ТЕХНИКАЛЫҚ ҒЫЛЫМДАР
ТЕХНИЧЕСКИЕ НАУКИ
TECHNICAL SCIENCE

CENTRAL ASIAN SCIENTIFIC JOURNAL
ISSN 3135-3061
2026. №4 (32), Том 2. С. 354–361
DOI: https://doi.org/10.5281/zenodo.24252417
Сайт журнала: https://cajournal.kz

УДК 004.056.53:004.85

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ОБНАРУЖЕНИЯ КИБЕРУГРОЗ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

В.В. Ким, Д.А. Хон, А.С. Кадыр, В.В. Цхай
Satbayev University, Алматы, Казахстан

Ким Владислав Вячеславович — студент 4-го курса образовательной программы, «Информационная безопасность», Институт автоматики и информационных технологий, Satbayev University, Алматы, Казахстан.

Хон Дмитрий Анатольевич — студент 4-го курса образовательной программы, «Информационная безопасность», Институт автоматики и информационных технологий, Satbayev University, Алматы, Казахстан.

Кадыр Ануар Сакенович — студент 4-го курса образовательной программы, «Информационная безопасность», Институт автоматики и информационных технологий, Satbayev University, Алматы, Казахстан.

Цхай Владимир Владиславович — студент 4-го курса образовательной программы, «Информационная безопасность», Институт автоматики и информационных технологий, Satbayev University, Алматы, Казахстан.

© В.В. Ким, Д.А. Хон, А.С. Кадыр, В.В. Цхай, 2026

Аннотация: Сравнение конфигураций систем обнаружения вторжений обычно сводится к одной итоговой оценке на одном наборе данных, что не позволяет разделить вклад проектных решений и выявить их взаимодействие. В работе протокол полного факторного эксперимента с повторениями, ранее проверенный авторами на синтетическом отклике, применён к реально измеренной F-мере гибридной системы обнаружения на наборе NSL-KDD: варьировались тип детектора, дообучение на данных защищаемого потока и условия трафика (известное распределение либо сдвиг с 17 новыми типами атак), всего 40 прогонов. Установлено, что на реальных данных дисперсии неоднородны (критерий Кочрена 0,858 при критическом 0,391), однородность восстанавливается логарифмированием недобора F-меры, а значимы все восемь эффектов модели, включая тройное взаимодействие. Преимущество ансамбля меняет знак в зависимости от условий: при известном распределении он уступает одиночной модели 1,7 п. п., при сдвиге без дообучения превосходит её на 8,8 п. п., а после дообучения на 5 % потока разница становится несущественной для практики (0,4 п. п.). Индекс стабильности популяции не превысил порог переобучения 0,25 ни по одному признаку, тогда как многомерный доменный классификатор обнаружил сдвиг с площадью под ROC-кривой 0,905.

Ключевые слова: обнаружение вторжений, машинное обучение, планирование эксперимента, полный факторный эксперимент, дрейф данных, ансамбль моделей, воспроизводимость, NSL-KDD.

Введение

Машинное обучение стало основным инструментом обнаружения сетевых вторжений, а объём публикаций достаточен для систематических обзоров [1]. На открытых эталонных наборах полнота выше 0,98 и доля ложных срабатываний ниже 1 % давно стали обычными,

однако в эксплуатации такие значения не воспроизводятся. В работе [2] показано, что задача обнаружения отличается от типовой классификации: выборка крайне несбалансированна, цена ошибок несимметрична, а представление о нормальном поведении неустойчиво во времени. Систематическое завышение публикуемых показателей подтверждено в реальных корпоративных сетях [3], а в [4] установлено, что модель остаётся состоятельной лишь при обучении на данных той сети, которую она защищает. Для Республики Казахстан проблема имеет и нормативное измерение: владельцы критически важных объектов информатизации обязаны вести непрерывный мониторинг событий информационной безопасности [5, 6], что без автоматизированного анализа практически неосуществимо.

