МТУСИ

Шелухин Олег Иванович, д.т.н., профессор
Раковский Дмитрий Игоревич, к.т.н.

РусКрипто’2026

Обнаружение сетевых атак на основе многозначных зависимостей

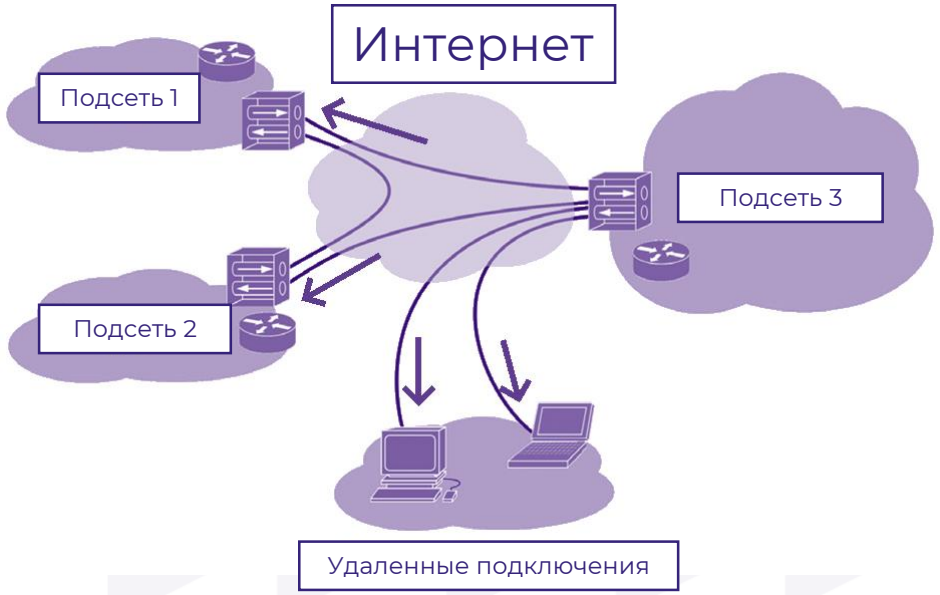
Шелухин Олег Иванович
д.т.н., профессор, заслуженный деятель науки РФ, заведующий кафедрой «Информационная безопасность»

Раковский Дмитрий Игоревич
к.т.н., доцент кафедры «Информационная безопасность»

Москва, 2026

Современные вызовы при создании систем обнаружения вторжений

№	Вызов	Обоснование
1	**Атаки «нулевого дня»**	Одна из главных проблем современных СОВ — невозможность детектирования ранее неизвестных деструктивных воздействий сигнатурными методами
2	**Многовекторные атаки**	Современные атаки используют несколько точек входа и методов одновременно, что требует анализа комбинированных угроз
3	**Высокий объём ложных срабатываний**	Требуется существенная настройка систем для минимизации false positive
4	**Недостаток размеченных данных**	Для обучаемых СОВ требуется обширная база размеченных данных
5	**Редкие аномальные события (PAC)**	Наиболее опасные события статистически редки, что создаёт дисбаланс классов в данных
6	**Шифрование трафика**	Ограниченные возможности анализа зашифрованного сетевого трафика



Контекст ранее проведенных исследований

Многозначная (*multi-label*) зависимость - одновременное соответствие нескольких классовых меток одному объекту.

Сферы, где активно используется свойство многозначности для принятия решений

Работа с текстом	• Тематическая рубрикация документов • Определение эмоциональной окраски текстовых сообщений
Компьютерное зрение	• Одновременное обнаружение нескольких объектов на одном изображении / кадре видео
Работа с профилями в социальных сетях	• Определение интересов и взглядов человека
Медицина	• Диагностика заболеваний • Определение наиболее эффективных лекарств и курса лечения пациента

При работе с целевым атрибутом классических алгоритмов МО допускаются ***ошибки первого и второго рода***, связанные с невозможностью принятия однозначного решения.

Источники возникновения многозначных зависимостей в компьютерных сетях

В следствие вредоносного воздействия	**Коллизии в следствие предобработки данных**
❑ Одновременная реализация КА на хосты КС, обладающие несколькими сетевыми интерфейсами; ❑ **Одновременная реализация КА на КС через несколько сетевых шлюзов;** ❑ **Однозначная КА + накопление ошибок в КС (аномальные состояния)**	❑ Преобразование данных, описывающих многопоточное функционирование КС, в табличную структуру; ❑ **Снижение размерности**; ❑ Передискретизация данных.

Алгоритм «Поиск полных дубликатов по атрибутному пространству»

1. В наборе данных производится поиск дубликатов по атрибутному пространству **A** с исключением целевых атрибутов.
2. Обнаруженные дубликаты группируются, игнорируя целевые атрибуты.
3. Если количество уникальных классовых меток в группе дубликатов >1, все записи группы считаются многозначными закономерностями.
4. Производится подсчет количества обнаруженных многозначных закономерностей.

Наименование авторского набора данных	Кол-во атрибутов в наборе данных, ед.	Кол-во уникальных классовых меток, ед.	Доля мзн. з. (без сокращения размерности)	Доля мзн. з. (10 наиболее информативных атрибутов)	Доля мзн. з. (5 наиболее информативных атрибутов)
SR-BH 2020	24	14	1	1,2 (+0,2%)	1,3 (+0,3%)
UNSW-NB15	45	10	16	18 (+2%)	37 (+21%)
NF-UQ-NIDS	46	21	$5 \cdot 10^{-3}$	42 (+41%)	86 (+85%)

Примеры реализаций многозначных КА

Пример	Примеры атак
Атака типа «Массовый отказ в обслуживании», DDoS Атаки с использованием ботнет-сетей	Атака на Dyn (2016) - DDoS через ботнет Mirai: Атака на DNS-провайдера Dyn с использованием IoT-устройств. Параллельно с DDoS велась фишинговая рассылка для распространения вредоносного ПО. * Атаки с использованием сочетания DDoS и Wiper, например – Hermeticwiper. Wiper-атака на Sony Pictures (2014) с использованием ВПО Destover*
Иные типы атак	Атаки с использованием программных и программно-аппаратных закладок в целевой КС; Эксплуатация КС злоумышленником через backdoor; лавинное распространение ВПО в корпоративной сети.

* *https://krebsonsecurity.com/2016/10/ddos-on-dyn-impacts-twitter-spotify-reddit/*
https://www.graphus.ai/blog/wiper-malware-the-nastiest-cyberthreat-in-town/
* Stephan HaggardJ.R. Lindsay North Korea and the Sony hack: Exporting instability through cyberspace // Analysis from the East-West Center. 2015. №117. С. 1-9.

