *Chen Y.* New Approaches for Calculating Moran's Index of Spatial Autocorrelation // PLoS one. 2013. no. 8. P. e68336. DOI: 10.1371/journal.pone.0068336.
8. *Freni P., Ferro E., Moncada R.* Tokenization and Blockchain Tokens Classification: a morphological framework // IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 2020, pp. 1-6, DOI: 10.1109/ISCC50000.2020.9219709.
9. *Marchewka-Bartkowiak K., Nowak K.* Get Tokenized... The Specificity of Personal Tokens in the Context of Tokenization and Axiological Categorization // Proceedings of the 3rd International Conference on Economics and Social Sciences. 2020. p. 823-831. DOI: 10.2478/9788395815072-081.
10. *Love J.* Intermagnet and the Global Community of Magnetic Observatories // AGU Fall Meeting Abstracts. 2006.
11. *Stocchero M.* Iterative deflation algorithm, eigenvalue equations, and PLS2 // Journal of Chemometrics. 2019. no.33. p.10.1002/cem.3144.
12. *Le H. A.* Trigger2B: A Tool Generating Event-B Models from Database Triggers // Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. 2021. vol. 343. pp.1-7. DOI: 10.1007/978-3030-67101-38.

13. *Lenz M.* Silent-Cron: A Cron Wrapper // Raku Fundamentals. 2020. pp. 59-80. DOI: 10.1007/978-1-4842-6109-5_6.
14. *Voss N., Becker T., Mencer O., Gaydadjiev G.* Rapid Development of Gzip with MaxJ // Lecture Notes in Computer Science. 2017. vol. 10216. p. 60-71. DOI: 10.1007/978-3-319-56258-2_6
15. *Mittal H.* Digital twin: An overview // CSI Communications. 2020. no. 44.
16. *Carver J., Penzenstadler B., Scheuner J., Staron M.* (Research) Insights for Serverless Application Engineering // IEEE Software. 2021. vol. 38. pp. 123-125. DOI: 10.1109/MS.2020.3028659.
17. *Mahmoudi N., Khazaei H.* Performance Modeling of Serverless Computing Platforms // IEEE Transactions on Cloud Computing. 2020. pp. 1-1. DOI: 10.1109/TCC.2020.3033373.
18. *Martins H., Araujo F., Cunha P. R.* Benchmarking Serverless Computing Platforms // Journal of Grid Computing. 2020. vol. 18. DOI: 10.1007/s10723-020-09523-1.
19. *Ali S., Alauldeen R., Ruqa A.* What is Client-Server System: Architecture, Issues and Challenge of Client-Server System (Review) // HBRP Publication. 2020. pp.1-6. DOI: 10.5281/zenodo.3673071.
20. *Zhu Y., Wu W., Li D.* Efficient Client Assignment for Client/Server Systems // IEEE Transactions on Network and Service Management. 2016. no. 13. pp. 1-10. DOI: 10.1109/TNSM.2016.2597269.
21. *Iskandar T., Lubis M., Kusumasari T., Lubis A.* Comparison between client-side and server-side rendering in the web development // IOP Conference Series Materials Science and Engineering. 2020. no. 801. pp. 012136. DOI: 10.1088/1757-899X/801/1/012136.
22. *Zeng Z., Zhang H.* A Study on Cache Strategy of CDN Stream Media // IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2020. pp. 1424-1429. DOI: 10.1109/ITAIC49862.2020.9338805.
23. *Sharma K., Marjit U., Biswas U.* Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet // Information Technology and Libraries. 2017. no.37. pp. 29-34. DOI:10.6017/ital.v37i3.10177.
24. *Kostinskiy S., Shaikhutdinov D., Narakidze N.* Loss Counter in Power Double Winding Transformers Implementing the Method of Conditionally Constant Coefficients in Online Mode Using the Information Platform // Известия высших учебных заведений. Электромеханика. 2020. no. 63. pp.79-85. DOI: 10.17213/0136-3360-20-5-79-85.
25. *Vorobev A.V., Pilipenko V.A., Enikeev T.A., Vorobeva G.R.* Geoinformation system for analyzing the dynamics of extreme geomagnetic disturbances from observations of ground stations // Computer Optics. 2020. no. 44(5). pp. 782-790. DOI: 10.18287/2412-6179-CO-707.
26. *Воробьев А.В., Воробьева Г.Р., Юсупова Н.И.* Концепция единого пространства геомагнитных данных // Труды СПИИРАН. 2019. в. 18, т. 2. с. 390–415.

Воробьев Андрей Владимирович — канд. техн. наук, доцент, кафедра геоинформационных систем факультета информатики и робототехники, Уфимский государственный авиационный технический университет (УГАТУ); старший научный сотрудник Геофизического центра РАН. Область научных интересов: геоинформационные технологии, цифровая обработка сигналов. Число научных публикаций — 147. geomagnet@list.ru; К.Маркса, 12, 450008, Уфа, Российская Федерация; р.т.: +7(917)345-2299.

