Архитектура защиты критических информационных инфраструктур на основе гибридного анализа данных и алгоритмов ИИ

Е. С. Митяков[1], А. И. Ладынин[2]

МИРЭА – Российский технологический университет

[1]iyao@mail.ru, [2]andrey.ladynin@hotmail.com

И. Д. Казакевич

Нижегородский государственный технический университет им. Р.Е. Алексеева;
МИРЭА – Российский технологический университет

kazakevich.igor2000@gmail.com

***Аннотация*. В статье рассматривается архитектура защиты критических информационных инфраструктур (КИИ) с использованием методов искусственного интеллекта (ИИ) и гибридного анализа данных. Предложенный подход сочетает спектральный и фрактальный анализ для выявления и нейтрализации киберугроз в реальном времени. Рассматриваются алгоритмы Isolation Forest, LOF, One-Class SVM и кластеризация K-средних для обнаружения аномалий в сетевом трафике и технологических процессах. Моделирование кибератак в энергосистеме демонстрирует эффективность предложенной архитектуры в обеспечении устойчивости КИИ к современным угрозам.**

Ключевые слова: критическая информационная инфраструктура, киберугрозы, искусственный интеллект, обнаружение аномалий, вейвлет-преобразование, спектральный анализ, машинное обучение, энергетическая безопасность.

I. Введение

Критическая информационная инфраструктура (КИИ) — совокупность информационных систем и сетей, нарушение работы которых может привести к значительным негативным последствиям для национальной безопасности, экономики, здравоохранения и общества. Современные угрозы, такие как кибератаки, программные сбои и целевые аномалии, требуют комплексного подхода к защите. В статье предлагается архитектура защиты критических информационных инфраструктур, объединяющая методы искусственного интеллекта (ИИ), спектральный и фрактальный анализ для обнаружения и нейтрализации угроз в режиме реального времени [1].

Использование искусственного интеллекта в таких системах должно соответствовать строгим критериям: скорость, точность, адаптивность. Алгоритмы должны обеспечивать анализ данных в реальном времени, быструю изоляцию аномалий, обработку мультимодальных данных (сетевой трафик, журналы событий, параметры работы оборудования), устойчивость к шумам и флуктуациям данных.

Для фильтрации случайных помех эффективным инструментом является вейвлет-преобразование, позволяющее выделить значимые сигналы из шума. Развитие методов анализа и защиты данных играет ключевую роль в обеспечении устойчивости критических информационных инфраструктур к современным угрозам. Формально вейвлет-преобразование W_x сигнала $x(t)$ на уровне j можно можно описать следующим выражением:

$$W_x(t,j) = \sum_t x(t)\psi_{j,k}(t)\ , \qquad (1)$$

где $\psi_{j,k}(t)$ — функция-вейвлет характеризующая уровень детализации j, сдвинутая и масштабированная версия материнского вейвлета.

Для каждого уровня разложения j получаем набор коэффициентов c_j, которые описывают различные временные масштабы сигнала:

$$c_j = W_x(t,j). \qquad (2)$$

На каждом уровне j вейвлет-разложения вычисляются мультифрактальные признаки, такие как среднее значение и дисперсия абсолютных значений коэффициентов c_j:

$$\mu_j = \frac{1}{N_j}\sum_{k=1}^{N_j}\left|c_{j,k}\right|, \qquad (3)$$

$$\sigma_j^2 = \frac{1}{N_j}\sum_{k=1}^{N_j}\left(\left|c_{j,k}\right| - \mu_j\right)^2, \qquad (4)$$

где N_j — количество коэффициентов на уровне j.

Эти признаки составляют вектор признаков для каждого временного ряда:

$$\text{features} = [\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \dots, \mu_m, \sigma_m^2]. \qquad (5)$$

Пусть F_i — вектор мультифрактальных признаков для i-го временного ряда, тогда множество признаков для всех временных рядов можно записать как матрицу:

$$\mathrm{F} = [\mathrm{F}_1, \mathrm{F}_2, \dots, \mathrm{F}_n]^T. \qquad (6)$$

Вейвлет-преобразование эффективно фильтрует флуктуации и выделяет значимые сигналы, анализируя динамику временных рядов на разных масштабах. Формирование вектора и матрицы признаков создаёт надёжную основу для алгоритмов ИИ, повышая точность обнаружения аномалий и позволяя оперативно реагировать на угрозы в КИИ.

