Раздел I. Кибератаки и их обнаружение

УДК 004.056 DOI 10.18522/2311-3103-2026-3-6-18

А.В. Балыбердин

МЕТОД ОБНАРУЖЕНИЯ КОМПЬЮТЕРНЫХ АТАК НА ОСНОВЕ МОДЕЛИ АУГМЕНТАЦИИ ДАННЫХ СЕТЕВОГО ТРАФИКА H-DDPM

Рассматривается задача повышения выявления компьютерных атак (КА) системой обнаружения вторжений при значительном дисбалансе данных сетевого трафика. На основе проведенного анализа методов снижения дисбаланса данных сделан вывод, что классические методы балансировки и генеративной аугментации не обеспечивают сохранение статистической структуры многомерных табличных данных, в том числе их фрактальных свойств и самоподобия, что снижает качество обучения классификатора. В работе предлагается метод обнаружения компьютерных атак (КА) на основе модели аугментации данных H-DDPM, представляющей собой модификацию диффузионной вероятностной модели DDPM, в которой дисперсия добавляемого гауссова шума в прямом процессе зависит от показателя Херста H для каждого класса КА. Метод включает предобработку данных, формирование временных рядов по скользящим окнам, оценку H методами DFA и R/S и генерацию синтетических данных для обучения классификатора LSTM. Метод оценивается по общим метрикам эффективности Accuracy, Recall, F1, ROC-AUC и G-means, а также Precision, Recall и F1-score для классов. Эксперименты проведены на наборах CSE-CSE-CIC-IDS2018 и UNSW-NB15. Выполнено сравнение с другими методами, такими как SMOTE, GAN и DDPM. Результаты экспериментов показывают, что H-DDPM повышает эффективность обнаружения КА, превосходя аналогичные методы по метрикам, чувствительным к дисбалансу. Также, на основе экспериментальных проверок показано, что непосредственно применение показателей Херста H для классов КА в модели H-DDPM повышает полноту и сбалансированное качество обнаружения КА. Отмечено, что H-DDPM оказывает влияние на классификацию КА, проявляющееся в росте ложноположительных срабатываний и снижении метрики ROC-AUC, что требует дополнительной настройки гиперпараметров модели классификатора и фильтрации синтетических данных.

Система обнаружения вторжений; сетевой трафик; аугментация данных; диффузионная модель; показатель Херста H; фрактальные свойства; самоподобие.

A.V. Balyberdin

A METHOD FOR DETECTING COMPUTER ATTACKS BASED ON H-DDPM NETWORK TRAFFIC DATA AUGMENTATION MODEL

This paper examines the problem of improving the detection of computer attacks (CA) by an intrusion detection system (IDS) under conditions of significant network traffic data imbalance. Based on an analysis of methods for reducing data imbalance, it is concluded that classical methods of balancing and generative augmentation do not preserve the statistical structure of multidimensional tabular data, including their fractal properties and self-similarity, which reduces the quality of classifier training. This paper proposes a method for detecting computer attacks (CA) based on the H-DDPM data augmentation model, a modification of the DDPM diffusion probabilistic model, in which the variance of the added Gaussian noise in the forward process depends on the Hurst exponent H for each CA class. The method includes data preprocessing, the formation of time series using sliding windows, H estimation using DFA and R/S methods, and the generation of synthetic data for training the LSTM classifier. The method is evaluated using the general performance metrics Accuracy, Recall, F1, ROC-AUC, and G-means, as well as Precision, Recall, and F1-score for each class. Experiments were conducted on the CSE-CSE-CIC-IDS2018 and UNSW-NB15 datasets. A comparison was made with other methods, such as SMOTE, GAN, and DDPM. The experimental results show that H-DDPM improves the efficiency of CA detection, outperforming similar methods in terms of imbalance-sensitive metrics. Furthermore, experimental validations demonstrate that directly using the Hurst H exponent for CA classes in the H-DDPM model improves the recall and balanced quality

of CA detection. It is noted that H-DDPM has an impact on CA classification, manifested by an increase in false positives and a decrease in the ROC-AUC metric, which requires additional tuning of the classifier model hyperparameters and filtering of synthetic data.

Intrusion detection system; network traffic; data augmentation; diffusion model; Hurst exponent; fractal properties; self-similarity.

Введение. Анализ отраслевых отчетов по кибербезопасности показывает устойчивое увеличение числа угроз и инцидентов информационной безопасности [1, 2]. Доступные злоумышленникам средства автоматизации позволяют оперативно формировать новые векторы кибератак на защищаемые информационные ресурсы организации. Система обнаружения вторжений (СОВ) является ключевым элементом защиты от внутренних и внешних угроз, так как осуществляют анализ сетевого трафика и на основе встроенных алгоритмов корреляции и классификации выявляют компьютерные атаки (КА) в сети [3].

В современных СОВ широко применяются классические методы машинного обучения, а также поведенческие методы обнаружения, основанные на анализе статистических характеристик сетевого трафика. Однако значительные объемы несбалансированных данных и разнообразие сетевого трафика снижают эффективность применения данных подходов, что обусловливает необходимость использования более устойчивых и точных методов обнаружения КА в сетевом трафике [4].

Модели глубокого обучения показали высокую эффективность для решения задач классификации и кластеризации в различных предметных доменах, что способствовало развитию данных моделей в области обнаружения КА в КС [5]. Однако для эффективного применения моделей глубокого обучения требуется значительный объем высококачественных и сбалансированных данных. Доступные открытые наборы данных, а также наборы, сформированные в реальных корпоративных сетях передачи данных, характеризуются существенным дисбалансом между нормальным сетевым трафиком и КА. Дисбаланс миноритарных классов приводит к смещению модели в сторону мажоритарных классов, вследствие этого часть КА может быть необнаруженной [6].

Для снижения влияния дисбаланса в классах и повышения эффективности обучения глубоких моделей существует два основных подхода: снижение количества данных и генерация синтетических данных, то есть аугментация данных. Для аугментации данных широко применяют генеративные модели на основе генеративно-состязательных сетей (GAN) и вариационных автоэнкодеров (VAE). Несмотря на значительные результаты применения генеративных моделей, при генерации структурированного многомерного табличного и временного трафика они по-прежнему сталкиваются с ограничениями. Ограничения связаны не только с разнородностью данных и высокой степенью коррелированности признаков, но и с тем, что в большинстве существующих подходов не учитываются фрактальные свойства и самоподобие данных, что снижает качество генерируемых синтетических данных.

Для преодоления рассмотренных ограничений в работе предлагается новая модель генеративной аугментации сетевого трафика H-DDPM, ориентированная на сохранение самоподобной структуры трафика и повышение эффективности обнаружения КА в КС. Для этого при обучении H-DDPM на реальном сетевом трафике в прямой диффузионный процесс к данным добавляется гауссовский шум с дисперсией, параметризуемой показателем Херста H, характеризующим самоподобие сетевого трафика. В обратном процессе модель восстанавливает синтетические образцы, согласованные с заданным значением H. Включение показателя Херста в диффузионный процесс позволяет лучше сохранять фрактальные свойства, а именно самоподобие сетевого трафика, что приводит к генерации более реалистичных и разнообразных синтетических данных. Экспериментальные результаты показывают, что модель H-DDPM демонстрирует повышенную стабильность генерации и лучше захватывает сложные шаблоны в данных по сравнению с классическими генеративными моделями.