Воробьева Гульнара Равилевна — канд. техн. наук, доцент, кафедра вычислительной математики и кибернетики факультета информатики и робототехники, Уфимский государственный авиационный технический университет (УГАТУ). Область научных

интересов: геоинформационные и веб-технологии, системы хранения и обработки информации. Число научных публикаций — 121. gulnara.vorobeva@gmail.com; К. Маркса, 12, 450008, Уфа, Российская Федерация; р.т.: +7(917)417-4111.

Поддержка исследований. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-00011а, Экспертного центра «Проектный офис развития Арктики» (Договор № 217-Г от 13 января 2021 г.), Министерства науки и высшего образования РФ в рамках выполнения работ по Государственному заданию ФГБОУ ВО УГАТУ # FEUE-2020-0007 по теме «Теоретические основы моделирования и семантического анализа процессов преобразования вихревых электромагнитных полей в инфокоммуникационных системах».

A. Vorobev, G. Vorobeva
**MODEL OF INFORMATION INTERACTION BETWEEN
ELEMENTS OF MULTILEVEL SYSTEM OF DIGITAL TWINS**

Vorobev A., Vorobeva G. Model of Information Interaction between Elements of Multilevel System of Digital Twins.

Abstract. One of the solutions to the problem of spatio-temporal data anisotropy is the use of a multilevel system of digital twins based on the corresponding industry models and the updated archive data base. The application of this approach has successfully proved itself in information systems for monitoring the parameters of the geomagnetic field and its variations, providing spatio-temporal interpolation of geomagnetic data with an accuracy of 0.81 nT in magnetically quiet periods. At the same time, the problem of information interaction between the levels of the system of digital twins remained unresolved, which is greatly aggravated by the constantly growing volume of data and their heterogeneous nature. The paper proposes a solution to the indicated problem by means of a formalized mechanism for packaging space-time information, in which the identification of data sources is performed on the basis of a hierarchical binary tokenization system. In addition, the proposed software implementation of such an approach is considered, a distinctive feature of which is the combination of traditional client-server and innovative serverless architectures to implement a highly loaded reactive web application for working with analyzed data. The main stages of the implementation of information interaction are highlighted and programmatically formalized - from obtaining initial information from its sources to verifying data, analyzing them, processing and forming the output information flow of the system. The results of the computational experiments carried out on the example of the problem of monitoring the parameters of the Earth's magnetic field and its variations confirmed the effectiveness of the proposed solutions, expressed both in increasing the reactivity of web-based applications and in increasing the computational speed of formation and filling of information storages that aggregate information from distributed heterogeneous sources.

Keywords: Digital Twins, Spatial Data, Information Interaction, Unified Information Space, Data Processing, Tokenization, Geographic Information Systems and Technologies

Vorobev Andrei — Ph.D., Associate Professor, Geoinformation Systems Department of Computer Science and Robotics Faculty, Ufa State Aviation Technical University (USATU); Senior Researcher, Geophysical Center of RAS. Research interests: geoinformation technologies, digital signal processing. The number of publications — 147. geomagnet@list.ru; 12, K. Marx, 450008, Ufa, Russian Federation; office phone: +7(917)345-2299.

Vorobeva Gulnara — Ph.D., Associate Professor, Computational Mathematics and Cybernetics Department of Computer Science and Robotics Faculty, Ufa State Aviation Technical University (USATU). Research interests: geoinformation and web technologies, systems of information storing and processing. The number of publications — 121. gulnara.vorobeva@gmail.com; 12, K. Marx, 450008, Ufa, Russian Federation; office phone: +7(917)417-4111.

Acknowledgements. The reported study was funded by RFBR, project number 20-07-00011a, and the Expert Center "Project Office for the Development of the Arctic" (Agreement No. 217-G dated January 13, 2021), and the Ministry of Education and Science of the Russian Federation as part of the basic part of the state assignment to higher education educational institutions #FEUE-2020-0007.