Часть причин связана с тем, как проводится сравнение. Типичная работа приводит одну итоговую оценку для каждой конфигурации на одном наборе данных. Такая схема не содержит оценки собственной погрешности, и разницу в один процентный пункт нельзя отличить от разброса между запусками. Она смешивает проектные решения, поскольку конфигурации обычно различаются сразу в нескольких отношениях. Наконец, она не видит взаимодействий: если выигрыш от решения зависит от условий эксплуатации, оценка, полученная при одних условиях, не переносится на другие. Для устранения этих недостатков авторами в препринте [7] предложен протокол сравнения на основе полного факторного эксперимента с повторениями и полной статистической обработкой. Протокол проверен на отклике, сгенерированном по модели с заранее известными коэффициентами, и восстановил их в пределах доверительных интервалов. Однако такая проверка устанавливает свойства протокола, а не свойства системы обнаружения, и вопрос о его применимости к реальным измерениям оставался открытым.

Направление развивается от классических классификаторов к глубоким архитектурам и ансамблям. В системе Kitsune [8] ансамбль автокодировщиков обучается без учителя непосредственно в защищаемой сети, а ошибка реконструкции служит признаком аномалии.

Цель настоящей работы — применить протокол факторного эксперимента к реально измеренной F-мере гибридной системы обнаружения и установить, как тип детектора, дообучение на данных защищаемого потока и сдвиг распределения трафика совместно влияют на качество обнаружения. Для этого решаются задачи: описать архитектуру системы, задающую варьируемые факторы; провести эксперимент 2^3 с пятью повторами; выполнить статистическую обработку и проверить её предпосылки на реальных данных; оценить пригодность индекса стабильности популяции для контроля дрейфа.

Материалы и методы

Объект исследования. Функциональная структура разрабатываемой системы приведена на рисунке 1. Она включает четыре уровня: сбор и нормализацию событий, формирование признаков, ансамбль детекторов, а также объяснение вердикта, реагирование и дообучение.

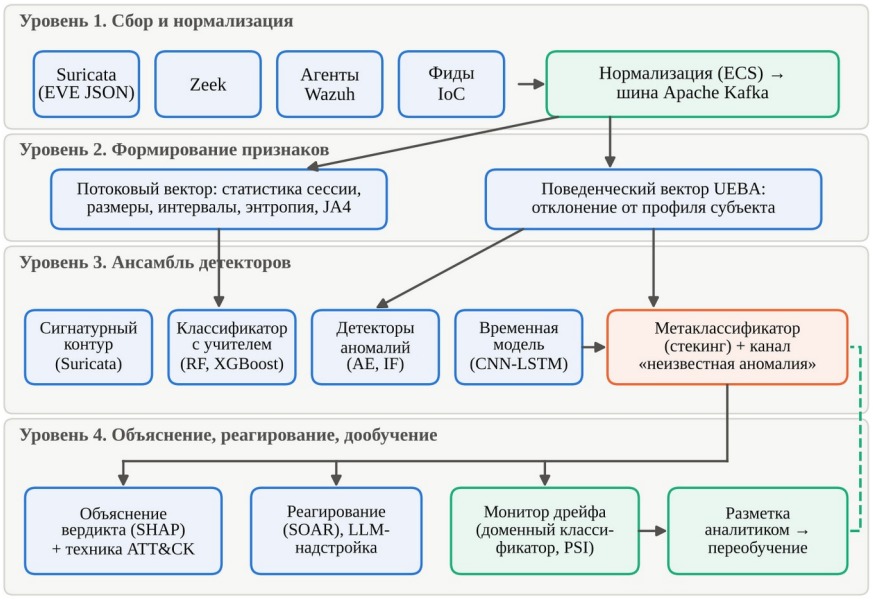


Рисунок 1 - Функциональная архитектура разрабатываемой системы

Вердикт формирует ансамбль четырёх разнородных детекторов — сигнатурного контура, классификатора с учителем на градиентном бустинге [9], детектора аномалий без учителя и модели временных зависимостей. Их оценки объединяются метаклассификатором, обучаемым по схеме стекинга [10]:

$$\hat{s} = g\left(s_1(x), s_2(x), s_3(x), s_4(x)\right), \quad (1)$$

где x — вектор признаков наблюдения; $s_k(x)$ — оценка k-го детектора; g — обучаемая функция агрегации. Чтобы исключить утечку между уровнями, метаклассификатор обучается на отдельной части выборки, которой базовые модели не видели. Поскольку метаклассификатор, обученный на данных известного распределения, склонен подавлять детектор аномалий, последний сохраняется как независимый канал «неизвестная аномалия»:

$$\hat{y}(x) = max\left([p_s(x) \geq 0{,}5]; [\varepsilon(x) \geq \tau_q]\right), \quad (2)$$

где p_s — вероятность атаки по метаклассификатору; $\varepsilon(x)$ — ошибка реконструкции автокодировщика; $[\cdot]$ — индикатор истинности условия; τ_q — q-квантиль ошибки реконструкции на легитимном обучающем трафике. Значение q выбирается по выборке метаклассификатора как наименьшее, при котором доля ложных срабатываний схемы не превышает 3 %; в эксперименте получено q = 0,967–0,972.