1. Постановка задачи. Модели глубокого обучения широко применяются для повышения эффективности обнаружения аномалий и компьютерных атак (КА) в КС. Однако значительный дисбаланс данных и недостаток высококачественных размеченных данных по классам КА существенно ограничивают эффективность данных моделей и тем самым снижают уровень защищенности в организации.

В связи с этим возникает задача разработки генеративной модели аугментации данных, позволяющей формировать реалистичные синтетические примеры КА и снижать влияние дисбаланса данных в обучающей выборке, в том числе на общедоступных наборах данных (CSE-CSE-CIC-IDS2018, UNSW-NB15 и т.д.).

В качестве основы модели используется диффузионная вероятностная модель DDPM, обладающая высокой устойчивостью в генерации синтетических данных и способностью создавать разнообразные варианты образцов сетевого трафика. В отличие от существующих диффузионных подходов, требуемая модель должна учитывать долговременную зависимость (самоподобие) сетевого трафика, обеспечивая сохранение его характерной статистической структуры в сгенерированных данных. Результатом работы модели являются сгенерированные синтетические данные, которые используются для обучения глубоких нейронных сетей, повышая тем самым полноту и качество обнаружения КА в КС.

Целью работы является разработка метода обнаружения КА в КС и модели аугментации данных, позволяющей генерировать синтетический сетевой трафик с сохранением его самоподобия и тем самым снижать влияние дисбаланса данных и повышать эффективность обнаружения КА в КС.

2. Релевантные работы. В настоящее время для эффективного применения моделей машинного обучения в задачах классификации, используют методы обучения с учителем, при этом реальные наборы данных часто характеризуются значительным дисбалансом данных в классах. В работе [7] автор подчеркивает, что несбалансированность данных при обучении влияет на обобщающуюся способность моделей машинного обучения, в особенности на модели глубокого обучения. Автор показывает, что проблема дисбаланса данных характерна для разных областей, таких как изображение, текст, графы, табличные данные и временные ряды. Для снижения влияния дисбаланса широко применяют методы аугментации данных, направленные на искусственное расширение обучающего набора, в том числе за счёт генерации дополнительных примеров для миноритарных классов. В [7] отдельно выделяется dataset-level аугментация, включающая генеративные модели на основе автоэнкодеров, GAN, диффузионных моделей, LLM и др. В статье [8] подчёркивается, что классические (эвристические) операции аугментации недостаточно эффективны для задач со сложными многомерными данными, при этом диффузионные модели демонстрируют более высокое качество сгенерированных изображений по сравнению с GAN. В статье [9] на основе 5 генеративных моделей (TVAE, CTGAN, TabDDPM, TabSyn, SMOTE / ucSMOTE) на 16 общедоступных наборах данных показано, что диффузионные модели в среднем превосходят классические подходы (SMOTE / ocSMOTE) по качеству синтетических данных, в особенности при детальной настройке гиперпараметров моделей.

Сетевой трафик представляет собой табличные данные, и для снижения дисбаланса целесообразно использовать диффузионные модели. В исследовании [10] автор для повышения точности обнаружения КА, генерирует синтетические данные на основе классической диффузионной модели DDPM с добавлением гауссовского шума для формирования новой обучающей выборки. Отмечено улучшение выявления КА по метрикам эффективности: Accuracy, Precision, Recall и F1. Для классов КА R2L и U2R точность достигает 0.692 и 0.610 соответственно, что требует дальнейшего улучшения и снижения времени генерации синтетических данных.

Для оптимизации времени аугментации на основе DDPM в работе [11] предлагается новый подход к применению диффузионной модели для аугментации признаков: в прямом процессе DDPM к признакам добавляется шум, а в обратном процессе предсказывается $\varepsilon_\theta(x_t, t)$ с последующим преобразованием методом Feature Masking в маску $\sigma(x_t, t) = \frac{1}{1+e^{-\varepsilon_\theta(x_t,t)}}$ и поэлементным перемножением с $\varepsilon_\theta(x_t, t)$. Отмечается, что аугментация признаков выполняется быстрее, чем аугментация данных, однако эффект по качеству незначителен: точность мультиклассовой классификации 99.89 → 99.93 (+ 0.04 п.п.), бинарной – 99.93 → 99.94 (+ 0.01 п.п.).

В работе [12] сетевой трафик из PCAP-файла преобразуется в признаки трёх уровней (все признаки, уровень приложений и уровень сессий), формируя тем самым RGB-изображения признаков. Улучшенная диффузионная модель NT-DDPM за счёт косинусного расписания шума обучается на этих RGB-изображениях. На наборах данных CTU, USTC-TFC, ISAC и UJS-IDS2024 модель увеличивает точность на 0.1–0.3 процентных пункта по метрикам Accuracy и F1, однако в работе отсутствует анализ по классам КА.

Несмотря на высокие значения Accuracy и F1, в рассмотренных выше работах не оцениваются метрики качества аугментированных данных: разнообразие, динамика (дрифт концепций) и временные зависимости синтетических данных. В продолжение исследований в работе [13] для модели DATG на основе DDPM со встроенным self-attention на каждом шаге диффузии используются следующие оценки качества аугментации: Jensen–Shannon divergence (JSD) и Continuous Ranked Probability Score (CRPS). Улучшение этих метрик на наборах данных ISCX-VPN2016 и UNSW-NB15 составляет 23–25% по сравнению с GAN. В работе рассматривается только нормальный трафик, в отличие от необходимых для задач обнаружения КА в КС.

Рассмотренные выше работы не учитывают фрактальные свойства сетевого трафика, в частности свойство самоподобия. Для более реалистичных и качественных аугментированных данных представляется целесообразным учитывать самоподобие сетевого трафика при генерации синтетических данных на основе диффузионной модели.

3. Метод обнаружения компьютерных атак на основе модели H-DDPM. ***3.1. Архитектура метода обнаружения КА.*** Метод обнаружения КА на основе модели H-DDPM состоит из следующих этапов: предобработка, расчет H, модель H-DDPM, оценка эффективности.