Примеры реализаций КА с накоплением последствий от их реализации

Пример	Примеры атак
Атаки типа "Low and Slow"	Low and Slow DDoS-атаки имитируют легитимный трафик, перегружая инфраструктуру медленно и целенаправленно Transport for London (TfL) Cyber Attack (2024)*
APT-атаки с долгосрочной компрометацией	Рост 136% в США**
Атаки, направленные на каскадные отказы критической инфраструктуры	Лавинообразное распространение ВПО, ransomware-атаки***

* https://medium.com/@ashu667/the-5-costliest-cybersecurity-incidents-of-2024-2025-lessons-losses-and-prevention-strategies-e27eb11477bf
** https://industrialcyber.co/threat-landscape/rsa-2025-trellix-cyberthreat-reveals-136-surge-in-apt-attacks-on-us-in-q1-2025-as-threat-landscape-intensifies/
*** industrialcyber.co/reports/half-of-2025-ransomware-attacks-hit-critical-sectors-as-manufacturing-healthcare-and-energy-top-global-targets/

Возникновение многозначных зависимостей в следствие коллизий

n	Классические наборы данных табличной структуры						
	A(n,)						Целевой атрибут
	a_{n1}	a_{n2}	...	$a_{n\lambda}$	...	$a_{n\Lambda}$	l
1	a_{11}	a_{12}	...	$a_{1\lambda}$	...	$a_{1\Lambda}$	l_1
2	a_{21}	a_{22}	...	$a_{2\lambda}$	...	$a_{2\Lambda}$	l_2
3	a_{31}	a_{32}	...	$a_{3\lambda}$	...	$a_{3\Lambda}$	l_3
...	...	...	...	...	...	...	...
N-1	$a_{N-1\,1}$	$a_{N-1\,2}$	...	$a_{N-1\,\lambda}$	...	$a_{N-1\,\Lambda}$	l_{N-1}
N	$a_{N\,1}$	$a_{N\,2}$	...	$a_{N\,\lambda}$	...	$a_{N\,\Lambda}$	l_N

Фрагмент базы данных «UNSW-NB15», иллюстрирующий проблему многозначности целевых атрибутов при «классическом» представлении табличной структуры данных при его анализе алгоритмами обучения с учителем

№	srip	proto	state	dur	...	stime	Целевой атрибут
1	175.45.176.2	ospf	INT	0.518061	...	1421927596	Exploits
2	175.45.176.2	ospf	INT	0.518061	...	1421927596	Exploits
3	175.45.176.2	ospf	INT	0.518061	...	1421927596	Reconnaissance
4	175.45.176.2	ospf	INT	0.518061	...	1421927596	Fuzzers
5	175.45.176.2	ospf	INT	0.518061	...	1421927596	Exploits
6	175.45.176.2	ospf	INT	0.518061	...	1421927596	Exploits
7	175.45.176.2	ospf	INT	0.518061	...	1421927596	Reconnaissance
8	175.45.176.2	ospf	INT	0.518061	...	1421927596	Fuzzers

Реализованные и апробированные результаты, являющиеся предпосылками к текущему исследованию

№	Результат
1	Исследование существующих наборов данных на наличие «скрытых» многозначных зависимостей
2	Модель табличного представления поведения КС, учитывающая многозначные зависимости
3	Разработка метода многозначной классификации компьютерных атак, алгоритма и архитектуры нейронной сети на ее основе и его апробация
4	Разработка программно-аппаратного комплекса для сбора телеметрии и имитационного моделирования компьютерных атак

Результаты получены для достижения цели работы: обеспечение ИБ путем повышения точности классификации сетевых атак в КС в условиях многозначности целевых атрибутов, маркирующих их тип.

Модель табличного представления поведения КС

$$\boldsymbol{D}_{NM} = \{(\boldsymbol{A}(n,\ ), \boldsymbol{L}(n,\ ));\ \boldsymbol{A} = (a_{n\lambda}), \boldsymbol{L} = (l_{n\xi}), \lambda = \overline{1,\Lambda}, \xi = \overline{1,\Xi}, n = \overline{1,N}, M = \Lambda + \Xi\}$$

- $\boldsymbol{A}(n,\ ) = (a_{n1}, a_{n2}, \ \dots\ , a_{n\Lambda})$ - n-ный вектор-строка матрицы атрибутов экспериментальных данных A, состоящая из Λ столбцов; $a_{ni} \in \boldsymbol{A}(n,\ )$ - метрическое значение λ-го атрибута на n-ой строке $\boldsymbol{D}_{NM}$;
- $\boldsymbol{L}(n,\ ) = (l_{n1}, l_{n2}, \ \dots\ , l_{n\Xi})$ - *n*-ый вектор-строка матрицы целевых атрибутов экспериментальных данных $\boldsymbol{L}$, состоящий из Ξ столбцов; $l_{n\xi} \in \boldsymbol{L}(n,\ )$ - ξ-я классовая метка на n-ой строке $\boldsymbol{D}_{NM}$; $l_{n\xi} \in \{0,1\}$.
- *N* – количество записей $\boldsymbol{D}_{NM}$; *M* – совокупное количество атрибутов в $\boldsymbol{D}_{NM}$; $M = \Lambda + \Xi$.

n	$\boldsymbol{D}_{NM}$											
	$\boldsymbol{A}(n,\ )$						$\boldsymbol{L}(n,\ )$					
	a_{n1}	a_{n2}	…	$a_{n\lambda}$	…	$a_{n\Lambda}$	l_{n1}	l_{n2}	…	$l_{n\xi}$	…	$l_{n\Xi}$
1	a_{11}	a_{12}	…	$a_{1\lambda}$	…	$a_{1\Lambda}$	l_{11}	l_{12}	…	$l_{1\xi}$	…	$l_{1\Xi}$
2	a_{21}	a_{22}	…	$a_{2\lambda}$	…	$a_{2\Lambda}$	l_{21}	l_{22}	…	$l_{2\xi}$	…	$l_{2\Xi}$
3	a_{31}	a_{32}	…	$a_{3\lambda}$	…	$a_{3\Lambda}$	l_{31}	l_{32}	…	$l_{3\xi}$	…	$l_{3\Xi}$
…	…	…	…	…	…	…	…	…	…	…	…	…
N-1	$a_{N-1\,1}$	$a_{N-1\,2}$	…	$a_{N-1\,\lambda}$	…	$a_{N-1\,\Lambda}$	$l_{N-1\,1}$	$l_{N-1\,2}$	…	$l_{N-1\,\xi}$	…	$l_{N-1\,\Xi}$
N	$a_{N\,1}$	$a_{N\,2}$	…	$a_{N\,\lambda}$	…	$a_{N\,\Lambda}$	$l_{N\,1}$	$l_{N\,2}$	…	$l_{N\,\xi}$	…	$l_{N\,\Xi}$

Модель требует предобработки данных алгоритмом «Поиска полных дубликатов по атрибутному пространству» (***слайд 4***).

