

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

УДК 004.422

DOI 10.21685/2072-3059–2016-4-1

П. П. Макарычев, Л. В. Гурьянов, Э. В. Цукарев

ВЕРИФИКАЦИЯ И ВАЛИДАЦИЯ ДАННЫХ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ МОНИТОРИНГА И УПРАВЛЕНИЯ ЭНЕРГОПОТРЕБЛЕНИЕМ

Аннотация.

Актуальность и цели. Контроль, учет и оперативное управление различными энергетическими потоками становятся все более актуальными и важными в связи с постоянным увеличением стоимости энергоресурсов и с необходимостью снижения себестоимости выпускаемой продукции. Получение и использование достоверной оперативной информации о потреблении энергоресурсов – необходимое условие эффективного решения задач энергосбережения и энергоэффективности. Применение комплексных автоматизированных систем сбора, мониторинга и управления потреблением энергоресурсов позволяет существенно сократить затраты и минимизировать ошибки при сборе данных с приборов учета ресурсов. Цель данного исследования – разработка алгоритмов верификации и валидации данных об энергопотреблении, обеспечивающих достоверность первичных данных.

Результаты. Предложены алгоритмы верификации и валидации, обнаруживающие недостоверные значения данных на основе анализа скорости изменения, адекватности и соотношения значений параметров, а также осуществляющие замену недостоверного значения на достоверное. Исследованы временные характеристики первичной обработки данных на основе предложенных алгоритмов в автоматизированной системе контроля и управления энергопотреблением.

Выводы. Полученные результаты исследований позволяют утверждать, что использование алгоритмов верификации и валидации обеспечивает решение задач получения достоверной оперативной информации о потреблении энергоресурсов.

Ключевые слова: верификация, валидация, сервер сбора и консолидации данных, энергосбережение, энергоэффективность, комплексная автоматизированная система.

P. P. Makarychev, L. V. Gur'yanov, E. V. Tsukarev

DATA VERIFICATION AND VALIDATION IN THE AUTOMATED SYSTEM OF ENERGY CONSUMPTION MONITORING AND MANAGEMENT

Abstract.

Background. Monitoring, accounting and operational management of various energy flows is becoming more relevant and important due to the constant increase in energy costs and the need to reduce production costs. Obtaining and using accurate real-time information on energy consumption is a necessary condition for effective settling of energy saving and energy efficiency problems. The use of integrated automated systems of energy consumption monitoring, management and data collection can significantly reduce costs and minimize errors in resource data counters. The purpose of this work is to develop algorithms for energy consumption data verification and validation, providing primary data credibility.

Results. The have proposed verification and validation algorithms that detect false data values based on the analysis of the rate of changes, adequacy and ratios of parameters and replace false values with true ones. The researchers investigated temporal characteristics of primary data processing on the basis of the proposed algorithms in the automated system of energy consumption control and management.

Conclusions. The obtained results allow to assert that the use of verification and validation algorithms provides solutions to the problems of obtaining accurate real-time information on energy consumption.

Key words: verification, validation, collection server and data consolidation, energy efficiency, integrated automated system.

Введение

Одним из эффективных решений задач энергосбережения и энергоэффективности в промышленности, социальной сфере и жилищно-коммунального хозяйства является использование комплексных автоматизированных систем сбора и анализа данных об энергопотреблении [1].

Использование консолидации данных в комплексных системах позволяет собирать данные из нескольких первичных источников и интегрировать их в одном хранилище. Такое хранилище может быть использовано для подготовки отчетности и проведения анализа энергопотребления или как источник данных для других приложений. Преимуществом консолидации данных является возможность верификации и трансформации значительных объемов данных в процессе их передачи от источника к хранилищу.

Особенно важной является верификация данных, полученных непосредственно с приборов учета энергоресурсов. Например, технический персонал неправильно подключил счетчик учета. В результате этого значение оперативного тега, ответственного за измеряемое значение параметра энергоресурса, будет неопределенно (в автоматизированной системе сбора данных оно равно NaN – Not a Number). Использование такого значения в процессах обработки данных, например в биллинге, приведет к ошибкам и сбоям при формировании счетов, отчетов и вычислении обобщенных показателей энергопотребления.