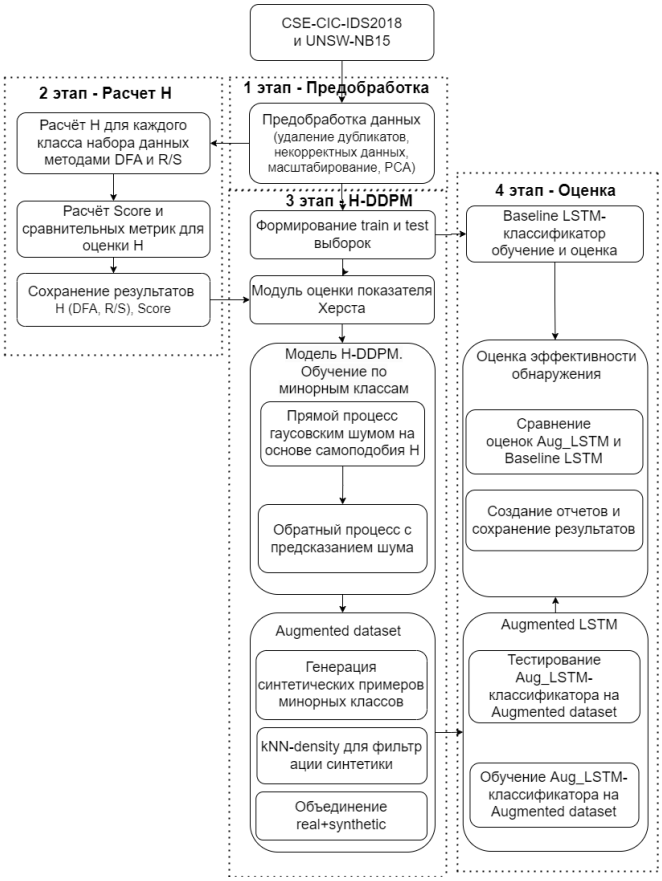

Рис. 1. Архитектура метода обнаружения КА на основе модели H-DDPM

На первом этапе предобработки выполняется удаление дубликатов и обработка пропущенных данных. Для приведения значений к одному виду производится линейное масштабирование значений признаков сетевого трафика в диапозоне [0;1] на основе алгоритма Min-Max Scaler $\tilde{x}_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}, i = 1, \dots, n, j = 1, \dots, d$, где i – событие, а j – признака сетевого трафика, x – значение признака сетевого трафика.

На втором этапе для предобработанных данных выполняется вычисление фрактального свойства самоподобия сетевого трафика для каждого класса. Для этого данные на основе скользящих окон преобразуются во временные ряды с разделением на равные сегменты. Для каждого временного ряда вычисляется показатель Херста H двумя методами: детрендированным флуктуационным анализом (DFA) и R/S анализом [14]. Для полученных показателей H вычисляется медианное значение для каждого класса набора данных.

На третьем этапе формируются выборки $X_{\text{train}}, X_{\text{test}}$ 80%, 20% соответственно. Для каждого класса используют оценки DFA и R/S вычисленные на предыдущем этапе, которые применяют для модификации расписания шумов для обучении модели H-DDPM (см. п.3.2). Обученная H-DDPM генерирует синтетические примеры $\tilde{x}$, которые проходят kNN-фильтрацию (фильтруются точки с наименьшей плотностью относительно X_{real}). Отфильтрованные данные объединяются в новый набор данных ($X_{\text{train}}^{\text{aug}}, y_{\text{train}}^{\text{aug}}$).

На заключительном четвертном этапе выполняется оценка эффективности по модели LSTM, обученной на реальных данных и на новом наборе данных. Эффективность определяется на основе оценок Accuracy, Recall, F1, G-mean, Roc Auc.

3.2. Модель аугментации минорных классов H-DDPM. Для снижения влияния дисбаланса и повышения качества синтетических данных разработана модель аугментации H-DDPM. Модель H-DDPM представляет собой диффузионную модель, в которой реализовано динамическое изменение дисперсии гауссовского шума в зависимости от показателя самоподобия сетевого трафика H для каждого класса.

На основе прямого диффузионного преобразования данных [15] модель H-DDPM для шага t-1 -> t можно представить следующим образом:

$$q(x_t|x_{t-1}) = N\left(x_t; \sqrt{1-\beta_t(H)} \cdot x_{t-1}, \beta_t(H) \cdot I\right), \quad (1)$$

где $\beta_t(H)$ – динамическая дисперсия шума, зависящая от показателя Херста H и шага t. Для сохранения свойства самоподобия сетевого трафика учитывается, что при H > 0,5 дисперсия шума растет медленнее, чем при H < 0,5. Для выполнения данного условия добавлен коэффициент затухания $\lambda > 0$ *в* дисперсию шума $\beta_t(H)$:

$$\beta_t(H) = \beta_{base}(t) \cdot exp\left(-\lambda \cdot (H - 0{,}5)\right), \quad (2)$$

где $\beta_{base}(t)$ – линейная базовая дисперсия.

Обозначим $\alpha_t(H) = 1 - \beta_t(H), \qquad \bar{\alpha}_t(H) = \prod_{s=1}^{t} \alpha_s(H)$.

где $\alpha_t(H)$ – коэффициент сигнала с учетом добавления шума на шаге t; $\bar{\alpha}_t(H)$ – накопленный коэффициент сохраненного сигнала с первого шага до t.

Прямое преобразование модели H-DDPM можно представить следующим образом:

$$q(x_t \mid x_0, H) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t(H)}\, x_0, (1 - \bar{\alpha}_t(H))I\right), \quad (3)$$

При обратном процессе модель H-DDPM последовательно восстанавливает данные из зашумлённых представлений, поэтапно удаляя добавленный гауссовский шум из прямого процесса. На каждом шаге t обратный переход аппроксимируется гауссовским распределением:

$$p_\theta(x_{t-1} \mid x_t, H) = \mathcal{N}(x_{t-1};\, \mu_\theta(x_t, t, H),\, \Sigma_t(H)), \quad (4)$$

где $\mu_\theta(x_t, t, H)$ – среднее, параметризованное нейросетевой моделью через предсказанный шум; $\Sigma_t(H)$ – дисперсия шума, который добавляется на шаге t, зависящий от расписания $\beta_t(H)$.

В обратном процессе H-DDPM используется *шумовая параметризация*: нейросеть $\varepsilon_\theta(x_t, t, H)$ аппроксимирует добавленный в прямом процессе шум ε. Тогда среднее обратного перехода выражается через ε_θ стандартной DDPM-формулой, модифицированной H-зависимыми коэффициентами:

$$\mu_\theta(x_t, t, H) = \frac{1}{\sqrt{\alpha_t(H)}}\left(x_t - \frac{\beta_t(H)}{\sqrt{1-\bar{\alpha}_t(H)}}\, \varepsilon_\theta(x_t, t, H)\right). \quad (5)$$

Дисперсия шума $\Sigma_t(H)$ задаётся аналитически в виде скалярной дисперсии, зависящей от H:

$$\Sigma_t(H) = \sigma_t^2(H)\, I, \qquad \sigma_t^2(H) = \frac{1-\bar{\alpha}_{t-1}(H)}{1-\bar{\alpha}_t(H)}\, \beta_t(H). \quad (6)$$