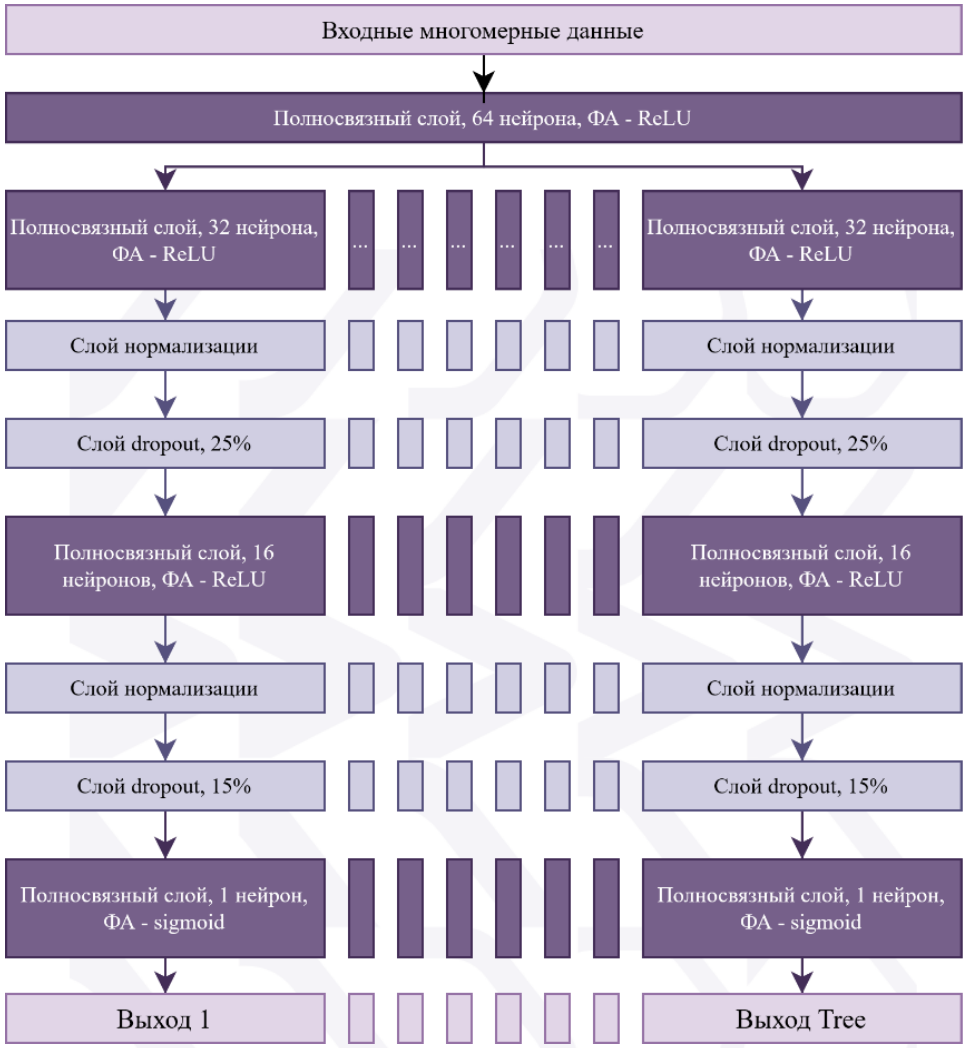
Метод и алгоритм многозначной классификации

Исходные данные преобразуются согласно предложенной модели

Для каждой $\hat{n}$ записи преобразованных данных:

- Запись дублируется Ξ раз, где Ξ – количество уникальных КА
- Для каждого дубля проводится последовательное нелинейное отображение в значение соответствующего целевого атрибута

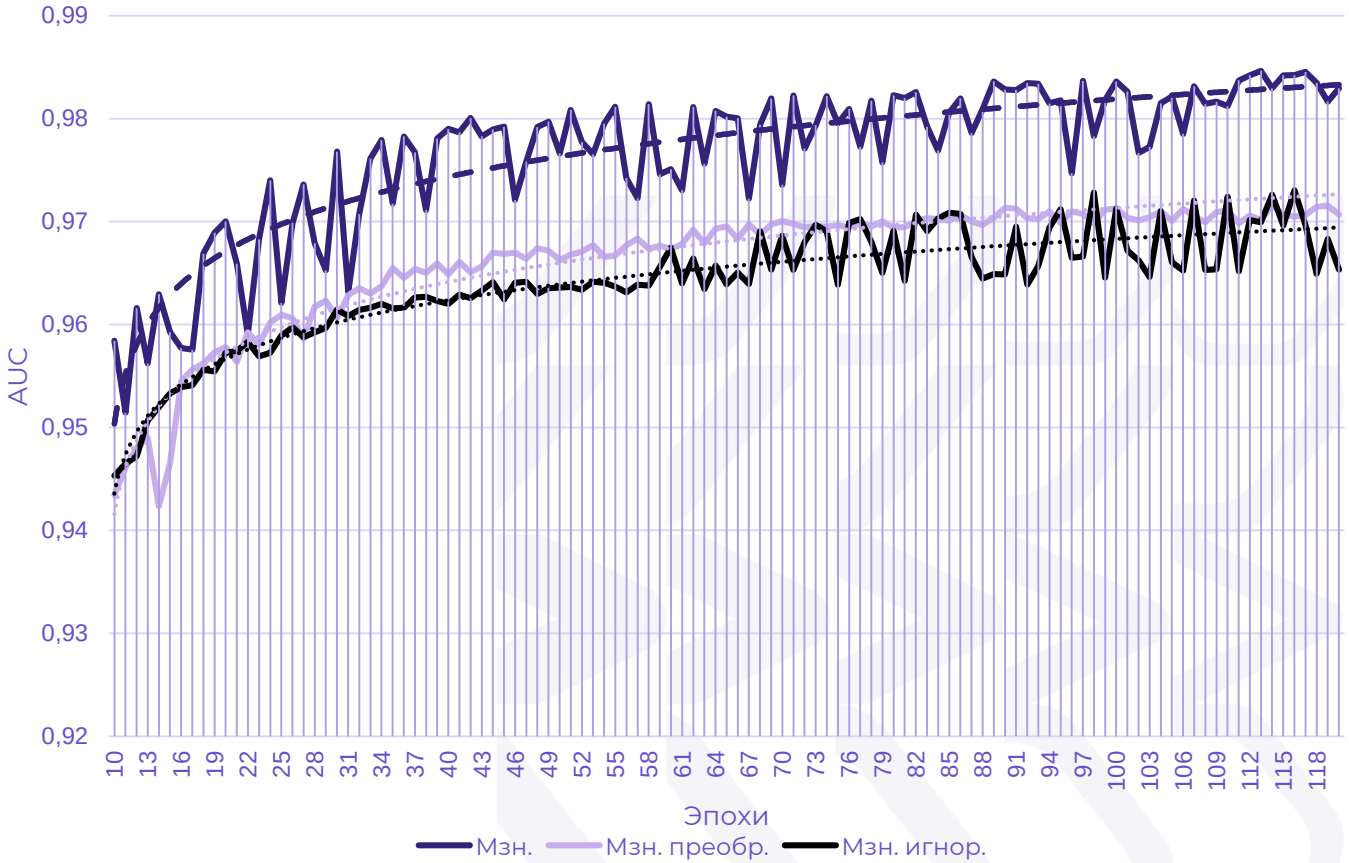
Результатом работы метода является вектор решений по каждому из Ξ КА, где «1» означает соответствие ξ-й КА $\hat{n}$-й записи, а «0» - ее отсутствие



Апробация АМК на наборе данных SR-BH 2020

Алгоритм	AUC_{ovo}
ИНС с множественным выходом (120 эпох, <мзн.>)	**0,983**
ИНС с множественным выходом (120 эпох, <мзн. преобр.>)	**0,971**
ИНС с множественным выходом (120 эпох, <мзн. игнор.>)	**0,965**
CatBoost с множественным выходом	0,913
LightGBM с множественным выходом	0,911
Одноступенчаный Clas.Chain LightGBM, учитывающий многозначность методом цепочек классификации	0,904
Одноступенчаный CatBoost, учитывающий многозначность методом цепочек классификации	0,902
Одноступенчаный CatBoost, учитывающий многозначность методом бинарной релевантности	0,901
Двуступенчатый CatBoost, учитывающий многозначность методом бинарной релевантности	0,9
Одноступенчаный LightGBM, учитывающий многозначность методом бинарной релевантности	0,898
Двуступенчатый LightGBM, учитывающий многозначность методом бинарной релевантности	0,898

Объем, ед.	918.176	**Типов КА, ед.**	14



Выигрыш на 10% за счет применения предложенной модели и учета нелинейных связей в АМК перед ансамблевыми алгоритмами многозначной классификации.