II. Постановка задачи

В современных КИИ для обнаружения и классификации аномальных состояний можно задействовать методы искусственного интеллекта. Рассмотрим основные алгоритмы, используемые в интегрированной системе защиты.

1. Isolation Forest для обнаружения резких аномалий. Isolation Forest представляет собой метод ансамблевого обучения, ориентированный на выявление выбросов за счет случайного разбиения выборки.

Пусть $x \in \mathbb{R}^d$. Определим $x = (x_1, x_2, \dots, x_d)$ как наблюдение в d-мерном пространстве, которое необходимо классифицировать как нормальное или аномальное. При этом xi – i-й признак (компонента) вектора x.

Алгоритм Isolation Forest строится на идее, что аномальные точки данных x легче изолировать по сравнению с нормальными точками.

В алгоритме применяется ансамбль случайных деревьев. Каждое дерево строится путём последовательного разделения подмножеств данных:

1. На каждом шаге случайным образом выбирается один из признаков x_i и случайное пороговое значение p_i, при этом данные разделяются на две части: если $x_i \le p_i$, то x идет в левую ветвь; иначе – в правую.

2. Этот процесс продолжается до тех пор, пока каждое наблюдение не будет полностью изолировано или пока не будет достигнута максимальная глубина дерева.

Для каждого наблюдения x оценивается его аномальность на основе средней длины пути $h(x)$, необходимой для изоляции x в деревьях изоляции. Меньшая длина пути указывает на то, что точку x легко изолировать, что характерно для аномалий. Оценка аномальности рассчитывается как [2]:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}, \tag{7}$$

где $E(h(x))$ – средняя длина пути для изоляции наблюдения x по всем деревьям, а $c(n)$ – корректирующий коэффициент, зависящий от размера выборки n:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}, \tag{8}$$

где $H(n)$ – гармоническое число, которое можно аппроксимировать как $H(n) \approx \ln(n) + \gamma$, где $\gamma \approx 0.5772156649$ – постоянная Эйлера–Маскерони.

Значение $s(x,n)$ близкое к единице указывает на высокую вероятность того, что точка x является аномальной. Если $s(x,n)$ близко к нулю, то точка x скорее всего нормальна.

2. LOF (Local Outlier Factor) для локальных отклонений. Метод LOF применяется для выявления отклонений, не обнаруживаемых глобальными алгоритмами, что особенно актуально для анализа трафика в IoT-сетях. Метод позволяет выявлять локальные аномалии на основе сравнения плотности данных в окрестностях каждой точки [3]. Локальный фактор выброса для каждой точки x_i рассчитывается следующим образом:

1. Определяется расстояние до ближайших соседей:

$$d_k(\mathrm{x}_i, \mathrm{x}_j) = \| \mathrm{x}_i - \mathrm{x}_j \|, \tag{9}$$

где k — количество ближайших соседей.

2. Определяется локальная плотность для точки x_i:

$$\mathrm{lrd}_k(\mathrm{x}_i) = \left(\frac{\sum_{j=1}^{k} \mathrm{reach\_dist}_k(\mathrm{x}_i, \mathrm{x}_j)}{k} \right)^{-1}, \tag{10}$$

где $\mathrm{reach\_dist}_k(\mathrm{x}_i, \mathrm{x}_j)$ — это расстояние, на которое нужно переместиться от x_i к x_j, чтобы достичь плотности x_j.

3. Расчет LOF:

$$\mathrm{LOF}_k(\mathrm{x}_i) = \frac{\sum_{j=1}^{k} \frac{\mathrm{lrd}_k(\mathrm{x}_j)}{\mathrm{lrd}_k(\mathrm{x}_i)}}{k}, \tag{11}$$

Значение $\mathrm{LOF}_k(\mathrm{x}_i)$, значительно превышающее 1, указывает на то, что точка x_i является аномальной.

Для визуализации результатов оценки аномалий полученные значения LOF инвертируются:

$$S_i = -\mathrm{LOF}_k(\mathrm{x}_i), \tag{12}$$

где S_i — аномальная оценка для точки x_i.

