30. *Leonov A.S., Sharapov I.V.* Vybor parametra regulyarizatsii Tikhonova v nekorrektnykh zadachakh, *Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki*, 2021, Vol. 61, No. 4, pp. 451-466. Available at: https://doi.org/10.31857/S004446692104008X.
31. *Montasari R.* A Standardised Digital Forensic Investigation Process Model (SDFIPM), *International Journal of Information Security*, 2021, Vol. 20, No. 2, pp. 195-211. Available at: https://doi.org/10.1007/s10207-020-00495-2.
32. *Alqahtani A., Sheldon F.T.* A Survey of Formal Methods and Their Applications in Cyber Security, *IEEE Access*, 2022, Vol. 10, pp. 68702-68719. Available at: https://doi.org/10.1109/ACCESS.2022.3186256.
33. *Kaddour J.* Causal Machine Learning: A Survey and Open Problems, *arXiv preprint*, 2022. Available at: https://doi.org/10.48550/arXiv.2206.15475.
34. *Lande D.* Adjustment of the analytic hierarchy process indicators using AI tools, *Information Technologies and Computer Engineering*, 2025, Vol. 22, No. 1, pp. 104-112.
35. *Sundaramoorthy S.* Cognitive bias in digital forensics: A review, *Forensic Science International: Digital Investigation*, 2021, Vol. 38, pp. 301138. Available at: https://doi.org/10.1016/ j.fsidi.2021.301138.
36. *Matveeva V.S.* Statisticheskiy metod obnaruzheniya lokal'nykh neodnorodnostey dannykh dlya rassledovaniya intsidentov informatsionnoy bezopasnosti: dis. … kand. tekhn. nauk [Statistical method for detecting local data inhomogeneities for information security incident investigation: Cand. of eng. sc. diss.]: 05.13.19. NIYaU MIFI. Moscow, 2015.
37. *Siboni S., & Cohen A.* Anomaly Detection for Individual Sequences with Applications in Identifying Malicious Tools, *Entropy*, 2020, 22 (6), 649. Available at: https://doi.org/10.3390/e22060649.
38. *Kim Juhwan & Son Baehoon & Yu Jihyeon & Yun Joobeom.* AI-Driven Prioritization and Filtering of Windows Artifacts for Enhanced Digital Forensics, *Computers, Materials & Continua*, 2024, 81, pp. 3371-3393. 10.32604/cmc.2024.057234.

Абрамов Евгений Сергеевич – Южный федеральный университет; e-mail: abramoves@sfedu.ru; г. Таганрог, Россия; тел.: +78634371905; к.т.н.; доцент.

Abramov Eugine Sergeevich – Southern Federal University; e-mail: abramoves@sfedu.ru; Taganrog, Russia; phone: +78634371905; cand. of eng. sc.; associate professor.

УДК 004.056 DOI 10.18522/2311-3103-2026-3-68-83

М.А. Киселев

СЦЕНАРНО- И РИСК-ВЗВЕШЕННАЯ МЕТОДИКА ДЕФИЦИТА ЛОГИРОВАНИЯ ДЛЯ SIEM С ВЕРОЯТНОСТНЫМ ОПОРНЫМ ПРОФИЛЕМ ИНТЕНСИВНОСТИ И ОЦЕНКОЙ ПРИГОДНОСТИ СОБЫТИЙ К КОРРЕЛЯЦИИ

Актуальность. *Эффективность корреляции в SIEM определяется не только полнотой поступления событий, но и их пригодностью к сценарному анализу. Практические дефекты логирования, включая недологирование, нарушение временной валидности, потерю корреляционно-значимых атрибутов, снижают качество детектирования и повышают вероятность ложных срабатываний.* ***Цель.*** *Разработать сценарно- и риск-взвешенную методику количественной оценки дефицита логирования как меры пригодности потока событий к корреляции.* ***Задачи.*** *Сформировать сценарно-обоснованное множество контролируемых классов событий, задать их риск-веса, построить вероятностный опорный профиль интенсивности, ввести критерии временной валидности и структурной пригодности записи, а также получить итоговый показатель с диагностической интерпретацией причин ухудшения качества логирования.* ***Методы.*** *Использованы сценарный анализ детект-контента, риск-взвешенная агрегация, построение baseline по историческим данным, контроль временной валидности событий и оценка полноты минимально достаточных профилей корреляционно-значимых атрибутов.* ***Результаты.*** *Предложенная методика расширяет базовый показатель дефицита логирования за счёт использования вероятностного опорного профиля интенсивности, учёта перелогирования и потери корреляционно-значимых атрибутов, а также введения диагностической декомпозиции причин деградации качества логирования. Экспериментальная апробация показала, что в baseline-режиме* $DL_{new} = 0{,}000$ *при* $DL_{old} = 0{,}297$*, что подтверждает отсутствие ложных срабатываний нового показателя на штатный поток; в сценарии дублирования* $DL_{old} = 0{,}000$ *при* $DL_{new} = 0{,}073$*, что демонстрирует способность нового показателя выявлять перелогирование, невидимое для базового показателя; при нарушении временной согласованности все показатели достигают 1{,}000, что соответствует полной непригодности потока к корреляции.* ***Выводы и значимость****. Научная новизна состоит в формализации дефицита логирования как дефицита пригодности событийного потока к корреляции с учётом сценарной значимости классов событий и введении диагностической декомпози-*