References

1. Vorobev A.V., Vorobeva G.R. [An approach to assessing the relative information efficiency of magnetic observatories in the INTERMAGNET network]. *Geomagnetizm i aeronomiya – Geomagnetism and Aeronomy*. 2018. vol. 58. no. 5. pp. 648–652. (In Russ.).
2. Gvishiani A.D., Lukyanova R.Yu., Soloviev A.A. [Geomagnetism: from the core of the Earth to the Sun]. Moscow: RAS, 2019. 186 p. (In Russ.).
3. Gvishiani A.D., Agayan S.M., Bogoutdinov Sh.R., Kagan A.I. [Gravitational smoothing of time series]. *Trudy Instituta matematiki i mekhaniki UrO RAN – Proceedings of the Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences*. 2011. vol.17. no. 2. pp. 62–70. (In Russ.).
4. Ryabova S.A. [Geomagnetic variations and variations in groundwater level and microseismic background in conditions of middle latitudes synchronous with them]: diss. for the degree of candidate of physical and mathematical sciences. Moscow, 2018 (In Russ.).
5. Kondrashov D., Shprits, Y., Ghil, M. Gap filling of solar wind data by singular spectrum analysis. *Geophys. Res. Lett.* 2010. No. 37. P. L15101 doi:10.1029/2010GL044138. (In Russ.).
6. Abbasova T.S. [Development of virtual instruments for creating digital twins] // *Informatsionno-tehnologicheskii vestnik – Information and technological bulletin*. 2019. no. 2(20). pp. 79–88. (In Russ.).
7. Chen Y. New Approaches for Calculating Moran’s Index of Spatial Autocorrelation. *PloS one*. 2013. no. 8. P. e68336. doi: 10.1371/journal.pone.0068336.
8. Freni P., Ferro E., Moncada R. Tokenization and Blockchain Tokens Classification: a morphological framework. *IEEE Symposium on Computers and Communications (ISCC)*, Rennes, France, 2020, pp. 1-6, doi: 10.1109/ISCC50000.2020.9219709.
9. Marchewka-Bartkowiak K., Nowak K. Get Tokenized... The Specificity of Personal Tokens in the Context of Tokenization and Axiological Categorization. *Proceedings of the 3rd International Conference on Economics and Social Sciences*. 2020. P. 823-831. doi:10.2478/9788395815072-081.
10. Love J. Intermagnet and the Global Community of Magnetic Observatories. *AGU Fall Meeting Abstracts*. 2006.
11. Stocchero M. Iterative deflation algorithm, eigenvalue equations, and PLS2. *Journal of Chemometrics*. 2019. No.33. P.10.1002/cem.3144.
12. Le H. A. Trigger2B: A Tool Generating Event-B Models from Database Triggers. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. 2021. vol 343. p.1-7. https://doi.org/10.1007/978-3-030-67101-3_8
13. Lenz M. Silent-Cron: A Cron Wrapper. *Raku Fundamentals*. 2020. pp. 59-80. doi:10.1007/978-1-4842-6109-5_6.
14. Voss N., Becker T., Mencer O., Gaydadjiev G. Rapid Development of Gzip with MaxJ. *Lecture Notes in Computer Science*. 2017. vol. 10216. p. 60-71. doi:10.1007/978-3-319-56258-2_6.
15. Mittal H. Digital twin: An overview. *CSI Communications*. 2020. no. 44.
16. Carver J., Penzenstadler B., Scheuner J., Staron M. (Research) Insights for Serverless Application Engineering. *IEEE Software*. 2021. vol. 38. pp.123-125. doi:10.1109/MS.2020.3028659.
17. Mahmoudi N., Khazaei H. Performance Modeling of Serverless Computing Platforms. *IEEE Transactions on Cloud Computing*. 2020. P. 1-10. doi:10.1109/TCC.2020.3033373.
18. Martins H., Araujo F., Cunha P. R. Benchmarking Serverless Computing Platforms. *Journal of Grid Computing*. 2020. vol.18. doi:10.1007/s10723-020-09523-1.

19. Ali S., Alauldeen R., Ruaa A. What is Client-Server System: Architecture, Issues and Challenge of Client-Server System (Review). *HBRP Publication*. 2020. P.1-6. doi: 10.5281/zenodo.3673071.
20. Zhu Y., Wu W., Li D. Efficient Client Assignment for Client/Server Systems. *IEEE Transactions on Network and Service Management*. 2016. no.13. pp.1-10. doi:10.1109/TNSM.2016.2597269.
21. Iskandar T., Lubis M., Kusumasari T., Lubis A. Comparison between client-side and server-side rendering in the web development. *IOP Conference Series Materials Science and Engineering*. 2020. no.801. p.012136. doi:10.1088/1757-899X/801/1/012136.
22. Zeng Z., Zhang H. A Study on Cache Strategy of CDN Stream Media. *IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Chongqing, China, 2020. pp. 1424-1429. doi: 10.1109/ITAIC49862.2020.9338805.
23. Sharma K., Marjit U., Biswas U. Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet. *Information Technology and Libraries*. 2017. no.37. pp. 29-34. doi:10.6017/ital.v37i3.10177.
24. Kostinskiy S., Shaikhutdinov D., Narakidze N. Loss Counter in Power Double Winding Transformers Implementing the Method of Conditionally Constant Coefficients in Online Mode Using the Information Platform. *Izvestiya vysshikh uchebnykh zavedeniy. Elektromekhanika - Proceedings of higher educational institutions. Electromechanics*. 2020. no.63. pp.79-85. doi:10.17213/0136-3360-2020-5-79-85. (In Russ.).
25. Vorobev A.V., Pilipenko V.A., Enikeev T.A., Vorobeva G.R. Geoinformation system for analyzing the dynamics of extreme geomagnetic disturbances from observations of ground stations. *Computer Optics*. 2020. no.44(5). pp.782-790. doi: 10.18287/2412-6179-CO-707.
26. Vorobev A.V., Vorobeva G.R., Yusupova N.I. [The concept of a single space of geomagnetic data]. *Trudy SPIIRAN - SPIIRAS Proceedings*. 2019. no. 18, vol. 2. pp.390-415. (In Russ.).