3. One-Class SVM для анализа журналов событий. Метод One-Class SVM обладает рядом особенностей, которые делают его особенно подходящим для задач обнаружения аномалий в критически важных системах, таких как электрические сети [4]. В отличие от других методов, One-Class SVM направлен на обучение модели, которая описывает распределение нормальных данных, и затем используется для выявления отклонений, которые не соответствуют этому распределению. Этот подход особенно полезен в условиях, где имеются ограниченные данные об аномальных состояниях или кибератаках, и основное внимание уделяется выявлению отклонений от нормального состояния системы.

Математически, метод One-ClassSVM строит гиперплоскость в пространстве признаков, которая отделяет все точки данных от начала координат, стремясь максимизировать расстояние между этой гиперплоскостью и наиболее близкими к ней точками данных. Цель состоит в том, чтобы все нормальные данные располагались по одну сторону гиперплоскости, а аномалии — по другую.

Формально, пусть $\mathbf{x}_i$ обозначает вектор признаков временного ряда, где $i = 1,2,\dots,n$. Модель One-Class SVM решает следующую задачу оптимизации:

$$\min_{\mathrm{w},\rho,\xi_i} \frac{1}{2} \| \mathbf{w} \|^2 + \frac{1}{\nu n} \sum_{i=1}^{n} \xi_i - \rho \tag{13}$$

при условии:

$$\left(\mathrm{w} \cdot \phi(\mathrm{x}_i)\right) \ge \rho - \xi_i, \quad \xi_i \ge 0, \quad i = 1,2,\dots,n. \tag{14}$$

Здесь $\mathbf{w}$ — вектор весов, ρ — смещение гиперплоскости, ξ_i — переменные разрывов (slackvariables), $\phi(\mathbf{x}_i)$ — функция отображения в высокоразмерное пространство признаков, а ν — гиперпараметр, контролирующий допустимую долю выбросов и сложность модели.

Результатом работы One-Class SVM является функция принятия решений:

$$f(\mathrm{x}) = \left(\mathrm{w} \cdot \phi(\mathrm{x})\right) - \rho. \tag{15}$$

Значения $f(\mathrm{x}) \ge 0$ указывают на потенциальные аномалии, тогда как значения $f(\mathbf{x}) < 0$ соответствуют нормальным данным.

4. Метод кластеризации К-средних. Метод К-средних предназначен для разбиения набора данных на k

кластеров, где каждый кластер характеризуется своим центром (центроидом).

Цель метода заключается в минимизации суммы квадратов расстояний между точками данных и центрами кластеров.

Пусть у нас есть набор данных $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, где каждая точка данных x_i является вектором признаков.

Расчет состоит из следующих шагов:

1. Выбирается число кластеров k, на которые нужно разбить данные.

2. Инициализация центроидов:

Инициализируются k начальных центроидов $\{\mu_1, \mu_2, \dots, \mu_k\}$, которые могут быть выбраны случайным образом из точек данных или другими методами, например K-Means++.

3. Назначение точек кластерам:

Для каждой точки данных x_i вычисляется расстояние до каждого из центроидов μ_j:

$$d(\mathrm{x}_i, \mu_j) = \| \mathrm{x}_i - \mu_j \|. \quad (16)$$

Точка x_i назначается кластеру с минимальным расстоянием:

$$C_i = \underset{j}{\operatorname{argmin}} d(\mathrm{x}_i, \mu_j), \quad (17)$$

где C_i — кластер, к которому относится точка x_i.

4. Обновление центроидов:

После назначения всех точек пересчитываются центроиды каждого кластера:

$$\mu_j = \frac{1}{|C_j|} \sum_{\mathrm{x}_i \in C_j} \mathrm{x}_i, \quad (18)$$

где $|C_j|$ — количество точек в j-м кластере, а μ_j — новое положение центроида.

5. Повторение шагов 3 и 4:

Шаги 3 и 4 повторяются до тех пор, пока не сойдется процесс (например, пока центроиды не перестанут изменяться или не будет достигнуто максимальное количество итераций).