Признаки вычисляются в потоковом режиме и не зависят от физической структуры конкретной сети: абсолютные адреса и номера портов заменяются производными характеристиками — принадлежностью адреса внутреннему или внешнему диапазону, категорией сервиса, частотой контактов в окне наблюдения. Дрейф входного распределения оценивается индексом стабильности популяции

$$PSI = \sum_{i=1}^{B} \left(p_i - q_i\right) \cdot \ln \frac{p_i}{q_i}, \quad (3)$$

где p_i и q_i — доли обучающей и текущей выборок, попавшие в i-й из B интервалов значений признака. Значения ниже 0,1 принято считать признаком стабильности, а значения от 0,25 и выше — основанием для переобучения. Индекс не требует разметки текущего трафика. Модуль объяснения вычисляет вклад каждого признака в прогноз по значениям Шепли [11] и сопоставляет его с техникой матрицы MITRE ATT&CK [12], что придаёт объяснению доказательное значение при расследовании инцидента.

Набор данных. Использован набор NSL-KDD [13]: 125 973 записи в обучающей части KDDTrain+ и 22 544 записи в тестовой части KDDTest+, каждое соединение описано 41 признаком. Тестовая часть содержит 37 типов атак, из которых 17 в обучении отсутствуют (3 750 записей, или 29,2 % атак), что воспроизводит появление неизвестных угроз и сдвиг

распределения. Категориальные признаки закодированы унитарным кодом, к счётным и объёмным признакам применено логарифмирование, числовые признаки стандартизованы; размерность вектора — 122. Обучающая часть делилась стратифицированно: 70 % — для базовых моделей (случайный лес из 200 деревьев, XGBoost из 300 деревьев глубиной 8, Isolation Forest [14] и автокодировщик 122–64–16–64–122, обучаемые только на легитимном трафике), 15 % — для метаклассификатора, 15 % — отложенная выборка.

План эксперимента. Принят полный факторный план 2^3; факторы и уровни их варьирования приведены в таблице 1.

Таблица 1 – Факторы и уровни варьирования

Фактор	Содержание	Нижний уровень (−1)	Верхний уровень (+1)
x_1	Тип детектора	Одиночная модель XGBoost	Гибридный ансамбль по (1), (2)
x_2	Дообучение на данных защищаемого потока	Нет	5 % потока, размеченные аналитиком, вес 5
x_3	Условия трафика	Известное распределение (отложенная выборка)	Сдвиг: KDDTest+ с новыми типами атак

Для каждого уровня x_3 поток случайно делился пополам: первая половина (9 448 записей при $x_3 = -1$ и 11 272 при $x_3 = +1$) служила для оценки, из второй при $x_2 = +1$ в обучение добавлялись 945 и 1 127 размеченных записей соответственно. Откликом y служила F-мера на оценочной половине, выраженная в процентах. Каждое из восьми сочетаний воспроизводилось пять раз с разными начальными значениями генератора случайных чисел, определяющими разбиение данных и инициализацию моделей, — всего 40 прогонов. Повторения нужны для оценки дисперсии воспроизводимости, без которой невозможна проверка адекватности.