Модель H-DDPM обучается на основе минимизации среднеквадратичного отклонения между истинным шумом ε, добавленным на шаге прямого процесса, и его оценкой $\varepsilon_\theta(x_t, t, H)$. За основу функции потерь модели H-DDPM используется оценка градиента *Score-based generative modeling* представленной в работе [16]. Таким образом, функцию потерь для модели H-DDPM можно записать следующим образом:

$$L(\theta) = \mathbb{E}_{t,x_0,\varepsilon}[\| \varepsilon - \varepsilon_\theta(x_t, t, H) \|_2^2]. \quad (7)$$

Минимизация функции потерь $L(\theta)$ по параметрам нейросетевой аппроксимирующей функции $\varepsilon_\theta(x_t, t, H)$, приближающей предсказанный шум $\varepsilon_\theta(x_t, t, H)$ к истинному шуму ε, добавленному в прямом диффузионном процессе, в сочетании с H-зависимым расписанием $\beta_t(H)$, учитывающим показатель Херста, обеспечивает генерацию синтетических данных для классов КА, которые лучше сохраняют самоподобную структуру сетевого трафика и, тем самым, обладают более высоким качеством.

4. Результаты экспериментов и метрики оценки эффективности. В данном разделе проводится экспериментальная оценка метода обнаружения КА на основе модели H-DDPM. Для этого представлены наборы данных для экспериментов и описаны метрики оценки эффективности. Результаты экспериментов по оценке эффективности метода обнаружения представлены в виде таблиц и графиков. Эксперименты выполнены на рабочем ноутбуке AMD Ryzen 7 8845HS w/ Radeon 780M Graphics (3.80 GHz), 32,0 ГБ, ОС Windows 11 Pro с использованием IDE Tool PyCharm 2.6.0, Programming Language Python3.9 Deep Learning Framework Deep Learning Framework Tensorflow 2.19.0.

4.1. Наборы данных CSE-CSE-CIC-IDS2018 и UNSW-NB15. В экспериментальной оценке эффективности метода используются два набора данных CSE-CIC-IDS2018 [17] и UNSW-NB15 [18]. Каждый набор данных разделен на две выборки: обучающая и тестовая выборки. Набора CSE-CIC-IDS2018 размечен, содержит 86 признаков, обучающая выборка состоит из 1,165,442 событий сетевого трафика, а тестовая – 499,476 (табл. 1). Набор UNSW-NB15 также размечен, содержит меньшее количество признаков – 49, обучающая выборка состоит из 715,827 событий сетевого трафика, а тестовая – 306,791 (табл. 2).

Таблица 1

Распределение данных по классам для набора CSE-CSE-CIC-IDS2018

Class	TRAIN	TEST
BENIGN	1,107,792	474,769
Botnet	515	221
Botnet - Attempted	2,847	1,220
DoS Slowloris	2,701	1,158
DoS Slowloris - Attempted	1,293	554
Infiltration	25	11
Infiltration - Attempted	32	13
Infiltration - Portscan	50,237	21,530
Sum	1,165,442	499,476

Таблица 2

Распределение данных по классам для набора UNSW-NB15

Class	TRAIN	TEST
Normal	705,931	302,543
Analysis	201	86
Backdoors	198	85
DoS	566	243
Exploits	2,818	1,208
Fuzzers	2,761	1,183
Generic	1,977	847
Reconnaissance	1,218	522
Shellcode	156	67
Worms	17	7
Sum	715,827	306,791

Для каждого класса формируется временной ряд на основе упорядочивания сетевых потоков по времени и их агрегации в скользящем окне $W = 10$ мин. с шагом $S = 5$ мин., где значения ряда задаются метрикой CountFlows (число потоков в окне) и, при необходимости, SumBytes/SumDuration. При отсутствии признаков времени используется оконная схема по событиям: окна фиксированной длины $W = 2000$ потоков со сдвигом $S = 1000$. Показатель H оценивается по каждому окну методами DFA и R/S. Результирующее значение H для каждого класса вычисляется как медиана по окнам и используется в модели H-DDPM.

4.2. Метрики оценки эффективности метода. Для оценки эффективности метода обнаружения КА в КС на основе модели H-DDPM в условиях несбалансированного сетевого трафика выбран следующий набор метрик, рекомендованный в обзорной работе [19]:

- Accuracy – базовая метрика оценивает общую точность классификации по всем событиям набора данных.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \tag{8}$$

где TP – количество истинно-положительных результатов; TN – количество истинно-отрицательных результатов; FP – количество ложноположительных результатов; FN – количество ложноотрицательных результатов.

- Precision – метрика точности обнаружения КА среди всех событий, классифицированных как КА.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{9}$$

- Recall – метрика полноты, отражающая долю корректно выявленных КА среди всех событий, относящихся к классу КА (включая пропущенные).

$$\text{Recall} = \frac{TP}{TP+FN} \tag{10}$$

- F1-мера – агрегированная метрика, отражающая баланс между точностью и полнотой обнаружения КА.

$$\text{F1} = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} = \frac{2TP}{2TP+FP+FN} \tag{11}$$

- G-means – показывает геометрическое среднее качества распознавания положительного и отрицательного классов. При сильном дисбалансе высокий G-means означает, что классификатор эффективно различает оба класса, в отличие от низкого значения, при котором предсказывается преимущественно мажоритарный класс.

$$\text{G-means} = \sqrt{\left(\frac{TP}{TP+FN}\right)\cdot\left(\frac{TN}{TN+FP}\right)} \tag{12}$$

- ROC-AUC – метрика, представляющая собой среднее по всем классам значение площади под ROC-кривой и определяет степень отличия каждого класса в наборе данных от остальных классов.

$$\text{ROC-AUC} \;=\; \frac{1}{K}\sum_{c=1}^{K}\text{AUC}_c\,, \tag{13}$$

где K – число классов (BENIGN и все типы КА), а AUC_c – площадь под ROC-кривой для класса c.

4.3. Экспериментальная оценка эффективности метода. *4.3.1. Оценка точности обнаружения.* Для экспериментальной оценки эффективности метода были выбраны две классические генеративные модели – GAN и DDPM, а также статистический метод аугментации минорных классов SMOTE. Оценка производилась по общим метрикам Accuracy, Recall, G-means, F1 и ROC-AUC, а также по каждому классу отдельно по метрикам Precision, Recall и F1-score.

Для модели H-DDPM во всех экспериментах использовалось стандартное линейное расписание β_t и экспоненциальная модуляция по показателю Херста H. Сеть ε_θ реализована на основе MLP и обучалась при неизменном количестве эпох. Синтетика после генерации

проходила одинаковую kNN-фильтрацию, затем один и тот же LSTM-классификатор обучался в идентичных условиях. В экспериментах использовался одинаковый объём синтетических данных для разных методов.