ции, позволяющей выделять доминирующий диагностический компонент деградации. Практическая значимость заключается в возможности количественно обосновывать приоритетные меры по совершенствованию источников, нормализации и правил корреляции в SIEM.

SIEM; дефицит логирования; корреляция событий ИБ; качество логирования; опорный профиль интенсивности.

M.A. Kiselev

SCENARIO- AND RISK-WEIGHTED LOGGING DEFICIT METHODOLOGY FOR SIEM WITH A PROBABILISTIC INTENSITY BASELINE AND ASSESSMENT OF EVENT SUITABILITY FOR CORRELATION

Relevance. *The effectiveness of correlation in SIEM is determined not only by the completeness of incoming events, but also by their suitability for scenario-based analysis. Practical logging defects, including under-logging, violations of temporal consistency, loss of mandatory attributes, and record duplication, degrade detection quality and increase the likelihood of false positives.* ***Objective.*** *To develop a scenario- and risk-weighted methodology for the quantitative assessment of logging deficit as a measure of the suitability of an event stream for correlation.* ***Tasks.*** *To form a scenario-justified set of controlled event classes, assign risk weights to them, construct a probabilistic baseline intensity profile, introduce criteria of temporal validity and structural suitability of records, and obtain an integral indicator with diagnostic interpretation of the causes of logging quality degradation.* ***Methods.*** *The study employs scenario analysis of detection content, risk-weighted aggregation, baseline construction from historical data, monitoring of event temporal validity, and assessment of the completeness of minimally sufficient profiles of correlation-significant attributes.* ***Results.*** *The proposed method extends the basic logging deficit indicator by introducing a probabilistic baseline intensity profile, sensitivity to over-logging, loss of correlation-significant fields, and a diagnostic decomposition of logging quality degradation causes. Experimental validation showed that in the baseline mode, DL_new = 0.000 while DL_old = 0.297, confirming the absence of false positives produced by the new indicator on a normal event stream. In the duplication scenario, DL_old = 0.000 while DL_new = 0.073, demonstrating the ability of the new inidcator to detect over-logging that remains invisible to the basic indicator. When temporal consistency is violated, all indicators reach 1.000, which corresponds to the complete unsuitability of the event stream for correlation.* ***Conclusions and significance.*** *The scientific novelty lies in the formalization of logging deficit as a deficit of event-stream suitability for correlation, taking into account the scenario significance of event classes and introducing a diagnostic decomposition that makes it possible to identify the dominant diagnostic component of degradation. The practical significance lies in the ability to quantitatively justify priority measures for improving event sources, normalization procedures, and correlation rules in SIEM.*

SIEM; logging deficit; security event correlation; logging quality; baseline intensity profile.

Введение. Современные SIEM-системы обеспечивают выявление инцидентов за счёт корреляции событий информационной безопасности (ИБ), поступающих из разнородных источников: операционных систем, средств защиты информации, сетевой инфраструктуры и прикладных сервисов. При этом качество корреляции определяется не только фактом поступления событий, но и способностью событийного потока поддерживать сценарии детектирования: события должны появляться в ожидаемых объёмах, содержать корреляционно-значимые атрибуты и иметь корректную временную привязку. На практике именно эти свойства чаще всего нарушаются в производственных контурах мониторинга: наблюдаются пропуски классов событий вследствие ошибок сбора или фильтрации, всплески и дублирование событий при сбоях агентов и транспортных компонентов, несогласованность временных меток события и времени приёма, а также снижение структурной пригодности событий к корреляции (частично заполненные поля, отсутствие идентификаторов субъектов/объектов взаимодействия, некорректные форматы значений) [1, 2]. Указанные дефекты приводят к двум критическим эффектам: (1) снижению полноты выявления атак (корреляционные правила не срабатывают при пропуске событий или потере ключевых атрибутов); (2) росту ложных срабатываний и перегрузке аналитиков центра мониторинга ИБ при появлении дублей и низкоинформативных/шумовых событий ИБ.