Метод K-средних минимизирует следующую функцию стоимости (функцию потерь):

$$J = \sum_{j=1}^{k} \sum_{\mathrm{x}_i \in C_j} \| \mathrm{x}_i - \mu_j \|^2, \quad (19)$$

где J — суммарное внутрикластерное отклонение, а $\| \mathrm{x}_i - \mu_j \|^2$ — квадрат евклидова расстояния между точкой данных и центроидом её кластера.

5. Спектральный анализ (PSD) для выявления частотных аномалий. Спектральный анализ, основанный на оценке спектральной плотности мощности (PSD), используется для выявления аномалий, проявляющихся в частотной области [5]. Основное выражение для вычисления PSD выглядит следующим образом: Спектральная плотность мощности $P(\omega)$ сигналов была рассчитана по следующей формуле:

$$P(\omega) = \frac{1}{N} \sum_{k=1}^{N} |X_k(\omega)|^2 , \quad (20)$$

где ω — частота, $X_k(\omega)$ — дискретное преобразование Фурье k-го сегмента сигнала, а N — количество сегментов.

Метод Уэлча используют для более точного определения спектра мощности. При этом сигнал разделяют на несколько частей, которые могут перекрываться. Затем для каждой части применяют преобразование Фурье. После этого вычисляют среднее значение спектров мощности всех сегментов. Это позволяет снизить влияние случайных помех и повысить стабильность оценки:

$$P_{\text{Welch}}(\omega) = \frac{1}{M} \sum_{m=1}^{M} P_m(\omega) , \quad (21)$$

где M — количество сегментов, а $P_m(\omega)$ — спектральная плотность мощности для m-го сегмента.

Чтобы глубже проанализировать спектральные характеристики и точно оценить разницу между гармоническими и аномальными сигналами, нужно вычислить интегральную энергию сигналов. Интегральная энергия сигнала E вычисляется путем интегрирования значений PSD $P(\omega)$ по всему диапазону частот ω:

$$E = \int_0^{\omega_{\max}} P(\omega)\, d\omega , \quad (21)$$

где $\omega_{\max}$ — максимальная частота, до которой выполняется интегрирование.

Этот переход к интегральной оценке энергии позволяет не только выявить, как энергия распределена по частотам, но и количественно оценить общую энергетическую составляющую сигналов, что является важным для более глубокого понимания их природы и для точной их классификации.

Для интеграции описанных алгоритмов в единую систему защиты КИИ предлагается многослойная архитектура, состоящая из следующих основных компонентов (рис. 1).

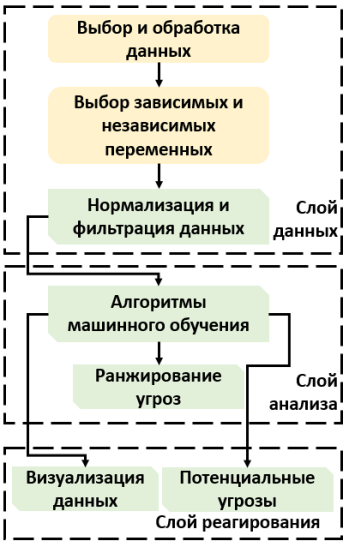

Рис. 1. Архитектура системы защиты КИИ на основе ИИ

Интеграция алгоритмов ИИ в КИИ требует комбинации методов, адаптированных к их природе [6]. Предложенная архитектура демонстрирует эффективность за счет использования Isolation Forest для оперативного обнаружения атак, применения

спектрального анализа для выявления скрытых угроз, минимизации ложных срабатываний через гибридный подход [7].

III. Моделирование аномалий в КИИ

Приведем экспериментальное моделирование кибератак на примере энергосистемы. Понимание аномалий в защите сетей электроснабжения невозможно без знаний особенностей технологического процесса. В этой связи, задачей исследования является проведение эксперимента, моделирующего отклонение напряжения в сети, и выбор такого метода анализа этой аномалии, который позволит максимально точно ее отличить от обычного аварийного режима работы электрической сети.