Результаты обнаружения КА методом на основе модели H-DDPM представлены в табл. 3 и 4. Результаты (табл. 3, 4) показывают, что H-DDPM обеспечивает наиболее высокую полноту обнаружения миноритарных классов и лучшее качество их распознавания (0.976 - Recall (macro), 0.975 - G-means) по сравнению с GAN и DDPM и также сопоставим с SMOTE. Отметим, что показатели H-DDPM по данным метрикам незначительно превышают показатели метода SMOTE. Вместе с тем метод на основе H-DDPM увеличивает число ложноположительных срабатываний (ЛПС), это связано с тем, что модель смещает оценку в сторону миноритарных классов, тем самым снижая пропуски КА ценой роста ЛПС. Для снижения ЛПС необходимы дополнительные методы фильтрации синтетических данных и настройки гиперпараметров модели LSTM.

Таблица 3

Сравнение методов аугментации обучающей выборки CSE-CSE-CIC-IDS2018

Метод	Accuracy	Recall (macro)	G-means	F1 (macro)	ROC-AUC (macro, OVR)
Без аугментации	0.998	0.695	0.001	0.684	0.998
SMOTE	0.990	0.967	0.965	0.618	0.989
Классический DDPM	0.997	0.552	0.000	0.569	0.998
Классический GAN	0.996	0.585	0.000	0.576	0.998
H-DDPM	0.952	**0.976**	**0.975**	0.458	0.996

Высокие значения метрик Recall и G-means для метода обнаружения на основе H-DDPM показывают, что данный подход позволяет выявлять КА в условиях значительного дисбаланса данных и низкого объема экземпляров КА, в отличие от методов, которые не выявили КА для данных классов. На рис. 2 представлены три минорных класса КА: Botnet – 736 экземпляров, Infiltration – 36 экз., Infiltration – Attempted – 45 экз. Метод на основе H-DDPM обнаружил эти классы и показал лучшие результаты по сравнению с другими методами аугментации данных (рис. 2).

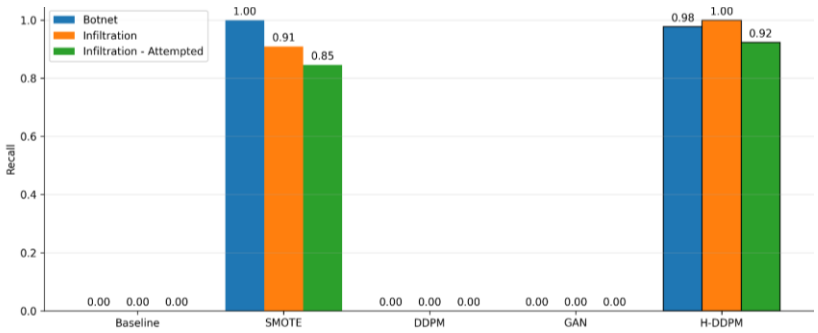

Рис. 2. Метрики обнаружения минорных классов КА для набора данных CSE-CSE-CIC-IDS2018

Дополнительно для проверки метода обнаружения КА на основе H-DDPM был взят набор данных UNSW-NB15 по которому выполнено сравнение метода без аугментации, SMOTE, а также генеративных моделей GAN, DDPM и предлагаемой модели H-DDPM по метрикам Accuracy, Recall (macro), G-means, F1 (macro) и ROC-AUC (macro, OVR) (табл. 4). На основе анализа результатов (табл. 4), H-DDPM обеспечивает наилучшие значения метрик, чувствительных к дисбалансу, достигая Recall (macro) = 0,554, G-means = 0,476 и F1 (macro) = 0,428, что превосходит SMOTE (0,511; 0,388; 0,397) и существенно выше результатов других моделей GAN и DDPM по G-means и F1.

При этом по метрике ROC-AUC macro у H-DDPM наблюдается снижение до 0,991 по сравнению с 0,997–0,998 у остальных подходов, что указывает на необходимость дополнительной настройки гиперпараметров модели.

Таблица 4

Сравнение методов аугментации обучающей выборки UNSW-NB15

Метод	Accuracy	Recall (macro)	G-means	F1 (macro)	ROC-AUC (macro, OVR)
Без аугментации	0,992	0,363	0,000	0,328	0,998
SMOTE	0,990	0,511	0,388	0,397	0,997
Классический DDPM	0,992	0,360	0,000	0,328	0,997
Классический GAN	0,996	0,332	0,000	0,302	0,998
H-DDPM	**0,989**	**0,554**	**0,476**	**0,428**	0,991

На рис. 3 показано, что H-DDPM существенно повышает полноту распознавания отдельных миноритарных классов (Shellcode – 0,60, Backdoor – 0,35, Analysis – 0,48, Worms – 0,14), тогда как метод без аугментации, DDPM и GAN не обнаруживают данные классы. В целом результаты подтверждают, что H-DDPM является наиболее эффективным методом аугментации для обнаружения КА в КС в условиях значительного дисбаланса данных на наборе данных UNSW-NB15.

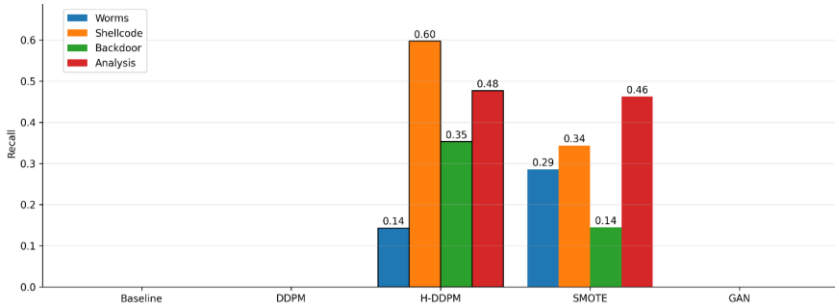

Рис. 3. Метрики обнаружения минорных классов КА для набора данных UNSW-NB15

В результате экспериментальной оценки на наборах CSE-CSE-CIC-IDS2018 и UNSW-NB15 показано, что предлагаемый метод H-DDPM обеспечивает наилучшую полноту и качество обнаружения классов КА, повышая Recall (macro) и G-means по сравнению с классическими генеративными моделями GAN и DDPM и демонстрирует результаты, сопоставимые или превосходящие метод SMOTE. Одновременно наблюдается рост ложноположительных срабатываний и снижение ROC-AUC macro для UNSW-NB15, что требует дополнительной настройки гиперпараметров модели LSTM и фильтрации синтетических данных. Таким образом, H-DDPM может рассматриваться как эффективный метод аугментации сетевого трафика для повышения устойчивости обнаружения КА в условиях значительного дисбаланса данных сетевого трафика.

4.3.2. Анализ влияния показателя Херста H модели H-DDPM на обнаружение КА в КС. Для проверки влияния показателя Херста H модели H-DDPM на обнаружение КА в КС проведены сравнительные эксперименты со следующими параметрами H: H для каждого класса на основе набора данных (per_class), перестановка H между выбранными классами (shuffled) и единый H_0 для всех классов (constant). В соответствии с данной методикой проведения экспериментов предполагается, что метрики обнаружения КА будут выше для рассчитанных показателей Херста H каждого класса набора данных модели H-DDPM (per_class) по сравнению с другими значениями показателей Херста H с режимами shuffled и constant.