Статистическая обработка выполнялась по стандартной последовательности активного эксперимента [15] при уровне значимости 0,05. Однородность построчных дисперсий s_u^2 проверялась критерием Кочрена

$$G = \frac{\max s_u^2}{\sum_{u=1}^{N} s_u^2}, \tag{4}$$

где $N = 8$ — число сочетаний уровней; критическое значение для восьми дисперсий с четырьмя степенями свободы составляет 0,391. Затем строилась модель с парными и тройным взаимодействиями

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3, \tag{5}$$

значимость коэффициентов проверялась критерием Стьюдента (критическое значение 2,037 при 32 степенях свободы), значимость модели — критерием Фишера, адекватность — сравнением дисперсии неадекватности с дисперсией воспроизводимости. Нормальность остатков проверялась критерием Шапиро — Уилка [16], мультиколлинеарность — коэффициентом инфляции дисперсии. При неоднородности дисперсий предусмотрено стабилизирующее преобразование [17] — логарифм недобора F-меры до 100 %:

$$z = \ln(100 - y). \tag{6}$$

По дельта-методу дисперсия z приближённо равна дисперсии y, делённой на $(100 - y)^2$, поэтому преобразование компенсирует сужение разброса по мере приближения F-меры к 100 %. Устойчивость выводов к возможному отклонению от нормальности дополнительно проверялась бутстрепом [18]: 20 000 повторных выборок с возвращением внутри каждого сочетания уровней.

Результаты и обсуждение

Средние значения отклика по сочетаниям уровней приведены в таблице 2 и показаны на рисунке 2.

Таблица 2 – Средние значения отклика по сочетаниям уровней (пять повторов)

№	x_1	x_2	x_3	F-мера, % ($\bar{y} \pm s$)	FPR, %	Полнота, %
1	−1	−1	−1	99,88 ± 0,03	0,08	99,86
2	+1	−1	−1	98,19 ± 0,23	3,14	99,92
3	−1	+1	−1	99,90 ± 0,02	0,08	99,88
4	+1	+1	−1	98,19 ± 0,23	3,14	99,92
5	−1	−1	+1	78,52 ± 0,70	2,80	66,01
6	+1	−1	+1	87,30 ± 1,99	4,24	80,00
7	−1	+1	+1	97,35 ± 0,12	2,33	96,52
8	+1	+1	+1	96,90 ± 0,21	4,36	97,10

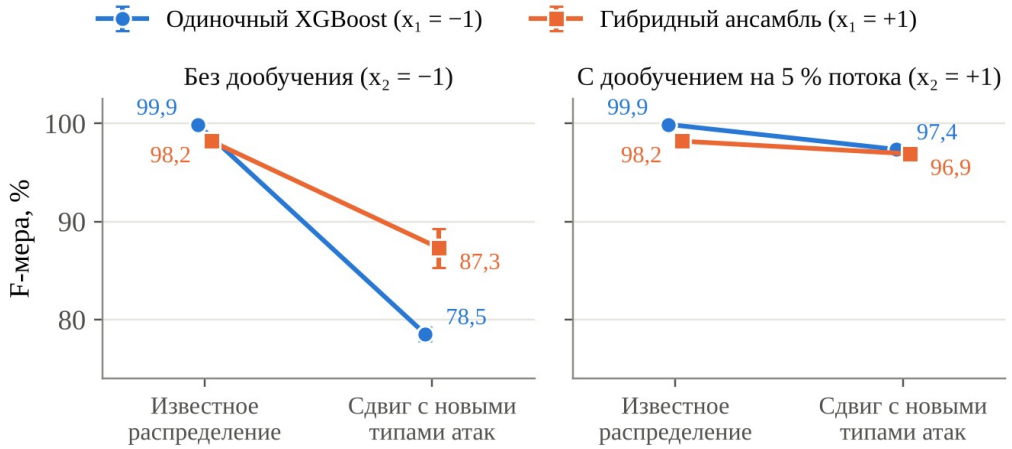


Рисунок 2 - Диаграмма взаимодействия факторов (среднее ± стандартное отклонение)

Сдвиг распределения без дообучения снижает F-меру одиночной модели с 99,88 до 78,52 %, а ансамбля — с 98,19 до 87,30 %. Дообучение на 5 % потока восстанавливает её до 97,35 и 96,90 % соответственно. При известном распределении ансамбль уступает одиночной модели 1,69 п. п.: канал аномалий поднимает долю ложных срабатываний с 0,08 до 3,14 %. В этих условиях вердикт ансамбля практически полностью определяется каналом аномалий, поэтому дообучение метаклассификатора на данных того же распределения его не меняет: результаты строк 2 и 4 таблицы 2 совпали во всех пяти повторах.