Для оценки эффективности алгоритмов ИИ были смоделированы три режима работы электрической сети (табл. 1) [8]. В нормальных условиях напряжение описывается синусоидальной функцией времени $U = f(t)$ с добавлением случайного шума, отражающего реальные флуктуации (рис. 2а). Для моделирования кибератаки был искусственно введен внезапный скачок напряжения на 0.3 кВ (рис. 1б). Аварийный режим работы с отклонением напряжения был смоделирован увеличением амплитуды синусоиды на определенный промежуток времени (рис. 1в)

ТАБЛИЦА I Различные режимы работы электрической сети

Режим	Формула / Описание	Длительность
Нормальный режим	$U(t) = U_{\text{ном}} \cdot sin(2\pi ft) + N(0, \sigma^2)$, $U\_\text{ном} = 0.4$ кВ, $f = 50$ Гц, $\sigma = 0.05$ кВ	*0–500 s*
Кибератака	Резкий скачок напряжения на $\Delta U = 0.3$ кВ в момент $t = 500$ с	*500–700 s*
Аварийный режим	$U(t) = (0.4 + 0.2 \cdot e^{-\frac{t-700}{100}}) \cdot sin(2\pi ft)$ – плавное увеличение амплитуды до $U_{\text{макс}} = 0.6$ кВ	*700–900 s*

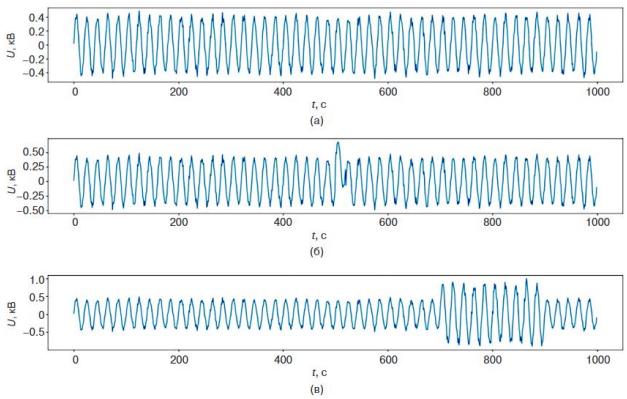

Рис. 2. Моделирование различных режимов работы электрической сети: нормальный режим работы (а); режим с кибератакой (б); обычный режим с отклонением напряжения (в)

В рамках данного исследования использованы несколько алгоритмов машинного обучения для анализа синтетических данных, моделирующих поведение электрической сети в условиях кибератаки и аварийного режима отклонения электрической нагрузки. На рис. 3 показаны тепловые карты аномалий, полученных с использованием различных алгоритмов ИИ.