Существующие практики контроля логирования в контуре SIEM преимущественно опираются на проверку охвата источников, факт наличия событий и сопоставление общего объёма потока с ожиданиями [3]. Эти проверки полезны на этапе подключения источника, однако не позволяют количественно и воспроизводимо ответить на ключевой вопрос для центра мониторинга ИБ: насколько текущий поток событий пригоден для корреляции конкретных сценариев

атак и какие дефекты логирования являются приоритетными для устранения с учётом риска. В результате, улучшения логирования часто выполняются реактивно и локально: устраняется локальный отказ доставки (например, прекращение поступления событий отдельных классов/каналов журналов), однако сохраняется неопределённость: приводит ли изменение к улучшению пригодности потока для корреляции.

Формулировка проблемы. В практических руководствах по лог-менеджменту подчёркивается, что ключевыми источниками ухудшения качества событийного потока как входных данных для корреляции являются временная валидность (inconsistent timestamps), неоднородность форматов и содержимого событий между источниками, а также риски потери данных при перегрузках и ограничениях инфраструктуры сбора [1]. Дополнительно при агентно-ориентированной схеме централизованного сбора событий, использующей подтверждаемую доставку, а также при отдельных стратегиях ротации файлов, возможны повторная отправка и дублирование событий либо потеря части записей, что снижает корректность последующей аналитики и корреляции [4, 5].

Следовательно, решаемая проблема заключается в отсутствии методики количественной оценки дефицита логирования как дефицита пригодности потока к корреляции – методики, обеспечивающей воспроизводимый показатель и одновременно различая типовые причины ухудшения качества событийного потока для корреляции – отклонение интенсивности от опорного профиля, нарушение валидности времени событий и снижение структурной пригодности записей к корреляции – а также обеспечивая приоритизацию работ по улучшению логирования в соответствии с риском сценариев атак**.**

Актуальность темы и практическая значимость. Актуальность задачи обусловлена тем, что в практике мониторинга событий ИБ улучшение логирования часто начинается с обеспечения «охвата» и базовой централизации событий ИБ. В рекомендациях по управлению событиями ИБ контроль прямо формализуется как обеспечение включённого логирования на активах, централизованный сбор и сопутствующие организационно-технические меры; при этом широко применяется показатель охвата (coverage) и оценка полноты подключения источников [3]. Такой вариант необходим как базовый уровень, однако он не даёт количественного ответа на вопрос: какие дефекты логирования мешают конкретным сценариям корреляции и что устранять в первую очередь, если ресурсы на доработки ограничены.

Для научной и прикладной области мониторинга событий ИБ это означает потребность в показателе, которая (а) связывает требования детектирования со строго определённым набором классов событий, (б) учитывает риск-значимость классов событий в сценарии, (в) обнаруживает не только недологирование, но и перелогирование/дубли как фактор ложных срабатываний и перегрузки, (г) обеспечивает диагностическую декомпозицию причин ухудшения качества входных данных для корреляции.

Краткий обзор состояния вопроса и источников. В источниках по лог-менеджменту, включая NIST SP 800-92, рассматриваются вопросы организации инфраструктуры журналирования, синхронизации времени, унификации процедур сбора и хранения журналов, а также риски потери данных при перегрузке каналов и пиковых нагрузках [1]. Практические фреймворки, в частности CIS Controls, связывают качество журналирования с полнотой охвата источников, централизованным сбором и регламентированным анализом журналов [3]. Документация по инструментам сбора и пересылки журналов, таким как Beats/Filebeat, указывает, что использование режима доставки *at-least-once* в ряде ситуаций может сопровождаться дублированием событий. Дополнительным источником искажений выступают отдельные схемы ротации лог-файлов, способные приводить как к повторной обработке ранее считанных данных, так и к частичной потере событий [4, 5]. Исследование журналов операционных систем и проблематика качества атрибутов событий систематизированы в [6]. Применение SIEM для защиты критических инфраструктур отражено в [7]. Таким образом, в литературе и прикладных руководствах подробно рассмотрены причины ухудшения качества логирования и базовые практики его контроля, однако сохраняется методологический разрыв [8, 9]: отсутствует единая сценарно-ориентированная количественная метиодика, привязанная к корреляции в SIEM и позволяющая воспроизводимо сравнивать пригодность событийного потока по сценариям атак и по классам событий.