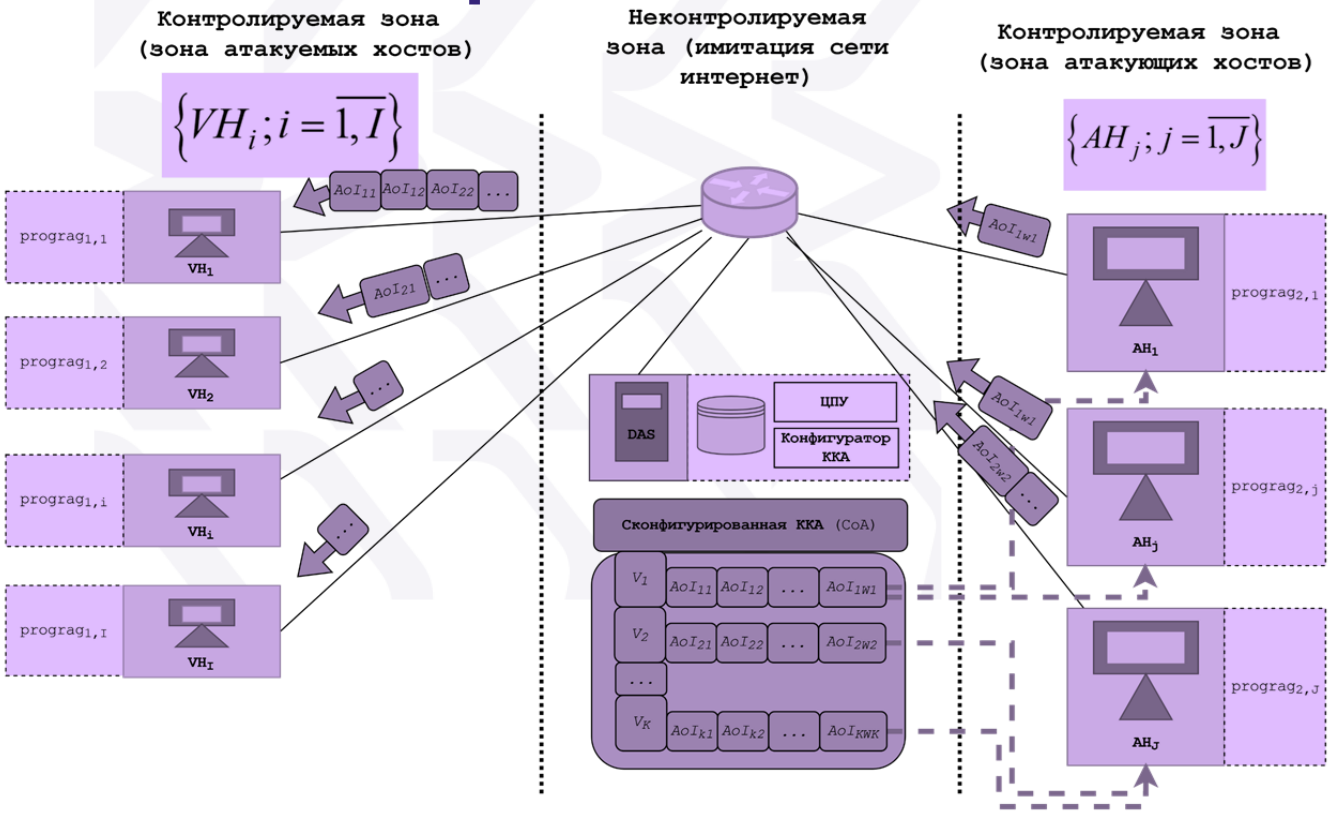
Апробация АМК на наборе данных UNSW-NB15

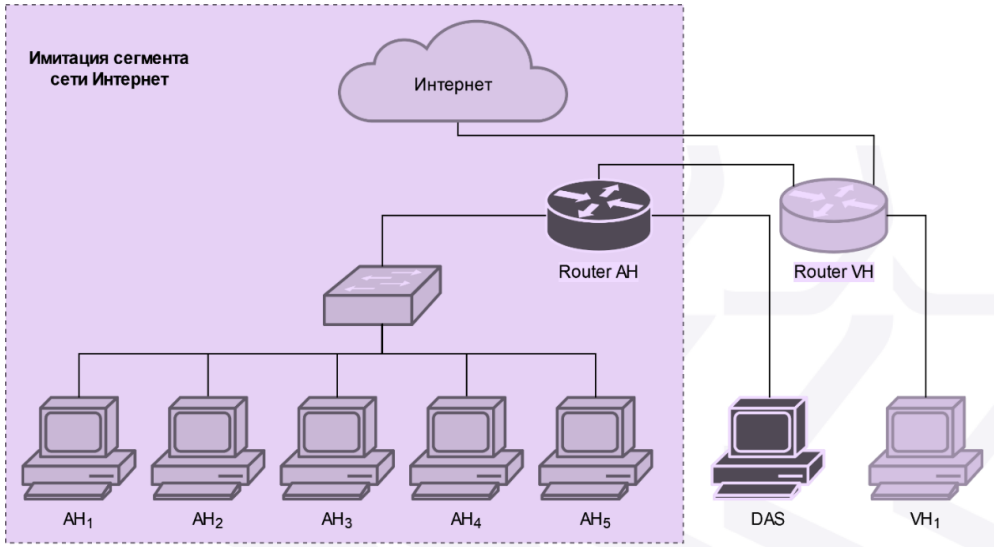
Объем, ед.	Типов КА, ед.
257.674	10

Алгоритм	Все атрибуты КС	10 наиболее значимых атрибутов КС	5 наиболее значимых атрибутов КС
	AUC_{ovo}	AUC_{ovo}	AUC_{ovo}
ИНС с множественным выходом (120 эпох, <мзн.>)	**0,95**	**0,93**	**0,90**
ИНС без множественного выхода (120 эпох, <мзн.>)	0,93	0,91	0,88
DecisionTree <мзн., модель>	0,93	0,92	0,87
ExtraTree <мзн., модель>	0,92	0,89	0,88
ExtraTrees <мзн., модель>	0,91	0,90	0,88
RandomForest <мзн., модель>	0,89	0,89	0,87
DecisionTree <мзн. игнор.>	0,84	0,83	0,76
ExtraTree <мзн. игнор.>	0,82	0,75	0,74
ExtraTrees <мзн. игнор.>	0,78	0,78	0,75
RandomForest мзн. игнор.>	0,77	0,83	0,76

Выигрыш до 13% за счет применения предложенной модели у всех алгоритмов и еще на 3% за счет учета нелинейных связей между атрибутным пространством и целевыми атрибутами в АМК