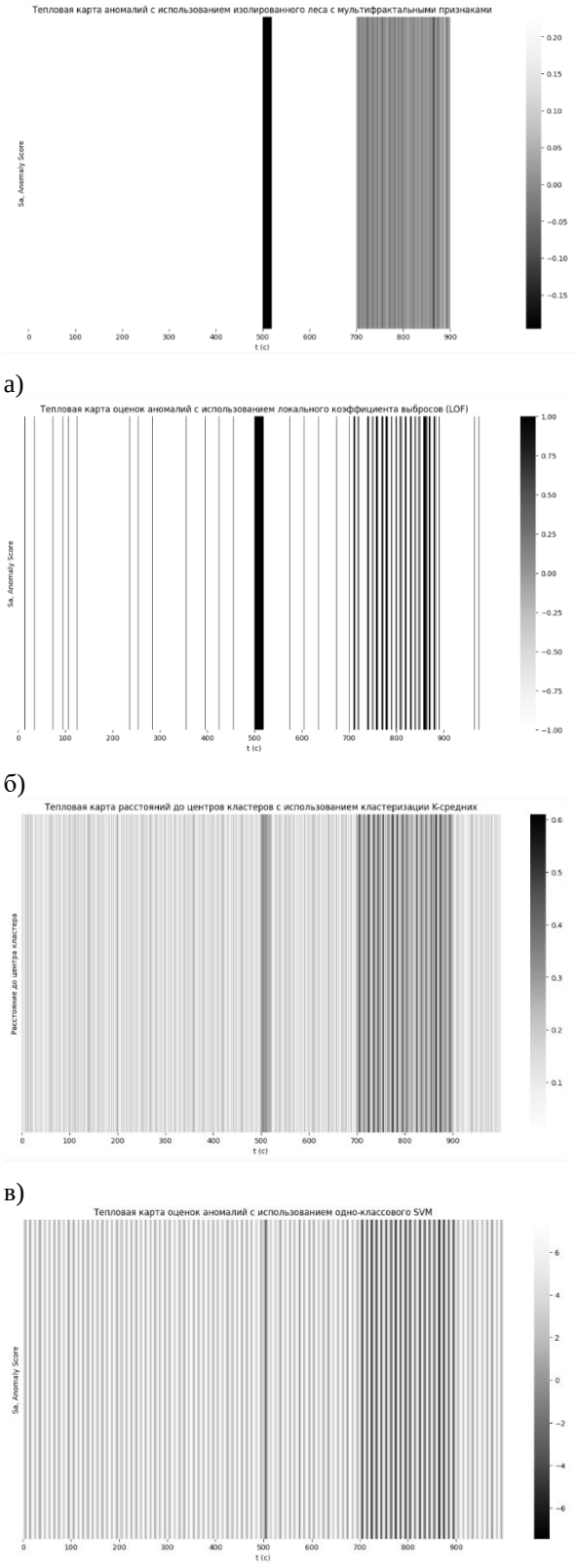

Рис. 3. Тепловая карта оценок аномалий

На основании проведенного анализа тепловых карт можно заключить, что различные методы обнаружения аномалий демонстрируют разную степень эффективности в контексте выявления киберугроз и других отклонений в электрических системах (табл. 2).

ТАБЛИЦА II РЕЖИМЫ РАБОТЫ ЭЛЕКТРИЧЕСКОЙ СЕТИ

Метод	**Точность**	**F1-Score**	**Время**	**Ложные срабатывания**
Isolation Forest	98%	0.96	20 ms	2%
LOF	89%	0.82	150 ms	15%
K-means	85%	0.78	50 ms	10%
One-Class SVM	92%	0.88	80 ms	5%

IV. Выводы

Метод Isolation Forest показал наилучшие результаты в обнаружении резких изменений, связанных с кибератаками, с высокой точностью и минимальными ложными срабатываниями. Метод LOF также эффективен, но его повышенная чувствительность к мелким отклонениям приводит к увеличению ложных срабатываний. Методы K-средних и One-Class SVM менее контрастны, но полезны для кластеризации данных и выявления как резких, так и плавных изменений. Таким образом, Isolation Forest является наиболее предпочтительным для выявления киберугроз, однако для комплексного анализа аномалий рекомендуется сочетание нескольких методов.

Список литературы

[1] Goodfellow, I., Y. Bengio, and A. Courville. *Deep Learning.* Cambridge: MIT Press, 2016. 800 p.

[2] Liu F.T., Ting K.M. & Zhou Z.-H. *Isolation Forest.* Proceedings of the 2008 IEEE International Conference on Data Mining. 2008, p. 413–422. DOI: 10.1109/ICDM.2008.17.

[3] Breunig M. M.; Kriegel H.-P.; Ng R. T.; Sander J. (2000). LOF: Identifying Density-based Local Outliers. *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data.* P. 93–104. doi:10.1145/335191.335388.

[4] Oliveri P. Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues. A tutorial. Analytica Chimica Acta. 982: 9-19. doi:10.1016/j.aca.2017.05.013. hdl:11567/881059.

[5] Welch, P. "The Use of Fast Fourier Transform for the Estimation of Power Spectra." *IEEE Transactions on Audio and Electroacoustics* 15, no. 2 (1967): 70–73.

[6] Иванов С.П., Смирнов К.О. Гибридные методы защиты критической инфраструктуры // Журнал "Кибербезопасность и защита информации". 2021. № 3. С. 72–85.

[7] Liu, F.T., K.M. Ting, and Z.H. Zhou. "Isolation Forest." *Proceedings of the 8th IEEE International Conference on Data Mining* (2008): 413–422.

[8] Кочергин С.В., Артемова С.В., Бакаев А.А., Митяков Е.С., Вегера Ж.Г., Максимова Е.А. Кибербезопасность смарт-сетей: сравнение подходов машинного обучения для обнаружения аномалий // Russian Technological Journal. 2024. Т. 12, № 6. С. 7–19. https://doi.org/10.32362/2500-316X-2024-12-6-7-19