Термины и определения. *Опорный профиль интенсивности (baseline)* – это эмпирически оценённая по историческим данным интервальная модель ожидаемого числа событий каждого класса в фиксированном окне T, используемая для выявления отклонений вниз (underlogging) и вверх (overlogging/дубли).

Дефицит логирования в рамках настоящей работы трактуется как снижение пригодности событийного потока к корреляции относительно сценарно-заданных требований к интенсивности, временной валидности и структурной полноте записей. Поэтому величина дефицита логирования может возрастать не только при недопоступлении событий, но и при их переизбыточности (дублировании/перелогировании), временной невалидности и утрате корреляционно-значимых атрибутов.

Структурная пригодность записи к корреляции – это степень наличия и валидности минимально достаточного набора корреляционно-значимых атрибутов $P_{s,j}$, необходимого для вычислимости условий правила, связывания событий и минимальной интерпретации результата.

Цель исследования. Цель работы – разработать методику сценарно- и риск-взвешенной количественной оценки дефицита логирования в SIEM, отражающую пригодность событийного потока к корреляции по заданным сценариям атак и обеспечивающую диагностическую декомпозицию причин ухудшения качества событийного потока для корреляции (интенсивность потока, валидность времени, структурная пригодность событий).

Задачи исследования. Для достижения цели необходимо решить следующие задачи:

1) сформировать сценарно-обоснованный набор контролируемых классов событий для выбранного сценария детектирования и определить механизм задания риск-весов классов;

2) построить вероятностный опорный профиль ожидаемой интенсивности потока по классам событий и определить критерий отклонений для выявления недологирования и перелогирования/дублирования;

3) определить критерий валидности времени событий с учётом допуска рассогласования/задержек доставки;

4) формализовать минимально достаточный профиль корреляционно-значимых атрибутов для пары «класс источника / класс события» и ввести показатель пригодности записи к корреляции как долю заполнения/валидности атрибутов профиля;

5) сформировать итоговый сценарно- и риск-взвешенный показатель дефицита логирования и разработать диагностическую декомпозицию причин ухудшения качества потока событий по компонентам (интенсивность, валидность времени, структурная пригодность) по классам событий, а также ввести контрольный показатель $DL_{\mathrm{scn}}(T)$ с равномерными весами для оценки влияния риск-взвешивания на итоговую оценку;

6) экспериментально подтвердить работоспособность методики на воспроизводимом стенде с контролируемыми нарушениями логирования и выполнить сопоставление с базовой версией показателя DL, введённой ранее [10].

Объект исследования – процесс формирования, передачи, нормализации и использования событий ИБ в контуре SIEM.

Предмет исследования – методика вычисления дефицита логирования как меры пригодности событийного потока к корреляции по сценариям с учётом риск-значимости классов событий и диагностикой причин ухудшения качества событийного потока как входных данных для корреляции.

Гипотеза исследования. Гипотеза состоит в следующем: если оценивать качество логирования не только по факту присутствия событий и общему объёму, а через интеграцию опорного профиля ожидаемой интенсивности по классам событий, критерия валидности времени и критерия структурной пригодности записи к корреляции, и затем агрегировать результаты по сценарию с риск-весами, то полученный показатель дефицита логирования будет (а) воспроизводимо различать основные типы дефектов логирования, (б) обеспечивать приоритизацию корректирующих мер по риску сценария и (в) демонстрировать преимущество над базовой версией DL [10] при нарушениях, связанных с дублированием, утратой корреляционно-значимых атрибутов и/или некорректностью их значений.

Обозначения и формальная постановка задачи. Рассматривается поток событий ИБ, поступающий в SIEM в виде нормализованных записей. Источники событий разделяются на классы $s \in S$ (например, Linux/Windows/Database/Network/EDR). Для каждого источника выделяется множество классов событий $j \in \{1, \dots, m\}$, соответствующих типам, используемым в корреляции (например, AUTH_FAIL, AUTH_SUCCESS, PROC_CREATE, NET_CONN). Анализ качества логирования выполняется на временных окнах длительности T(в практических контурах удобно использовать $T = 1$ час).

Для учёта циклического поведения нагрузок вводится индекс «час недели» $h \in \{1, \dots, 168\}$, вычисляемый по времени окончания окна анализа (1):

$$h = 24 \cdot weekday + hour + 1, \quad (1)$$

где $weekday \in \{0, \dots, 6\}$, $hour \in \{0, \dots, 23\}$.