Применяемый подход согласуется с логикой экспериментов, используемой в работе [20], где вклад ключевого компонента оценивается путём сравнения качества полной модели и варианта без данного компонента.

Эксперименты проведены на основе набора данных UNSW-NB15. Для каждого класса набора UNSW-NB15 рассчитывается показатель Херста *H* методами DFA и R/S. Результаты вычислений представлены на рис. 4. Более устойчивые значения *H* соответствуют методу R/S, так как имеет незначительный разброс в оценках *H*.

Для проверки влияния выбраны мажоритарный класс КА Fuzzers (обучающая train – 2,761 экз.) и миноритарный Shellcode (выборка train – 156 экз.). Для оценки влияния *H* взята метрика CountFlows, которая соответствует количеству событий для каждого класса КА набора данных UNSW-NB15 выбранного окна, используемого для расчета показателя Херста *H*. Для оценки влияния H выбираем среднее значение между дельтой *H* класса Fuzzer КА и *H* класса Shellcode КА (рис. 5).

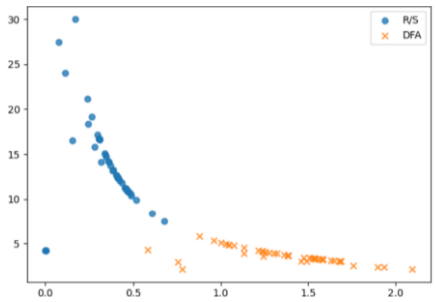

Рис. 4. Показатели Херста H методами DFA и R/S для КА набора данных UNSW-NB15

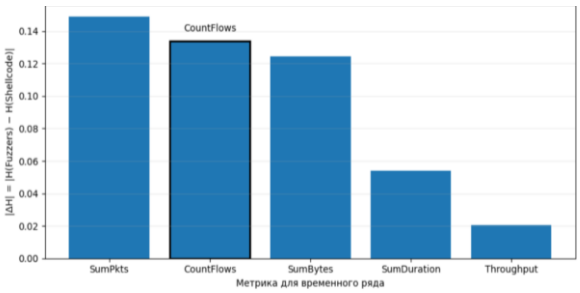

Рис. 5. Дельта H для различных метрик (признаков) набора данных UNSW-NB15

Для базового режима per_class метода R/S и метрики CountFlows используем вычисленные *H* для набора данных UNSW-NB15 для классов КА Fuzzers и Shellcode. Для режима shuffled значения *H* меняются между данными классами, а для режима constant считается среднее значение H по классам Fuzzers и Shellcode.

Результаты экспериментов по проверке влияния показателя Херста *H* на метрики обнаружения КА представлены в табл. 5. Режим per_class корректного соответствия показателя H классу существенно повышает полноту и качество распознавания миноритарных классов КА, что выражается в резком росте показателя G-Mean (табл. 5).

Таблица 5

Общие метрики оценки обнаружения в зависимости от режимов

Режим	Accuracy	F1_macro	Recall_macro	G-Mean	ROC-AUC
per_class	**0.988895**	0.402202	**0.528563**	**0.448658**	**0.985082**
constant	0.987135	0.408182	0.529337	0.035421	0.965564
shuffled	0.986382	0.401937	0.521844	0.034731	0.962679

Перемешивание показателей *H* (shuffled) или их унификация (constant) приводит к деградации метрики G-Mean (см. табл. 5) и низким значениям метрики полноты Recall для миноритарных классов КА (табл. 6).

Таблица 6

Метрика полноты Recall для различных классов КА

Класс	per_class recall	constant recall	shuffled recall
Shellcode	**0.5522**	0.4328	0.4030
Worms	**0.1429**	0.0000	0.0000
Fuzzers	0.6729	0.6915	0.7743

На основе результатов проведенных экспериментов по анализу влияния показателя Херста H можно сделать вывод, что применение показателя Херста *H* для каждого класса в прямом процессе диффузионной модели H-DDPM оказывает непосредственно положительное влияние и повышает полноту и сбалансированное качество обнаружения КА в КС при значительном дисбалансе данных.

Заключение. В работе проведен анализ современных методов, направленных на повышение полноты обнаружения КА и снижения дисбаланса классов. Анализ данных методов показал, что одним из наиболее эффективных способов является применение генеративных моделей, в особенности диффузионных, которые позволяют выполнять аугментацию данных КА.

Научная новизна работы заключается в усовершенствовании метода обнаружения КА в КС, отличающегося от аналогичных методов использованием модели H-DDPM, которая обеспечивает создание синтетических образцов КА на основе диффузионной модели с сохранением свойств самоподобия сетевого трафика, что позволяет повысить полноту обнаружения и снизить влияние дисбаланса классов.

Метод обнаружения на основе модели H-DDPM показал лучшие результаты по метрикам полноты обнаружения Recall (macro) и G-means. Для набора данных CSE-CSE-CIC-IDS2018 метрика Recall (macro) увеличились с 0,552 до 0,976, а G-means с 0 до 0,975 в сравнении с методами без аугментации и с аугментацией генеративных моделей. Дополнительно модель H-DDPM была проверена на наборе данных UNSW-NB15, метрика Recall (macro) увеличилась с 0,332 до 0,554, а G-means с 0 до 0,476.

Одновременно отмечено незначительное увеличение числа ложноположительных срабатываний и снижение значения метрики ROC-AUC (для CSE-CSE-CIC-IDS2018 снижение с 0.998 до 0.996, UNSW-NB15 снижение с 0,998 до 0,991), что свидетельствует о необходимости дополнительной настройки гиперпараметров классификатора и фильтрации синтетических данных.

Анализ влияния показателя Херста H в модели H-DDPM показал, что генерация синтетических образцов КА с сохранением самоподобия сетевого трафика позволяет также повысить метрики полноты обнаружения КА в КС G-Mean с 0.034731 до 0.448658.