ПАК для сбора телеметрии и имитационного моделирования КА





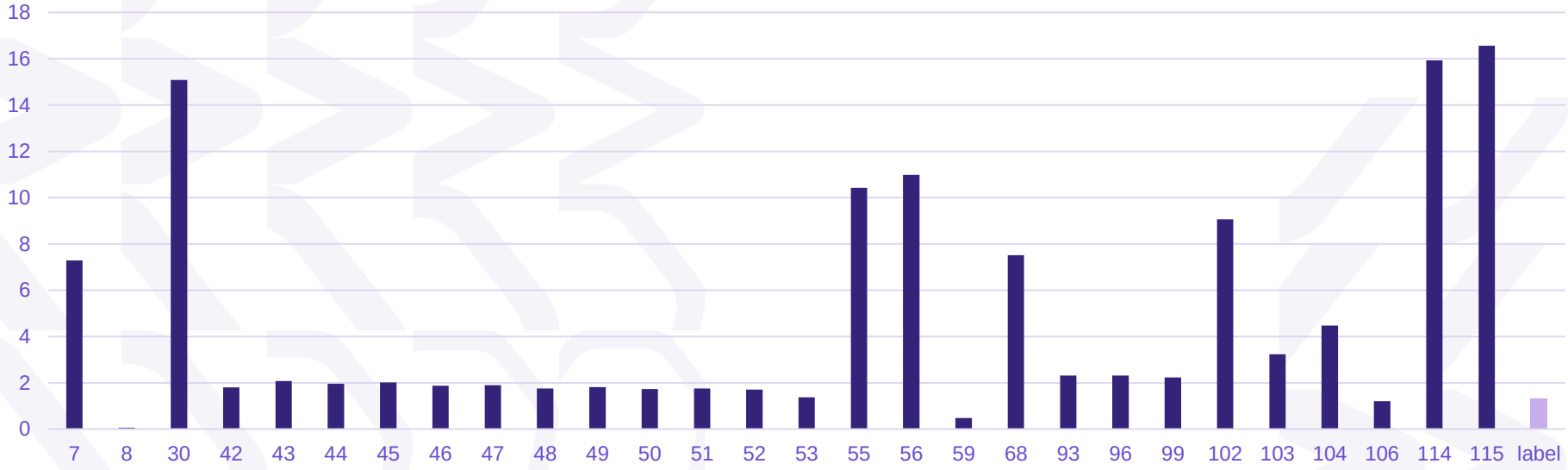
$$T = \{VH_i; i = \overline{1,I}\} \cup \{AH_j; j = \overline{1,J}\} \cup DAS \cup Router$$
$$AL = \{attack_k; k = \overline{1,K}\}.$$
$$AoI_k: \langle IS, IE, attack_k, ah, tar, dur, int, etcp \rangle$$
$$\vec{V}_k = (AoI_{kw}; w = \overline{1,W_k}), CoA = (\vec{V}_k; k = \overline{1,K})$$

С использованием ПАК собран набор данных

Количество записей экспериментальных данных в итоговом наборе составило **263 388** ед. по 125 атрибутам

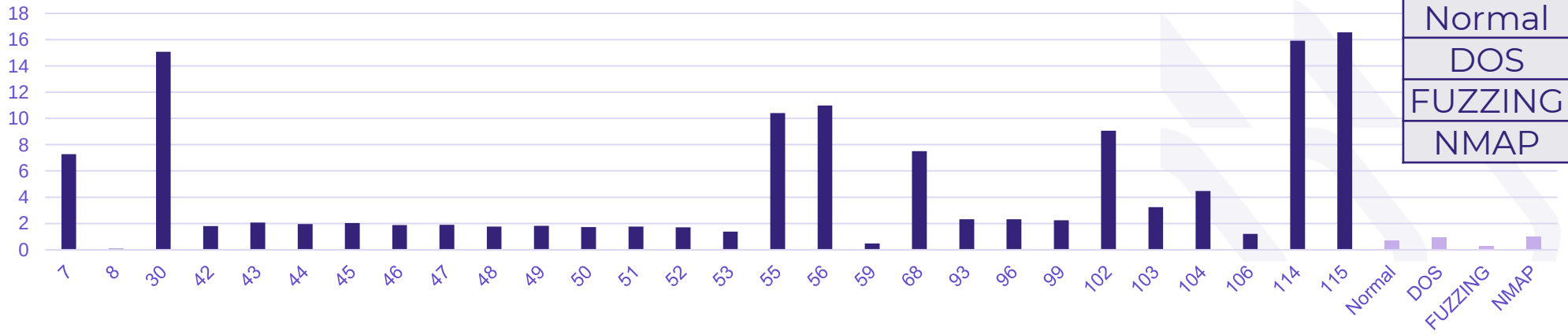
Обнаружение синергического эффекта от учета многозначных зависимостей

Многоклассовое представление (<мзн. игнор.>), энтропия каждого атрибута, H(a)



	H(a)
label	**1,33**

Многозначное представление, энтропия каждого атрибута, H(a)



	H(a)	sum
Normal	0,713708	**2,86**
DOS	0,916397	
FUZZING	0,244767	
NMAP	0,990095	

Связь многозначных зависимостей с редкими аномальными событиями в первом приближении

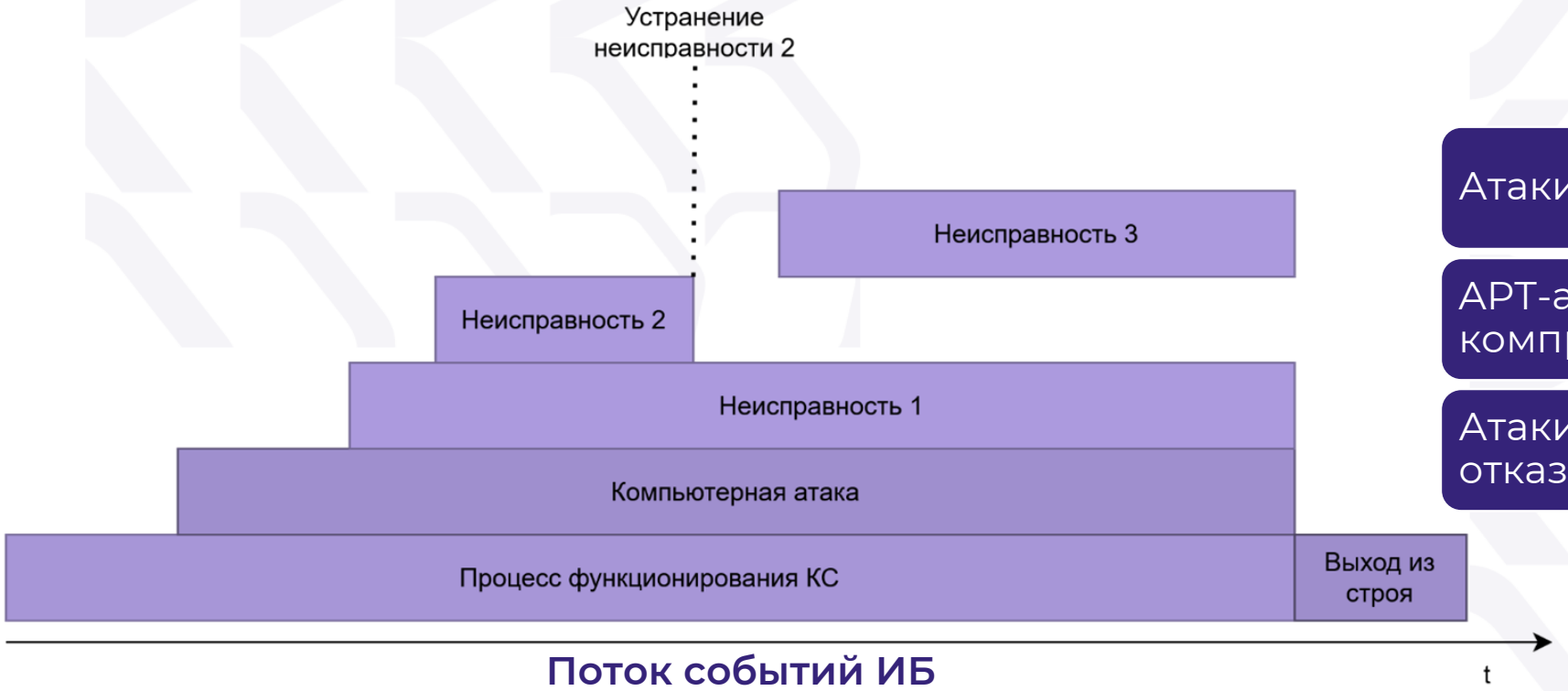
Пусть задано универсальное множество событий КС D, наблюдаемых в процессе функционирования системы. Введём следующие подмножества:

- $D_M \subset D$ - подмножество многозначных событий, для которых выполняется условие: $\forall d \in D_M: |L(d) \geq 2|$, где $L(d) \subseteq \mathcal{L}$ - множество классовых меток, ассоциированных с событием d, $\mathcal{L}$ - алфавит уникальных типов КА.
- $D_R \subset D$ - подмножество редких событий, определяемое через вероятностную меру $\mathbb{P}$: $\forall d \in D_R: \mathbb{P}(d) < \varepsilon, \varepsilon \leq 0{,}01$, где ε - порог редкости, установленный согласно рубрикатору редкости событий.
- $D_A \subset D$ - подмножество аномальных (деструктивных) событий, семантически определяемое в предметной области информационной безопасности как события, нарушающие свойства доступности, целостности или конфиденциальности информации.

Тогда **редкое аномальное событие** формально определяется как пересечение трёх множеств:

$$RAE = D_M \cap D_R \cap D_A$$

Проявление редких аномальных событий в многозначной трактовке



$$RAE = D_M \cap D_R \cap D_A$$

- Атаки типа "Low and Slow«
- APT-атаки с долгосрочной компрометацией
- Атаки, направленные на каскадные отказы критической инфраструктуры

Ключевое ограничение текущих результатов и методы их устранения

Ограничение: метод и алгоритм многозначной классификации относятся к области контролируемого обучения - что не позволяет обнаруживать атаки нулевого дня.
Компромисс: ИНС на базе автокодировщиков.

Для проверки идеи выдвигаются две гипотезы:

1. принципиально возможно ли обнаружение многозначных событий в наборах данных, связанных с функционированием КС, без знания классовой метки?
2. (***если подтверждается гипотеза 1***) возможно ли выявление аномалий в низкоразмерном пространстве скрытого слоя автокодировщиков, связанных с проявлением многозначных зависимостей в исходном атрибутном пространстве?

Ключевой вопрос, который решается с подтверждением первой гипотезы: возможно ли принципиально отделить многозначные КА от однозначных атак без знания меток классов.

Гипотеза H_1

Определим ошибку реконструкции для каждого объекта как среднеквадратичное отклонения

$$e_n = \sqrt{\frac{1}{\lambda}\sum_{i=1}^{N}\|a_n - \hat{a}_n\|}^2$$

Сведем ее в вектор ошибок реконструкции: $E = (e_n), n = \overline{1,N}$. Матрица многозначных меток классов, где Ξ – количество типов КА (целевых атрибутов), $\Xi \geq 1$:

$$L = \left(l_{n\xi}\right) \in \{0,1\}^{N\times\Xi}$$

Количество одновременных атак для n-го объекта:

$$k_n = \sum_{\xi}^{\Xi} l_{n\xi}$$

объединим в вектор $K = (k_n), n = \overline{1,N}$.

Разобьём выборку на три непересекающихся подмножества:

$$G_0 = (n \in \overline{1,N} | k_n = 0)\ G_1 = (n \in \overline{1,N} | k_n = 1)\ G_{\geq 2} = (n \in \overline{1,N} | k_n \geq 2)$$

Обозначим эмпирические функции распределения ошибок реконструкции для каждой группы:

$$\hat{F}_0(e) = \frac{1}{|G_0|}\sum_{n\in G_0} Ind(e_n \leq e)\ \hat{F}_1(e) = \frac{1}{|G_1|}\sum_{n\in G_1} Ind(e_n \leq e)\ \hat{F}_{\geq 2}(e) = \frac{1}{|G_{\geq 2}|}\sum_{n\in G_{\geq 2}} Ind(e_n \leq e)$$

Гипотеза H_1

Гипотеза H1 (принципиальная возможность обнаружения многозначных атак) принимается, если выполняются следующие условия:

Условие 1. Распределение $\hat{F}_{\geq 2}$ образует отдельный «хвост» с доминированием больших значений ошибок по сравнению с $\hat{F}_0$, $\hat{F}_1$.

Условие 2. Статистический критерий Колмогорова-Смирнова

$$D_{KS}\left(\hat{F}_1, \hat{F}_{\geq 2}\right) = \sup_{e \in R} \left|\hat{F}_0(e) - \hat{F}_{\geq 2}(e)\right| > D_\alpha,$$

$$D_{KS}\left(\hat{F}_0, \hat{F}_{\geq 2}\right) = \sup_{e \in R} \left|\hat{F}_0(e) - \hat{F}_{\geq 2}(e)\right| > D_\alpha$$

$$D_\alpha = \frac{0.05}{3} = 0.0167$$

Условие 3. Критерий взаимной информации

$$I(E,K) = \sum_{e \in \varepsilon} \sum_{k \in \kappa} \hat{p}(e,k) \log \frac{\hat{p}(e,k)}{\hat{p}(e)\hat{p}(k)} > \varepsilon, \varepsilon = 0.1 \text{ бит}$$

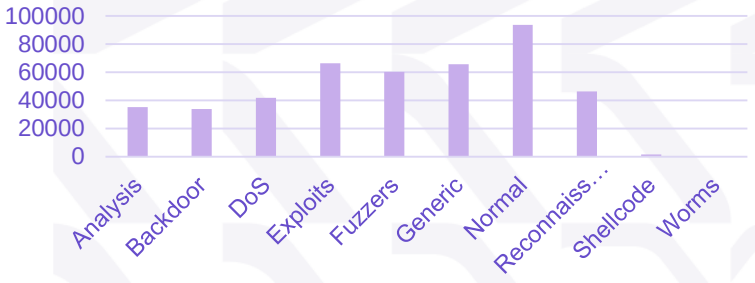
$\hat{p}(e,k)$ - оценка совместной вероятности ошибки реконструкции e и количества атак k.

Если все 3 условия выполняются, то **гипотеза H_1 подтверждается**: ошибки реконструкции автокодировщика содержат статистически значимую информацию о количестве одновременных атак, что позволяет принципиально выявлять многозначные события без использования меток классов.