Таким образом, верификация консолидируемых данных в комплексных системах контроля и управления энергопотреблением является важной и актуальной. Верификация первичных данных должна выполняться всегда, независимо от способа получения данных: по расписанию (заданному интервалу опроса приборов учета), инициативному опросу или через web-сервис [2, 3].

Исследование возможных путей решения задач первичной обработки данных об энергопотреблении привело к выводу об использовании двух этапов: верификации и валидации данных.

ГОСТ Р ИСО 9000–2008 определяет эти термины следующим образом: «Верификация – подтверждение на основе представления объективных свидетельств того, что установленные требования были выполнены. Валидация – подтверждение на основе представления объективных свидетельств того, что требования, предназначенные для конкретного использования или применения, выполнены». В рамках комплексной автоматизированной системы контроля и управления энергопотреблением эти термины можно интерпретировать следующим образом:

- **верификация** – проверка полученного значения технологического тега на достоверность;

- **валидация** – замена полученного недостоверного значения тега на достоверное значение.

Для установки признака достоверности V значения технологического тега можно использовать различные функции $F_i(\cdot)$, $i = 1, 2, \dots$. Все эти функции должны удовлетворять требованию: принимать значение 1 при выполнении условий достоверности, 0 – в случае невыполнения условий. Выбор вида функции зависит от параметра потребляемого ресурса, включая электроэнергию, холодную и горячую воду и т.д.

1. Алгоритмы верификации

1.1. Алгоритм верификации по скорости изменения параметра

Алгоритм предназначен для определения достоверности текущего значения параметра энергопотребления по скорости изменения этого значения. Для расчета допустимой скорости изменения значения параметра задается временной интервал, а также величина допустимого изменения значения за заданный интервал в единицах изменения параметра. Функция имеет вид

$$F_1(Z) = \begin{cases} 1, & \text{если } |(Z_k - Z_{k-1}) / \Delta t| \leq Z_d / \Delta t; \\ 0 & \text{иначе,} \end{cases}$$

где Z_k – текущее значение параметра; Z_{k-1} – предыдущее значение параметра; Δt – разность меток времени между текущим и предыдущим значениями параметра; Z_d – допустимое изменение параметра за заданный интервал; Δt_d – заданный интервал времени для допустимого изменения параметра.

Пример. Инженер автоматизированной системы управления технологическим процессом определил допустимое изменение значения измеряемого параметра энергопотребления в 100 единиц на временном промежутке в 1 мин. При поступлении нового значения в данном алгоритме вычисляется разность между текущим (новым) и предыдущим значениями, а также разность между их метками времени. Предположим, что интервал между метками времени двух последовательных значений составляет 1 мин, а разность между значениями параметра – 200 единиц. Следовательно, текущее значение будет признано недостоверным, так как разность превышает максимальное допустимое значение изменения в 100 единиц. Если разность между метками времени текущего и предыдущего значений составляет 4 мин, а разность между значениями – 200 единиц, то текущее значение параметра будет признано досто-

верным, так как приращение значения в минуту составляет 50 единиц, которое не превышает максимальное допустимое значение.

1.2. Алгоритм верификации по связанному параметру

Алгоритм позволяет сравнивать значение измеренного параметра энергопотребления со сторонним параметром. В этом случае функция может быть задана следующим образом:

$$F_2(Z) = \begin{cases} 1, & \text{если } ((Z_k \geq Z_{k-1}) \wedge (((P_k \geq P_{k-1}) \wedge (S_t = 0)) \vee \\ & \vee ((P_k \leq P_{k-1}) \wedge (S_t = 1)))) \vee ((Z_k \leq Z_{k-1}) \wedge (((P_k \leq P_{k-1}) \wedge \\ & \wedge (S_t = 0)) \vee ((P_k \geq P_{k-1}) \wedge (S_t = 1))))); \\ 0 & \text{иначе,} \end{cases}$$

где P_k – текущее значение связанного параметра; P_{k-1} – предыдущее значение связанного параметра; S_t – признак связанности (может иметь значения 0 или 1). Если значение S_t равно нулю, применяется правило прямо пропорциональной связанности. При значении S_t равном единице, применяется обратно пропорциональная связанность.

Пример 1. Расход воды прямо пропорционально зависит от давления в трубопроводе. Имеется датчик, измеряющий давление воды. Измеренное значение фиксируется в технологическом параметре «Тег1». Значение другого датчика, отвечающего за расход воды в трубопроводе, присваивается параметру «Тег2». При анализе изменения значений этих параметров применяется правило прямо пропорциональной связанности.