В дальнейшем планируется исследовать свойства признакового пространства аугментированных данных сетевого трафика для повышения качества обучения классификатора и эффективности обнаружения КА в КС в условиях значительного дисбаланса данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Check Point Research. Киберугрозы за 2025 год: исследование мировых кибератак за второй квартал 2025. – URL: https://www.itsec.ru/news/check-point-research-iberataka-za-2025-predstavil-issledovaniye-mirovih-kiberatak-za-vtoroy-kvartal-2025 (дата обращения: 17.03.2025).
2. Positive Technologies. Актуальные киберугрозы: IV квартал 2024 года и I квартал 2025 года. – URL: https://ptsecurity.com/research/analytics/aktualnye-kiberugrozy-iv-kvartal-2024-goda-i-kvartal-2025-goda/ (дата обращения: 17.03.2025).
3. *Шелухин О.И., Сакалема Д.Ж., Филинова А.С.* Обнаружение вторжений в компьютерные сети (сетевые аномалии): учеб. пособие / под ред. О.И. Шелухина. – М.: Горячая линия-Телеком, 2018. – 220 с. – ISBN 978-5-9912-0323-4. – Лань: электронно-библиотечная система. – URL: https://e.lanbook.com/book/111119.
4. *Abdelkhalek A., Mashaly M.* Addressing the class imbalance problem in network intrusion detection systems using data resampling and deep learning // J Supercomput. – 2023. – 79. – P. 10611-10644. – https://doi.org/10.1007/s11227-023-05073-x.
5. *Zhang Y., Muniyandi R.C., Qamar F.* A Review of Deep Learning Applications in Intrusion Detection Systems: Overcoming Challenges in Spatiotemporal Feature Extraction and Data Imbalance // Appl. Sci. – 2025. – 15, 1552. – https://doi.org/10.3390/app15031552.
6. *Leevy J.L., Khoshgoftaar T.M., Bauder R.A. et al.* A survey on addressing high-class imbalance in big data // J Big Data. – 2018. – 5, 42. – https://doi.org/10.1186/s40537-018-0151-6.
7. *Wang Z. et al.* A Comprehensive Survey on Data Augmentation // in IEEE Transactions on Knowledge and Data Engineering. – doi: 10.1109/TKDE.2025.3622600.
8. *Yunhao Chen, Zihui Yan, Yunjie Zhu*. A comprehensive survey for generative data augmentation // Neurocomputing. – 2024. – Vol. 600, 128167. – ISSN 0925-2312. – https://doi.org/10.1016/ j.neucom.2024.128167. https://www.sciencedirect.com/science/article/pii/S092523122400938X.
9. *G. Charbel N. Kindji, Lina M. Rojas-Barahona, Elisa Fromont, Tanguy Urvoy.* Tabular data generation models: An in-depth survey and performance benchmarks with extensive tuning // Neurocomputing. – 2025. – Vol. 658, 131655. – ISSN 0925-2312. – https://doi.org/10.1016/ j.neucom.2025.131655.
10. *Tang B., Lu Y., Li Q., Bai Y., Yu J., Yu, X.* A Diffusion Model Based on Network Intrusion Detection Method for Industrial Cyber-Physical Systems // Sensors. – 2023. – 23, 1141. – https://doi.org/ 10.3390/s23031141.
11. *Yang Y., Tang X., Liu Z., Cheng J., Fang H., and Zhang C.* Diff-IDS: A Network Intrusion Detection Model Based on Diffusion Model for Imbalanced Data Samples // Comput. Mater. Contin. – 2025. – Vol. 82, No. 3. – P. 4389-4408. – https://doi.org/10.32604/cmc.2025.060357.

12. *Saihua Cai, Yingwei Zhao, Jiaao Lyu, Shengran Wang, Yikai Hu, Mengya Cheng, Guofeng Zhang*. DDP-DAR: Network intrusion detection based on denoising diffusion probabilistic model and dual-attention residual network // Neural Networks. – 2025. – Vol. 184, 107064. – ISSN 0893-6080. – https://doi.org/10.1016/j.neunet.2024.107064.
13. *Wang Z., Guan Z., Liu X., Qiao M., Sun X., Li J.* Diffusion–Attention Traffic Generation: Traffic Generation Based on the Fusion of a Diffusion Model and a Self-Attention Mechanism // Electronics. – 2025. – 14, 1977. – https://doi.org/10.3390/electronics14101977.
14. *Котенко И., Саенко И., Лаута О., Крибель А.* Метод раннего обнаружения кибератак на основе интеграции фрактального анализа и статистических методов // Первая миля. – 2021. – № 6 (98). – С. 64-71. – DOI 10.22184/2070-8963.2021.98.6.64.70. – EDN KRIUAD.
15. *Jonathan Ho, Ajay Jain, and Pieter Abbeel.* Denoising diffusion probabilistic models // In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS '20). Curran Associates Inc., Red Hook, NY, USA, 2020. – Article 574. – P. 6840-6851.
16. *Yang Song and Stefano Ermon.* Generative modeling by estimating gradients of the data distribution // Proceedings of the 33rd International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, 2019. – Article 1067. – P. 11918-11930.
17. *Leevy J.L., Khoshgoftaar T.M.* A survey and analysis of intrusion detection models based on CSE-CSE-CIC-IDS2018 Big Data // J Big Data. – 2020. – 7, 104. – https://doi.org/10.1186/s40537-020-00382-x.
18. *Moustafa N., Slay J.* The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set // Inf. Secur. J. A Glob. Perspect. – 2016. – 25. – P. 18-31.
19. *Kadam V., & Verma R.* A Critical Review of Performance Metrics in Intrusion Detection Systems // Journal of Engineering Science and Technology Review. – 2025. – 18 (1). – P. 199-209.
20. *Morehead A., Cheng J.* Geometry-complete diffusion for 3D molecule generation and optimization // Commun Chem. – 2024. – 7, 150. – https://doi.org/10.1038/s42004-024-01233-z.

REFERENCES

1. Check Point Research. Kiberugrozy za 2025 god: issledovanie mirovykh kiberatak za vtoroy kvartal 2025 [Check Point Research. Cyber Threats for 2025: a study of global cyberattacks in second quarter 2025]. Available at: https://www.itsec.ru/news/check-point-research-iberataka-za-2025-predstavil-issledovaniye-mirovih-kiberatak-za-vtoroy-kvartal-2025 (accessed 17 March 2025).
2. Positive Technologies. Aktual'nye kiberugrozy: IV kvartal 2024 goda i I kvartal 2025 goda [Positive Technologies. Current cyber threats: fourth quarter of 2024 and first quarter of 2025. Available at: https://ptsecurity.com/research/analytics/aktualnye-kiberugrozy-iv-kvartal-2024-goda-i-kvartal-2025-goda/ (accessed 17 March 2025).
3. *Shelukhin O.I., Sakalema D.Zh., Filinova A.S.* Obnaruzhenie vtorzheniy v komp'yuternye seti (setevye anomalii): ucheb. posobie [Network intrusion detection (network anomalies): a textbook], ed. by O.I. Shelukhina. Moscow: Goryachaya liniya-Telekom, 2018, 220 p. ISBN 978-5-9912-0323-4. Lan': elektronno-bibliotechnaya sistema. Available at: https://e.lanbook.com/book/111119.
4. *Abdelkhalek A., Mashaly M.* Addressing the class imbalance problem in network intrusion detection systems using data resampling and deep learning, *J Supercomput*, 2023, 79, pp. 10611-10644. Available at: https://doi.org/10.1007/s11227-023-05073-x.
5. *Zhang Y., Muniyandi R.C., Qamar F.* A Review of Deep Learning Applications in Intrusion Detection Systems: Overcoming Challenges in Spatiotemporal Feature Extraction and Data Imbalance, *Appl. Sci.*, 2025, 15, 1552. Available at: https://doi.org/10.3390/app15031552.
6. *Leevy J.L., Khoshgoftaar T.M., Bauder R.A. et al.* A survey on addressing high-class imbalance in big data, *J Big Data*, 2018, 5, 42. Available at: https://doi.org/10.1186/s40537-018-0151-6.
7. *Wang Z. et al.* A Comprehensive Survey on Data Augmentation, *in IEEE Transactions on Knowledge and Data Engineering*. doi: 10.1109/TKDE.2025.3622600.
8. *Yunhao Chen, Zihui Yan, Yunjie Zhu.* A comprehensive survey for generative data augmentation, *Neurocomputing*, 2024, Vol. 600, 128167. ISSN 0925-2312. Available at: https://doi.org/10.1016/ j.neucom.2024.128167. https://www.sciencedirect.com/science/article/pii/S092523122400938X.
9. *G. Charbel N. Kindji, Lina M. Rojas-Barahona, Elisa Fromont, Tanguy Urvoy.* Tabular data generation models: An in-depth survey and performance benchmarks with extensive tuning, *Neurocomputing*, 2025, Vol. 658, 131655. ISSN 0925-2312. Available at: https://doi.org/10.1016/ j.neucom.2025.131655.
10. *Tang B., Lu Y., Li Q., Bai Y., Yu J., Yu, X.* A Diffusion Model Based on Network Intrusion Detection Method for Industrial Cyber-Physical Systems, *Sensors*, 2023, 23, 1141. Available at: https://doi.org/ 10.3390/s23031141.
11. *Yang Y., Tang X., Liu Z., Cheng J., Fang H., and Zhang C*. Diff-IDS: A Network Intrusion Detection Model Based on Diffusion Model for Imbalanced Data Samples, *Comput. Mater. Contin.*, 2025, Vol. 82, No. 3, pp. 4389-4408. Available at: https://doi.org/10.32604/cmc.2025.060357.