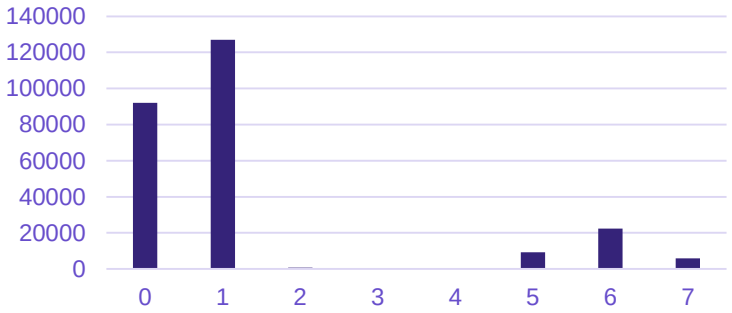
Детализация эксперимента для проверки H_1

UNSW-NB15	
Объем, ед.	Типов КА, ед.
257.674	10

Проведен поиск полных дубликатов по атрибутному пространству (Слайд 4)

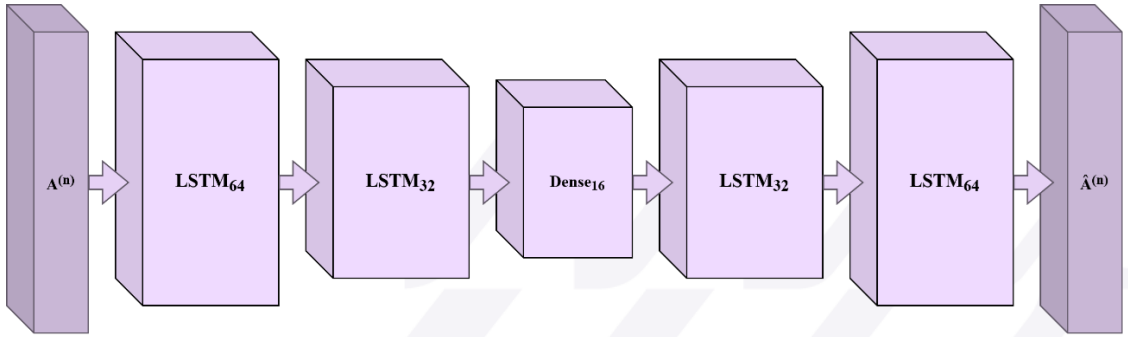


Распределение классовых меток после выявления всех скрытых многозначных закономерностей



Распределение количества классовых меток, присваиваемых каждой записи

Эксперимент №1. Разделение проведено без перемешивания для сохранения темпоральных связей между записями.



Условие 1. Распределение $\hat{F}_{\geq 2}$ образует отдельный «хвост» с доминированием больших значений ошибок по сравнению с $\hat{F}_0, \hat{F}_1$.

Условие 2. Статистический критерий Колмогорова-Смирнова

$$D_{KS}(\hat{F}_1, \hat{F}_{\geq 2}) = \sup_{e \in R} |\hat{F}_0(e) - \hat{F}_{\geq 2}(e)| > D_\alpha,$$

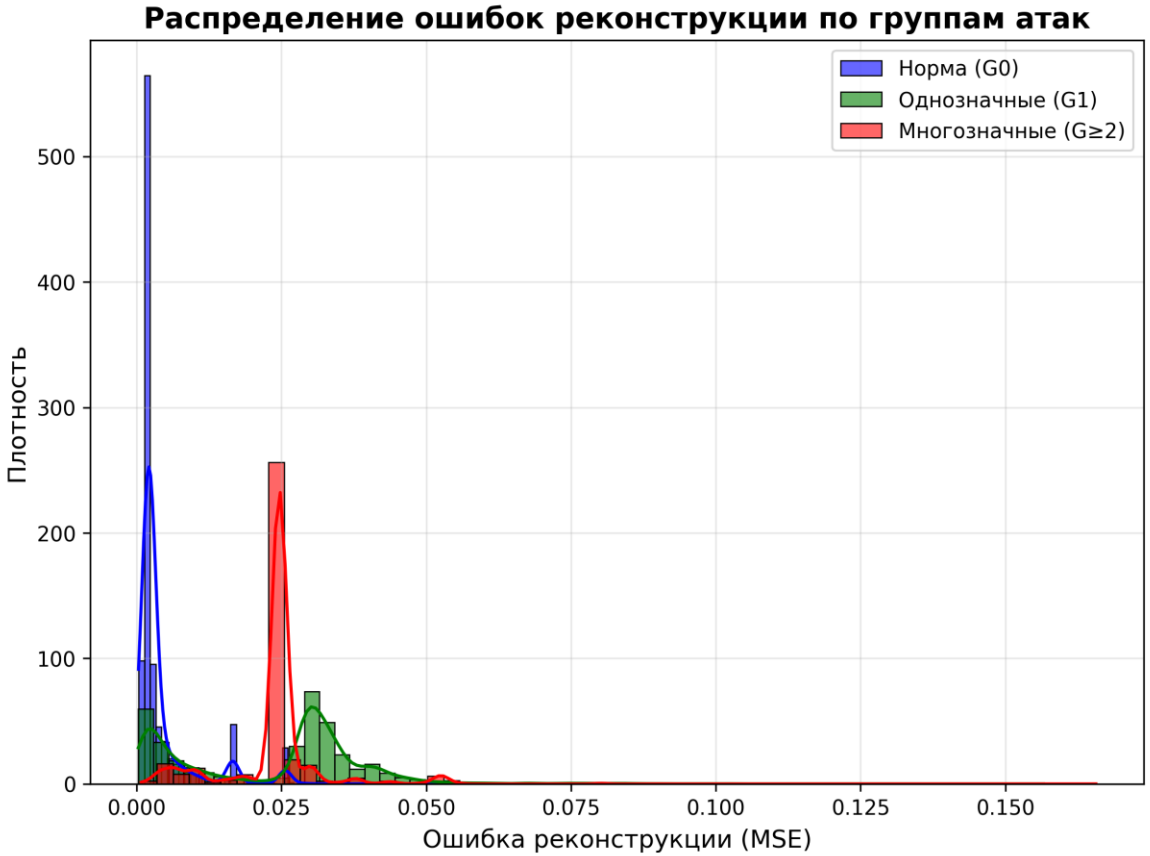
$$D_{KS}(\hat{F}_0, \hat{F}_{\geq 2}) = \sup_{e \in R} |\hat{F}_0(e) - \hat{F}_{\geq 2}(e)| > D_\alpha$$