Текущее значение параметра «Тег1» оценивается как достоверное значение в следующих случаях:

– если значение «Тег1» больше предыдущего значения «Тег1» или равно ему и текущее значение «Тег2» также больше или равно предыдущему значению «Тег2»;

– если текущее значение «Тег1» меньше или равно его предыдущему значению и текущее значение «Тег2» также меньше или равно предыдущему значению «Тег2».

В противном случае текущее значение «Тег1» считается недостоверным.

Пример 2. На производстве имеется две емкости накопительного типа. В ходе технологического процесса жидкость полностью переливается из первой емкости во вторую. На каждой из емкостей установлен датчик, измеряющий объем жидкости (значение объема, полученное датчиком на первой емкости – «Тег1», значение объема, полученное датчиком на второй емкости – «Тег2»). На временном интервале, когда происходит переход жидкости из первой емкости во вторую, оперативные значения «Тег1» будут уменьшаться, а оперативные значения «Тег2» – увеличивающимися.

При первичной обработке значений «Тег1» и «Тег2» по алгоритму связанных параметров применяется обратно пропорциональная связанность. В этом случае значения «Тег2» будут признаваться достоверными, если текущее значение «Тег2» больше его предыдущего значения, а текущее значение «Тег1» меньше его предыдущего значения.

Если во время переливания жидкости происходит утечка (жидкость не попадает во вторую емкость), оперативные значения «Тег2» не будут увеличиваться, и алгоритм признает оперативные значения «Тег2» недостоверными.

Пример 3. Для определения достоверности значения параметра технологического процесса используются связанные значения, хранимые в базе данных (БД) (значения в БД могут формироваться сторонней или дублирующей системой). «Тег1» – значение параметра первичной обработки данных, «Тег2» – значение, определенное в БД. Для каждого полученного значения «Тег1» алгоритм осуществляет доступ к требуемой БД и считывает значение «Тег2». Далее с учетом типа связанности происходит сравнение текущих значений этих параметров с их предыдущими значениями, и как результат устанавливается достоверность значения «Тег1».

1.3. Алгоритм верификации по интегральному значению

Алгоритм предназначен для определения достоверности текущего значения параметра интегрального типа (например, объем потребленной воды, мощность потребленной электрической энергии). Функция оценки признака достоверности имеет вид

$$F_3(Z) = \begin{cases} 1, & Z_k \geq Z_{k-1}; \\ 0 & \text{иначе.} \end{cases} \quad (1)$$

Из (1) следует, что текущее значение считается достоверным, если оно равно или больше предыдущего значения, в противном случае значение признается недостоверным. Данный алгоритм не имеет настроечных параметров, так как в расчете достоверности используются только текущее и предыдущее значения параметра.

1.4. Алгоритм верификации по заданному интервалу

Алгоритм предназначен для определения достоверности значения текущего значения параметра по его вхождению в заданный интервал. Текущее значение считается достоверным, если оно попадает в заданный интервал, в противном случае – недостоверным:

$$F_4(Z) = \begin{cases} 1, & \text{если } Z_{\min} \leq Z_k \leq Z_{\max}; \\ 0 & \text{иначе,} \end{cases}$$

где Z_{\min} – нижняя граница интервала; Z_{\max} – верхняя граница интервала.

2. Алгоритмы валидации

2.1. Алгоритм валидации по среднему значению на интервале

Алгоритм валидации по среднему значению на интервале производит замену недостоверного значения точки на среднее значение за период времени, заданный относительно последнего значения, используемого в процедуре валидации. На первом этапе вычисляем сумму текущих значений $S_1^*(Z)$ достоверных точек по следующей формуле:

$$S_1^*(Z) = \sum_{k=1}^n Z_k (V = 1).$$

На втором этапе производится вычисление среднего значения на интервале $S_1(Z)$:

$$S_1(Z) = S_1^*(Z) / m,$$

где m – количество достоверных точек.

2.2. Алгоритм валидации по последнему достоверному значению

Алгоритм валидации по последнему достоверному значению производит замену недостоверного значения параметра на последнее достоверное, которое сохраняется в контексте анализа.