12. *Saihua Cai, Yingwei Zhao, Jiaao Lyu, Shengran Wang, Yikai Hu, Mengya Cheng, Guofeng Zhang.* DDP-DAR: Network intrusion detection based on denoising diffusion probabilistic model and dual-attention residual network, *Neural Networks*, 2025, Vol. 184, 107064. ISSN 0893-6080. Available at: https://doi.org/10.1016/j.neunet.2024.107064.
13. *Wang Z., Guan Z., Liu X., Qiao M., Sun X., Li J*. Diffusion–Attention Traffic Generation: Traffic Generation Based on the Fusion of a Diffusion Model and a Self-Attention Mechanism, *Electronics*, 2025, 14, 1977. Available at: https://doi.org/10.3390/electronics14101977.
14. *Kotenko I., Saenko I., Lauta O., Kribel' A.* Metod rannego obnaruzheniya kiberatak na osnove integratsii fraktal'nogo analiza i statisticheskikh metodov [Method of early detection of cyberattacks based on the integration of fractal analysis and statistical methods], *Pervaya milya* [First Mile], 2021, No. 6 (98), pp. 64-71. DOI 10.22184/2070-8963.2021.98.6.64.70. EDN KRIUAD.
15. *Jonathan Ho, Ajay Jain, and Pieter Abbeel.* Denoising diffusion probabilistic models, *In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS '20). Curran Associates Inc., Red Hook, NY, USA, 2020*, Article 574, pp. 6840-6851.
16. *Yang Song and Stefano Ermon.* Generative modeling by estimating gradients of the data distribution, *Proceedings of the 33rd International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, 2019*, Article 1067, pp. 11918-11930.
17. *Leevy J.L., Khoshgoftaar T.M.* A survey and analysis of intrusion detection models based on CSE-CSE-CIC-IDS2018 Big Data, *J Big Data*, 2020, 7, 104. Available at: https://doi.org/10.1186/s40537-020-00382-x.
18. *Moustafa N., Slay J.* The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set, *Inf. Secur. J. A Glob. Perspect*, 2016, 25, pp. 18-31.
19. *Kadam V., & Verma R.* A Critical Review of Performance Metrics in Intrusion Detection Systems, *Journal of Engineering Science and Technology Review*, 2025, 18 (1), pp. 199-209.
20. *Morehead A., Cheng J.* Geometry-complete diffusion for 3D molecule generation and optimization, *Commun Chem*, 2024, 7, 150. Available at: https://doi.org/10.1038/s42004-024-01233-z.

Балыбердин Алексей Викторович – Финансовый университет при Правительстве РФ; e-mail: balyberdinav@gmail.com; г. Москва, Россия; аспирант; https://orcid.org/0009-0005-1264-9290.

Balyberdin Alexey Viktorovich – Financial University under the Government of the Russian Federation; e-mail: balyberdinav@gmail.com; Moscow, Russia; graduate student; https://orcid.org/0009-0005-1264-9290.

УДК 004.056.5 DOI 10.18522/2311-3103-2026-3-18-31

К.С. Дунюшкина, И.В. Машкина

РАЗРАБОТКА МОДЕЛИ УГРОЗ В ИНФОРМАЦИОННОЙ СРЕДЕ ГИПЕРКОНВЕРГЕНТНОЙ ИНФРАСТРУКТУРЫ НА ОСНОВЕ ПОСТРОЕНИЯ СЦЕНАРИЕВ МНОГОКОМПОНЕНТНЫХ АТАК

Целью данного исследования является разработка формализованной модели угроз безопасности информации гиперконвергентной инфраструктуры (Hyper-Converged Infrastructure – HCI) на основе построения сценариев многокомпонентных атак с использованием EPC-моделирования (Event-driven Process Chain (EPC), с учетом архитектурных особенностей гетерогенной HCI, каскадного характера распространения угроз, специфических объектов воздействия угроз и мультитенантной модели доступа. В работе применены метод системного анализа, принцип методологии ARIS (Architecture of Integrated Information Systems), а также метод формализации сценариев многокомпонентных атак. Технической базой послужили: методика ФСТЭК России по оценке угроз безопасности информации, реестр угроз безопасности информации и сведения об уязвимостях из банка данных угроз ФСТЭК и международной базы (NVD). В результате исследования выявлены ключевые архитектурные особенности гиперконвергентной инфраструктуры как объекта защиты: тесная интеграция компонентов (вычислений, хранения, сети), программная определяемость компонентов, мультитенантность, каскадный характер распространения угроз. Предложена классификация нарушителей по уровню привилегий. Разработаны два формализованных сценария многокомпонентных атак на специфические объекты воздействия угроз в HCI. Сценарии представлены в виде EPC-диаграмм, что позволяет не только визуализировать действия нарушителя, наглядно представить используемые им тактики и техники, но и численно оценить вероятности реализации сценариев. На основе исследования структурно-функциональных характеристик HCI и с учетом модели оценки угроз ФСТЭК разработана обобщенная модель угроз гиперконвергентной инфраструктуры с примерами заполнения для некоторых объектов воздействия. Научная новизна исследования