Однородность дисперсий. Для исходного отклика критерий Кочрена равен 0,858 при критическом 0,391, и гипотеза однородности отвергается: наибольшая построчная дисперсия превышает наименьшую примерно в 8 200 раз. Вблизи 100 % разброс F-меры сжат, а при сдвиге без дообучения результат ансамбля зависит от того, какие из новых атак попадут в зону канала аномалий при данном разбиении. Протокол [7], проверенный на синтетическом отклике с постоянной дисперсией, этой ситуации не встречал. После преобразования (6) критерий равен 0,330 < 0,391, однородность принимается, а дисперсия воспроизводимости составляет 0,0238 при 32 степенях свободы. Дальнейшая обработка выполнена для *z*.

Регрессионная модель. Благодаря ортогональности плана все коэффициенты имеют одинаковую стандартную ошибку 0,0244, а порог значимости эффекта составляет 0,099. Оценки приведены в таблице 3.

Таблица 3 – Коэффициенты модели для отклика z и их значимость (t-крит. = 2,037)

Коэффициент	Оценка	t	95 % бутстреп-интервал	Вывод
b_0	0,552	22,66	0,507…0,593	значим
b_1	0,656	26,93	0,615…0,700	значим
b_2	−0,450	−18,45	−0,493…−0,407	значим
b_3	1,373	56,32	1,331…1,416	значим
b_{12}	0,099	4,06	0,058…0,142	значим
b_{13}	−0,751	−30,83	−0,795…−0,711	значим
b_{23}	−0,424	−17,42	−0,466…−0,380	значим
b_{123}	0,074	3,03	0,031…0,117	значим

Все восемь коэффициентов значимы, и ни один бутстреп-интервал не содержит нуля, поэтому итоговая модель совпадает с полной (коэффициенты округлены до сотых):

$$\hat{z} = 0{,}55 + 0{,}66\,x_1 - 0{,}45\,x_2 + 1{,}37\,x_3 + 0{,}10\,x_1x_2 - 0{,}75\,x_1x_3 - 0{,}42\,x_2x_3 + 0{,}07\,x_1x_2x_3. \qquad (7)$$

Переход фактора с нижнего уровня на верхний умножает недобор F-меры в среднем в $\exp(2b_j)$ раз. Сдвиг распределения увеличивает недобор в 15,6 раза, дообучение уменьшает его в 2,5 раза. Однако средний эффект типа детектора (b_1 = 0,656) малоинформативен, поскольку его перекрывают взаимодействия b_{13} = −0,751 и b_{123} = 0,074. По таблице 2 при известном распределении ансамбль увеличивает недобор с 0,12 до 1,81 %, при сдвиге без дообучения уменьшает его с 21,48 до 12,70 %, а после дообучения меняет лишь с 2,65 до 3,10 %. Сравнение по одной итоговой оценке в зависимости от условий показало бы, что ансамбль лучше, хуже или равноценен одиночной модели. Взаимодействие b_{23} = −0,424 означает, что дообучение тем полезнее, чем сильнее сдвиг: при известном распределении оно практически ничего не даёт.

Проверка модели. Модель (7) объясняет 99,4 % дисперсии отклика и значима по критерию Фишера: 788,1 при критическом 2,31. Проверку адекватности выполнить невозможно: число значимых коэффициентов равно числу сочетаний уровней ($d = N = 8$), и для дисперсии неадекватности не остаётся степеней свободы. На реальных данных значимо даже тройное взаимодействие, то есть план 2^3 исчерпал свою информационную ёмкость. Для проверки адекватности его необходимо расширить, например введя третий уровень фактора x_2 (доля дообучения 0, 1 и 5 %).

Критерий Шапиро — Уилка даёт W = 0,940 при p = 0,034, то есть гипотеза нормальности остатков на уровне 0,05 формально отвергается; отклонение создают «тяжёлые хвосты» строки 6 таблицы 2, где разброс наибольший. Поскольку бутстреп-интервалы, не опирающиеся на предположение о нормальности, не содержат нуля, выводы о значимости эффектов остаются в силе.

Контроль дрейфа. Значения индекса (3) для признаков с наибольшим сдвигом между KDDTrain+ и KDDTest+ показаны на рисунке 3.