Возможна ситуация, когда среди поступивших значений параметра нет достоверных значений. В этом случае данный алгоритм не произведет замену недостоверных значений и сформирует предупреждающее сообщение в протокол событий. Данная ситуация наиболее вероятна при первом старте автоматизированной системы, когда отсутствуют реальные достоверные данные.

Данный алгоритм не имеет настроечных параметров.

2.3. Алгоритм по пользовательскому значению

Алгоритм валидации по пользовательскому значению производит замену недостоверного значения параметра на значение, которое определено в настройках алгоритма.

Значение, определенное пользователем в настройках, может задаваться как для группы однотипных параметров, так и для каждого параметра индивидуально.

2.4. Алгоритм «Ручной ввод»

Алгоритм валидации «Ручной ввод» применяется для отложенной замены значений параметров энергопотребления, записанных в БД.

3. Компонент первичной обработки данных сервера консолидации данных

Комплексная автоматизированная система контроля и управления энергопотреблением строится по иерархическому принципу и включает уровни приборов учета, измерительных и исполнительных механизмов; сбора и консолидации данных; визуализации, мониторинга и управления.

Архитектура такой автоматизированной системы на примере измерительно-вычислительного комплекса (ИВК) «ЭнергоКруг» приведена на рис. 1.

Базовые компоненты ИВК – это контроллеры сбора данных DevLink, серверы сбора, предварительной обработки и консолидации данных WideTrack, серверы визуализации и обработки данных DataRate. Эти компоненты позволяют гибко организовать уровни сбора, консолидации, хранения и визуализации данных, а также обеспечить предоставление результатов измерений и анализа данных энергопотребления персоналу и руководству предприятия в необходимой и удобной форме.