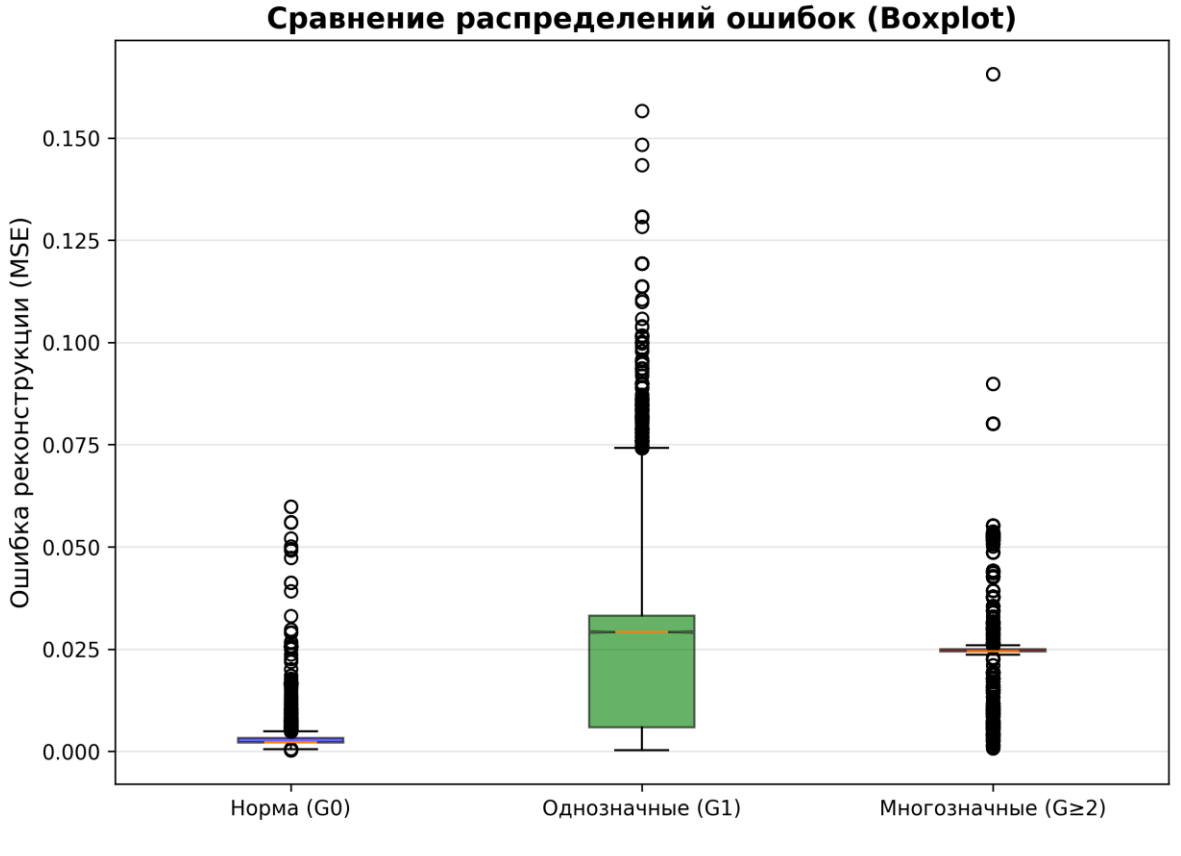
Условие 3. Критерий взаимной информации

$$I(E,K) = \sum_{e \in \varepsilon} \sum_{k \in \kappa} \hat{p}(e,k) \log \frac{\hat{p}(e,k)}{\hat{p}(e)\hat{p}(k)} > \varepsilon$$

В дополнение, выполнялся точный тест Фишера для проверки связи между количеством атак и ошибкой реконструкции автокодировщика.

Результаты проверки критерия Колмогорова-Смирнова





Пара сравниваемых критериев	Результат теста Колмогорова-Смирнова,	p_{value}
G_0 vs G_1	0.6	$< 10^{-6}$
G_0 vs $G_{\geq 2}$	0.84	$< 10^{-6}$
G_1 vs $G_{\geq 2}$	0.5	$< 10^{-6}$

Параметр	Значение
Взаимная информация	0.354

Результаты эксперимента – точный тест Фишера

Направление дальнейших исследований

Апробация первой гипотезы на иных наборах данных. Планируется уточнить на более обширном множестве наборов данных, связанных с событиями информационной безопасности во время функционирования КС.

Формулирование и проверка гипотезы №2. Вторая гипотеза, имеет фундаментальное значение для теории обнаружения РАС, утверждала возможность выявления аномалий непосредственно в низкоразмерном латентном пространстве автокодировщика, где должны проявляться нелинейные корреляции, отражающие синергетический эффект многозначных зависимостей в исходном атрибутном пространстве. *Её подтверждение позволит обосновать найденный эффект на слайде 18.*

Модификация определения РАС вводом временной оси:

$$RAE = D_M \cap D_R \cap D_A$$

Учёт временной эволюции атрибутов позволит выявлять аномальные траектории в пространстве состояний системы, что существенно повышает чувствительность системы обнаружения к сложным многоэтапным сценариям.

Выводы

1. Проведен анализ специфических особенностей многозначной классификации в задачах обнаружения и классификации КА. Доля многозначных зависимостей составила от 1,3% до 86%.
2. Разработаны модель табличного представления профилей функционирования КС, учитывающую многозначность целевых атрибутов, связанных с реализацией КА. Разработаны и апробированы многозначный метод обнаружения и классификации КА в КС, заключающийся в многозначном отображении пространства атрибутов в множество классовых меток, и алгоритм на его основе, выигрыш в точности от которого достигает 13% по метрике AUC_{ovo}, 6% по метрике AUC_{ovo} перед многозначными алгоритмами классификации.
3. Автокодировщики демонстрируют **принципиальную применимость** для выявления многозначных атак и РАС через анализ ошибок реконструкции. Полученные результаты обосновывают необходимость дальнейшего исследования возможности обнаружения аномалий в латентном пространстве автокодировщиков, особенно в контексте выявления нелинейно скоррелированных событий и синергетических эффектов многозначных зависимостей в исходном атрибутном пространстве.
    1. Критерий Колмогорова-Смирнова выявил статистически значимые различия между распределениями ошибок реконструкции всех групп, включая критическое сравнение однозначных и многозначных атак ($D \approx 0.5$; $p < 10^{-6}$ с поправкой Бонферрони), что свидетельствует о принципиальной разделимости данных по количеству одновременных атак.
    2. Взаимная информация между ошибкой реконструкции и количеством атак составила $I \approx 0.39$ бит, что подтверждает наличие статистически значимой связи между латентными представлениями и многозначностью событий.
    3. Точный тест Фишера подтвердил статистически значимую связь ($p<0.05$) между нормальным трафиком и всеми типами атак. Отсутствие значимых различий между однозначными и многозначными атаками объясняется не отсутствием различий, а экстремальным классовым дисбалансом - группы с 2–4 метками составляют менее 0.5% выборки (247, 57 и 63 записи соответственно), что делает применение точного теста Фишера статистически некорректным для этих подгрупп.

БЛАГОДАРЮ ЗА ВНИМАНИЕ!

Контакты

Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики»

111024, г. Москва, ул. Авиамоторная, 8А, URL: https://mtuci.ru

Кафедра: Информационная безопасность (ИБ),
URL: https://mtuci.ru/about_the_university/structure/769/
+7(495)957-77-99 (доб.136), **ib@mtuci.ru**

Шелухин Олег Иванович
д.т.н., профессор, заслуженный деятель науки РФ, заведующий кафедрой «Информационная безопасность»,
sheluhin@mail.ru

Раковский Дмитрий Игоревич
к.т.н., доцент кафедры «Информационная безопасность», **d.i.rakovskiy@mtuci.ru**

Список наиболее значимых публикаций авторов по теме исследования

- Шелухин О.И., Раковский Д.И. Многозначная классификация и прогнозирование М.: Горячая линия - Телеком, 2026. 296 с.
- Шелухин, О. И. Редкие аномальные события. проблемы обнаружения и обработки / О. И. Шелухин, Д. И. Раковский // Наукоемкие технологии в космических исследованиях Земли. – 2025. – Т. 17, № 4. – С. 54-68. – DOI 10.36724/2409-5419-2025-17-4-54-68. – EDN GDQHTU.
- Шелухин, О. И. Разработка программно-аппаратного комплекса моделирования многозначных компьютерных атак / О. И. Шелухин, Д. И. Раковский // Вопросы кибербезопасности. – 2024. – № 4(62). – С. 116-130. – DOI 10.21681/2311-3456-2024-4-116-130. – EDN FIWNMU.
- Шелухин, О. И. Многозначная классификация компьютерных атак с использованием искусственных нейронных сетей с множественным выходом / О. И. Шелухин, Д. И. Раковский // Труды учебных заведений связи. – 2023. – Т. 9, № 4. – С. 97-113. – DOI 10.31854/1813-324X-2023-9-4-97-113. – EDN KVBUQJ.
- Sheluhin, O. I. New Algorithm for Predicting the States of a Computer Network Using Multivalued Dependencies / O. I. Sheluhin, A. V. Osin, D. I. Rakovsky // Automatic Control and Computer Sciences. – 2023. – Vol. 57, No. 1. – P. 48-60. – DOI 10.3103/s0146411623010091. – EDN FUIVBX.