© Central Asian Scientific Journal, 2026

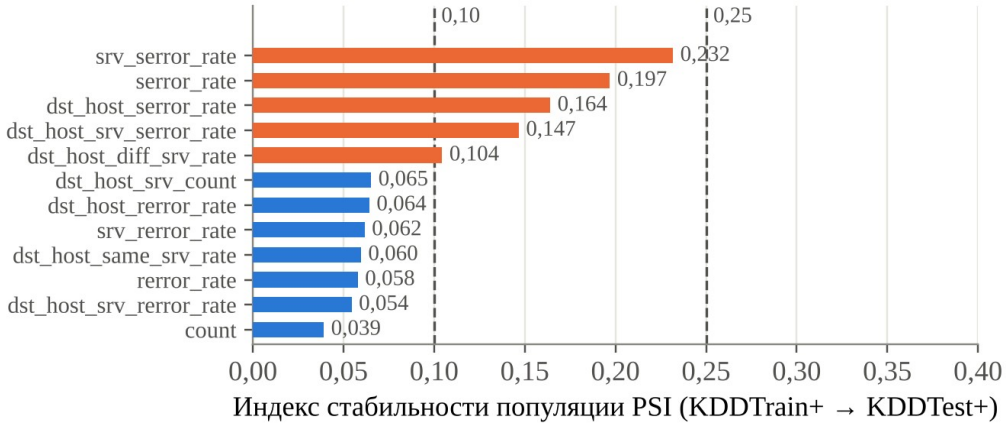


Рисунок 3 - Индекс стабильности популяции по признакам; пунктир — пороги 0,10 и 0,25

Ни один из 38 числовых признаков не достиг порога переобучения 0,25; лишь 5 признаков попали в зону умеренного сдвига от 0,10 до 0,25. Правило, основанное только на индексе стабильности, не инициировало бы переобучение, хотя F-мера одиночной модели при этом снизилась на 21,4 п. п. Многомерный доменный классификатор [19], обучаемый отличать записи эталонного потока от текущего, различает их с площадью под ROC-кривой 0,905, тогда как для двух случайных половин обучающей выборки она равна 0,502. Следовательно, сдвиг проявляется в сочетаниях признаков, а не в отдельных признаках. Индекс стабильности полезен лишь как интерпретируемое дополнение, указывающее, какие признаки сместились. С учётом эффекта b_{23} переобучение целесообразно запускать по событию, с участием аналитика и проверкой новой версии модели в теневом режиме.

Полученные результаты имеют ограничения. Набор NSL-KDD синтетичен и не содержит современного зашифрованного трафика, поэтому абсолютные значения отклика не переносятся на реальную сеть; переносимы прежде всего структура эффектов и выводы о самом протоколе. Сигнатурный контур и модель временных зависимостей в эксперименте не реализованы, а второй этап протокола [7] — оценка скорости деградации во времени — не воспроизведён, так как записи набора не упорядочены по времени.

Заключение

Протокол полного факторного эксперимента с повторениями, ранее проверенный на синтетическом отклике, применён к реально измеренной F-мере гибридной системы обнаружения киберугроз. Эксперимент 2^3 из 40 прогонов показал, что на реальных данных нарушается предпосылка однородности дисперсий (критерий Кочрена 0,858 при критическом 0,391) и её следует восстанавливать логарифмированием недобора F-меры, после чего значимыми оказываются все восемь эффектов, включая тройное взаимодействие. Насыщенность модели делает проверку адекватности невозможной и указывает на необходимость расширения плана.

По существу установлено, что преимущество ансамбля не является постоянной величиной: при известном распределении он уступает одиночной модели 1,7 п. п., при сдвиге с новыми типами атак без дообучения превосходит её на 8,8 п. п., а после дообучения на 5 % размеченного потока разница составляет 0,4 п. п. в пользу одиночной модели. Индекс стабильности популяции не выявил сдвига, снизившего F-меру более чем на 20 п. п., тогда как доменный классификатор обнаружил его с площадью под ROC-кривой 0,905. Для разрабатываемой системы отсюда следуют два проектных решения: многомерный монитор дрейфа и событийный контур дообучения с участием аналитика.

Дальнейшая работа предполагает применение протокола к трафику действующей сети и к другим наборам данных, реализацию сигнатурного контура и временной модели, а также оценку устойчивости ансамбля к состязательным воздействиям.