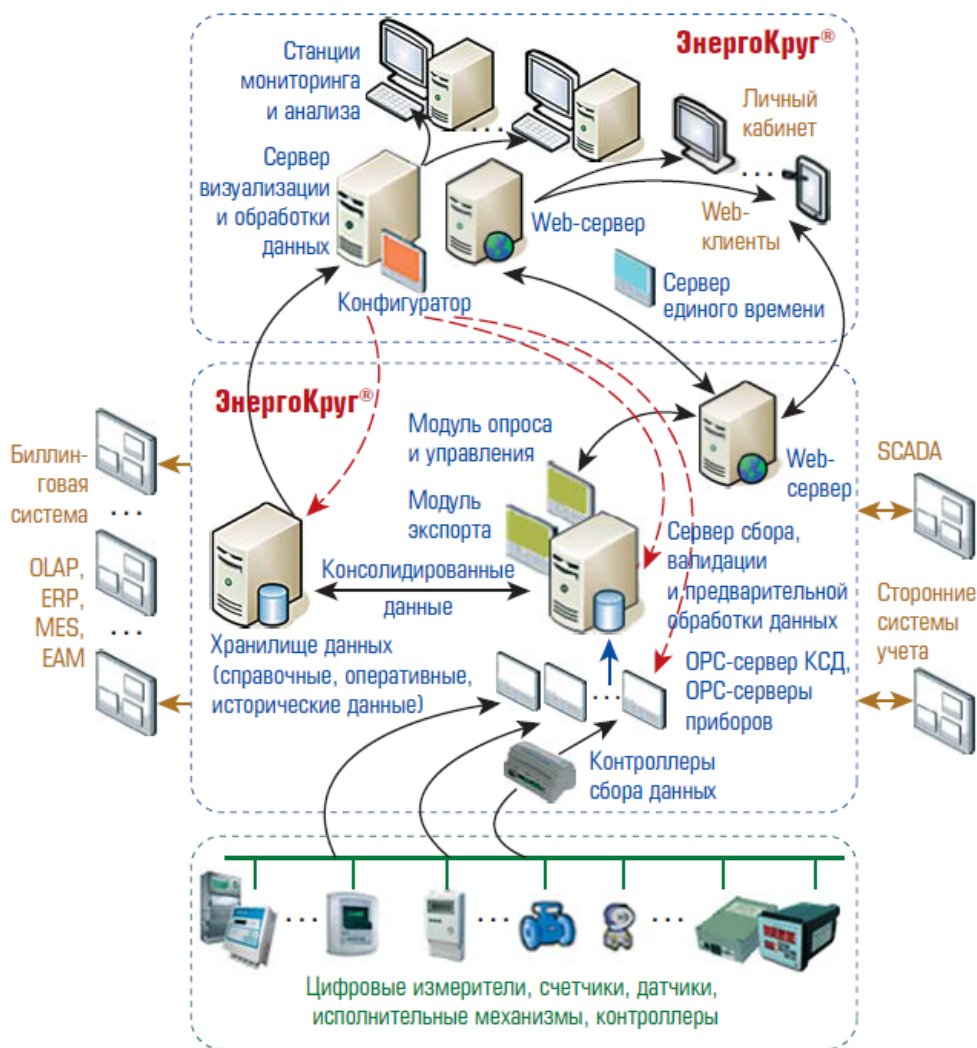


Рис. 1. Архитектура ИБК «ЭнергоКруг»

Выполнять первичную обработку данных на нижних уровнях ИБК (контроллер, OPC Сервер) проблематично, так как сложно настраивать алгоритмы верификации и валидации из программного обеспечения верхнего уровня.

Лучше и проще всего сделать первичную обработку данных в виде отдельного компонента на уровне Сервера сбора, предварительной обработки и консолидации данных (рис. 2).

Компонент первичной обработки данных может включать в себя один или несколько алгоритмов верификации и валидации данных. Структура компонента легко расширяема для добавления новых алгоритмов верификации и валидации.

Конфигуратор Сервера сбора, предварительной обработки и консолидации данных позволяет выполнять как групповую настройку алгоритмов, так и независимое конфигурирование для каждого параметра (тега). Конфи-

гурация алгоритмов сохраняется в БД. Это позволяет редактировать настройки алгоритмов не только конфигуратором Сервера сбора, предварительной обработки и консолидации данных, но и сторонним программным средством.

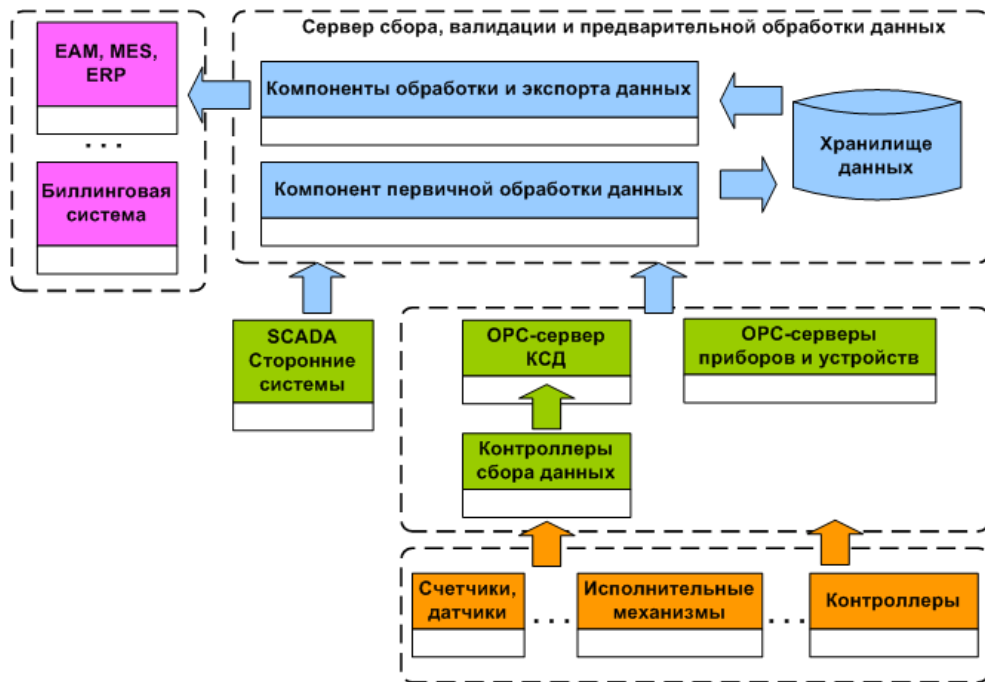


Рис. 2. Компонент первичной обработки данных в структуре Сервера сбора данных

4. Исследование алгоритмов верификации и валидации

Для получения сравнительных характеристик работы компонента первичной обработки данных был использован типовый проект автоматизированной системы контроля и управления энергопотреблением, созданной на основе ИВК «ЭнергоКруг». В проекте был подключен компонент первичной обработки, в котором для верификации использовался алгоритм по скорости изменения параметра, для валидации – алгоритм по среднему значению на интервале.