Список литературы:

1. Jacob S. L., Sultana Habibullah P. A Systematic Analysis and Review on Intrusion Detection Systems Using Machine Learning and Deep Learning Algorithms // Journal of Computational and Cognitive Engineering. – 2025. – Vol. 4, No. 2. – P. 108–120. – DOI: 10.47852/bonviewJCCE42023249.

2. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection // Proceedings of the IEEE Symposium on Security and Privacy. – Oakland, 2010. – P. 305–316. – DOI: 10.1109/SP.2010.25.

3. Apruzzese G., Colajanni M., Ferretti L. [et al.] On the Effectiveness of Machine and Deep Learning for Cyber Security // Proceedings of the 10th International Conference on Cyber Conflict (CyCon). – Tallinn, 2018. – P. 371–390. – DOI: 10.23919/CYCON.2018.8405026.

4. Горюнов М. Н., Мацкевич А. Г., Рыболовлев Д. А. Синтез модели машинного обучения для обнаружения компьютерных атак на основе набора данных CICIDS2017 // Труды Института системного программирования РАН. – 2020. – Т. 32, вып. 5. – С. 81–94.

5. Об информатизации : Закон Республики Казахстан от 24 ноября 2015 года № 418-V [Электронный ресурс]. – URL: https://adilet.zan.kz/rus/docs/Z1500000418 (дата обращения: 10.11.2026).

6. Об утверждении единых требований в области информационно-коммуникационных технологий и обеспечения информационной безопасности : Постановление Правительства Республики Казахстан от 20 декабря 2016 года № 832 [Электронный ресурс]. – URL: https://adilet.zan.kz/rus/docs/P1600000832 (дата обращения: 10.11.2026).

7. Kim V., Khon D., Kadyr A., Tskhai V. A Full Factorial Design for the Comparative Evaluation of Intrusion Detection Configurations: Methodology and Controlled Validation : препринт. – Almaty : Satbayev University, 2026. – 10 p. – URL: https://zenodo.org/records/24158881.

8. Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection // Proceedings of the Network and Distributed System Security Symposium (NDSS). – San Diego, 2018. – 15 p. – DOI: 10.14722/ndss.2018.23204.

9. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – San Francisco, 2016. – P. 785–794. – DOI: 10.1145/2939672.2939785.

10. Wolpert D. H. Stacked generalization // Neural Networks. – 1992. – Vol. 5, No. 2. – P. 241–259. – DOI: 10.1016/S0893-6080(05)80023-1.

11. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems 30 (NeurIPS 2017). – Long Beach, 2017. – P. 4765–4774.

12. MITRE ATT&CK: Enterprise Matrix [Электронный ресурс] / The MITRE Corporation. – URL: https://attack.mitre.org (дата обращения: 10.11.2026).

13. Tavallaee M., Bagheri E., Lu W., Ghorbani A. A. A Detailed Analysis of the KDD CUP 99 Data Set // Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA). – Ottawa, 2009. – P. 1–6. – DOI: 10.1109/CISDA.2009.5356528.

14. Liu F. T., Ting K. M., Zhou Z.-H. Isolation Forest // Proceedings of the 8th IEEE International Conference on Data Mining (ICDM). – Pisa, 2008. – P. 413–422. – DOI: 10.1109/ICDM.2008.17.

15. Адлер Ю. П., Маркова Е. В., Грановский Ю. В. Планирование эксперимента при поиске оптимальных условий. – 2-е изд., перераб. и доп. – М. : Наука, 1976. – 279 с.

16. Shapiro S. S., Wilk M. B. An Analysis of Variance Test for Normality (Complete Samples) // Biometrika. – 1965. – Vol. 52, No. 3/4. – P. 591–611. – DOI: 10.2307/2333709.

17. Box G. E. P., Cox D. R. An Analysis of Transformations // Journal of the Royal Statistical Society. Series B (Methodological). – 1964. – Vol. 26, No. 2. – P. 211–252.

18. Efron B., Tibshirani R. J. An Introduction to the Bootstrap. – New York : Chapman & Hall, 1993. – 436 p.

19. Rabanser S., Günnemann S., Lipton Z. C. Failing Loudly: An Empirical Study of Methods for Detecting Dataset Shift // Advances in Neural Information Processing Systems 32 (NeurIPS 2019). – Vancouver, 2019. – P. 1394–1406.