Количество технологических тегов в типовом проекте изменялось от 1 000 до 10 000. Исследование первичной обработки данных осуществлялось для значений десятиминутных тегов. За цикл опроса считывается шесть значений для каждого тега, итого считывается от 6 000 до 60 000 значений (в зависимости от количества тегов в проекте) в час для всех тегов.

После первичной обработки данные сохранялись в БД. В процессе тестовых прогонов проекта измерялось время, прошедшее от момента старта проекта до момента времени, когда все данные будут записаны в БД.

Данные исследования сведены в табл. 1 и 2. Результаты табл. 1 представлены в графическом виде на рис. 3.

Увеличение времени обработки данных при верификации и валидации – это неизбежные затраты времени, учитываемые при консолидации данных в комплексной автоматизированной системе.

Таблица 1

Время первичной обработки данных

Количество тегов в проекте	Количество значений	Время обработки с верификацией и валидацией, с	Время обработки только с верификацией, с	Время обработки без использования верификации и валидации, с
1000	6000	205	180	150
2000	12000	256	193	162
3000	18000	283	216	178
4000	24000	324	241	192
5000	30000	386	267	208
6000	36000	425	291	224
7000	42000	463	330	241
8000	48000	491	360	276
9000	54000	512	410	297
10000	60000	521	433	310

Таблица 2

Количество текущих значений тегов, требующих верификации, полученных за один цикл опроса

Количество тегов в проекте	Количество значений	Количество недостоверных тегов в проекте после верификации и валидации	Количество недостоверных тегов в проекте без верификации и валидации
1000	6000	0	5
2000	12000	0	8
3000	18000	0	10
4000	24000	0	14
5000	30000	0	18
6000	36000	0	23
7000	42000	0	29
8000	48000	0	33
9000	54000	0	38
10000	60000	0	47

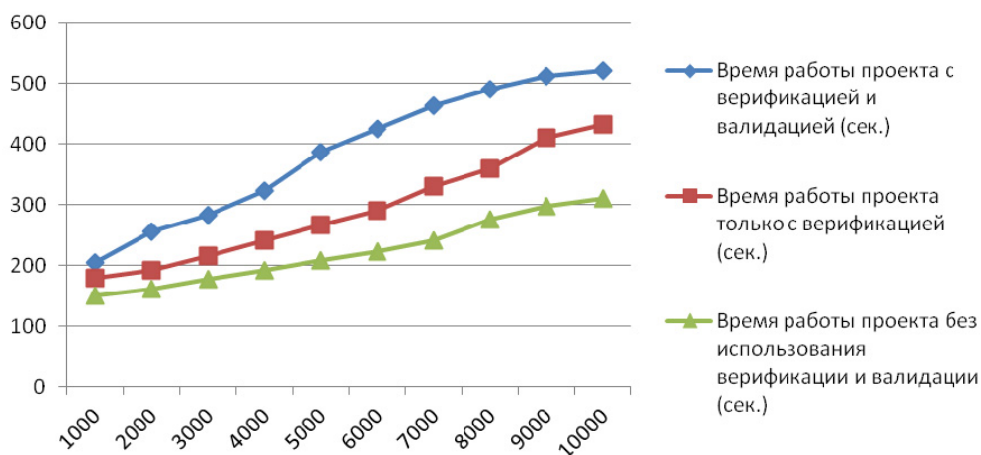


Рис. 3. Сравнительное время первичной обработки данных

Однако, как показывают данные исследования, увеличение времени сбора и обработки данных в проекте с верификацией и валидацией существенно не влияет на производительность комплекса технических средств, на которых эксплуатируется Сервер сбора, предварительной обработки и консолидации данных, так как максимально допустимое время работы одного цикла опроса типового проекта не должно превышать 30 мин. Ни один из приведенных замеров не превысил это значение.

При прогоне проекта также измерялось количество тегов (табл. 2), требующих верификации (значениями этих тегов были NaN, Null, плюс бесконечность или минус бесконечность). Следует отметить, что измерения количества тегов, требующих верификации, производились непосредственно перед записью значений тегов в БД.

Приведенные данные в табл. 2 показывают, что использование алгоритмов верификации и валидации полностью устраняет некорректные текущие значения тегов.

Заключение

Использование алгоритмов верификации и валидации данных обеспечивает решение важных задач энергосбережения и энергоэффективности, которые связаны с получением и использованием достоверной оперативной информации о потреблении энергоресурсов, текущем состоянии промышленных, социальных объектов и объектов жилищно-коммунального хозяйства, инженерных коммуникаций.

Обнаружение недостоверных данных при сборе и консолидации помогает выявлять нарушения режимов и факты нерегламентированного использования энергоресурсов, предаварийные и аварийные ситуации при эксплуатации оборудования. Это позволяет современным предприятиям и организациям эффективно осуществлять свою деятельность в условиях жесткой экономики денежных средств, материальных ресурсов и оптимизации всех производственных процессов в масштабе района, города, региона.

Список литературы

1. **Ицкович, Э. Л.** Основные положения рационального построения систем автоматизации энергообъектов разных классов / Э. Л. Ицкович // Автоматизация и IT в энергетике. – 2013. – № 7. – С. 2–9.
2. **Цукарев, Э. В.** Реализация инициативного опроса на Сервере консолидации технологических данных / Э. В. Цукарев // Открытые инновации – вклад молодежи в развитие региона : сб. материалов регионального молодежного форума : в 2 т. – Пенза : Изд-во ПГУ, 2013. – Т. 1. – С. 207–208.
3. **Цукарев, Э. В.** Разработка Сервера-шлюза получения технологических данных в гетерогенной сети / Э. В. Цукарев // Современные проблемы компьютерных наук (СПКН-2013) : сб. материалов I Междунар. науч.-практ. конф. – Пенза : Изд-во ПГУ, 2013. – С. 61–63.
4. **Лукин, Д. В.** Адаптация систем сбора данных к входным перегрузкам / Д. В. Лукин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2008. – № 2 (6). – С. 47–55.

References

1. Itskovich E. L. *Avtomatizatsiya i IT v energetike* [Automation and IT in the energy sector]. 2013, no. 7, pp. 2–9.

2. Tsukarev E. V. *Otkrytye innovatsii – vklad molodezhi v razvitie regiona: sb. materialov regional'nogo molodezhnogo foruma: v 2 t.* [Open innovations – youth's contribution into regional development: proceedings of the Regional youth forum]. Penza: Izd-vo PGU, 2013, vol. 1, pp. 207–208.
3. Tsukarev E. V. *Sovremennye problemy komp'yuternykh nauk (SPKN-2013): sb. materialov I Mezhdunar. nauch.-prakt. konf.* [Modern problems of computer sciences (SPKN-2013): proceedings of I International scientific and practical conference]. Penza: Izd-vo PGU, 2013, pp. 61–63.
4. Lukin D. V. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki* [University proceedings. Volga region. Engineering sciences]. 2008, no. 2 (6), pp. 47–55.

Макарычев Петр Петрович

доктор технических наук, профессор,
заведующий кафедрой математического
обеспечения и применения ЭВМ,
Пензенский государственный
университет (Россия, г. Пенза,
ул. Красная, 40)

E-mail: makpp@yandex.ru

Makarychev Petr Petrovich

Doctor of engineering sciences, professor,
head of sub-department of software
and computer application, Penza State
University (40 Krasnaya street,
Penza, Russia)

Гурьянов Лев Вячеславович

кандидат технических наук, доцент,
кафедра математического обеспечения
и применения ЭВМ, Пензенский
государственный университет (Россия,
г. Пенза, ул. Красная, 40)

E-mail: leo8087@yandex.ru

Gur'yanov Lev Vyacheslavovich

Candidate of engineering sciences, associate
professor, sub-department of software
and computer application, Penza
State University (40 Krasnaya
street, Penza, Russia)

Цукарев Эдуард Владимирович

аспирант, Пензенский государственный
университет (Россия,
г. Пенза, ул. Красная, 40)

E-mail: tsukarev.ed@gmail.ru

Tsukarev Eduard Vladimirovich

Postgraduate student, Penza State
University (40 Krasnaya street,
Penza, Russia)

УДК 004.422

Макарычев, П. П.

Верификация и валидация данных в автоматизированной системе мониторинга и управления энергопотреблением / П. П. Макарычев, Л. В. Гурьянов, Э. В. Цукарев // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2016. – № 4 (40). – С. 5–15. DOI 10.21685/2072-3059-2016-4-1