Каждому сценарию детектирования сопоставляется индекс u и сценарно-обоснованное подмножество классов событий J_u, входящее в нормативное супермножество J_{norm} (нормативное супермножество всех допустимых классов событий, задаваемое политикой мониторинга SOC и охватывающее все классы, потенциально используемые в детект-контенте):

$$J_u \subseteq J_{norm}.$$

Для приоритизации качества логирования в рамках сценария вводятся риск-веса $w_{u,j}$ для классов $j \in J_u$, удовлетворяющие условиям неотрицательности и нормировки:

$$w_{u,j} \geq 0, \sum_{j \in J_u} w_{u,j} = 1.$$

Тем самым задаётся сценарно- и риск-взвешенная оценка, в которой ухудшение качества наблюдаемого потока событий критичных классов оказывает больший вклад в итоговый показатель.

Пусть $E_{j,T}$ – множество событий класса j, наблюдаемых в окне T, а $N_{h,j}$ – число событий данного класса в окне T, соответствующем часу недели h. На основе исторических данных для каждой пары (h, j) строится вероятностный опорный профиль интенсивности (baseline) потока, задаваемый допустимым интервалом $[L_{h,j}, U_{h,j}]$ и типичным уровнем $MU_{h,j}$. Отклонения наблюдаемой интенсивности от baseline используются для выявления недологирования ($N_{h,j} < L_{h,j}$) и перелогирования/дублирования ($N_{h,j} > U_{h,j}$).

Корректность временной привязки событий описывается через две временные метки: t_{event} – время события на источнике и t_{ingest} – время поступления события в индексатор (в практике – поле @timestamp) [11]. Вводится допуск расхождения Δ между t_{event} и t_{ingest}, на основе которого определяется валидность времени отдельной записи $A_e \in \{0,1\}$ и агрегированная валидность по классу $A_j(T)$ как доля валидных событий в окне [12].

Пригодность записи к корреляции [13] формализуется через минимально достаточный профиль корреляционно-значимых атрибутов $P_{s,j}$ для пары «класс источника s / класс события j». Для каждой записи e вычисляется показатель пригодности K_e как доля заполненных и валидных атрибутов профиля, а для класса событий j в окне T– средняя пригодность $K_j(T)$.

Требуется построить сценарно- и риск-взвешенный показатель дефицита логирования, который агрегирует качество потока по классам $j \in J_u$ и обеспечивает диагностируемую декомпозицию причин ухудшения качества сценарно-обусловленного потока событий по компонентам: (1) отклонение интенсивности относительно baseline; (2) валидность времени; (3) структурная пригодность событий к корреляции. В качестве базового уровня сопоставления используется показатель DL, введённый ранее [10] (далее – DL_{old}); в настоящей работе методика развивает этот показатель в виде компонентного представления качества логирования, обеспечивающей чувствительность к перелогированию/дублированию и снижению полноты и корректности корреляционно-значимых атрибутов.

Методика количественной оценки дефицита логирования. Разработанная методика предназначена для количественной оценки пригодности событийного потока к корреляции в SIEM по заданному сценарию детектирования. Методика оценивает качество потока событий по трём диагностическим компонентам качества логирования:

1) интенсивность потока относительно опорного профиля – фиксирует недологирование и перелогирование/дублирование;

2) валидность времени – фиксирует рассогласование временных меток события и времени приёма, критичное для окон корреляции;

3) структурная пригодность записи – фиксирует наличие обязательных корреляционно-значимых атрибутов и корректность их значений.

Общая структура разработанной методики, включая состав входных данных, вычислительные блоки и виды выходных результатов, представлена на рис. 1.

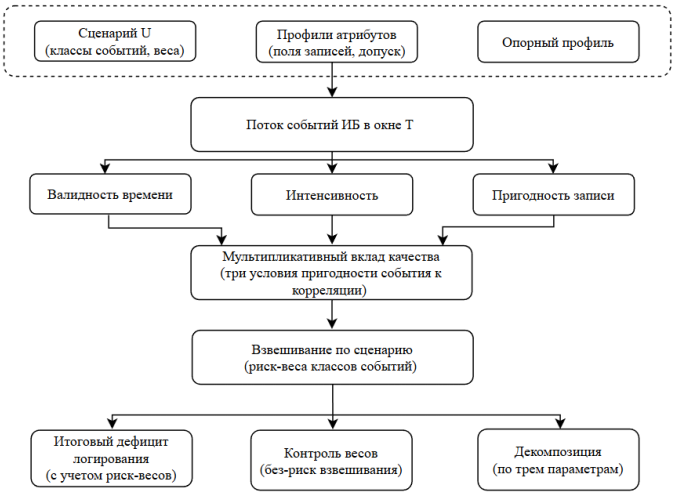

Рис. 1. Структура методики количественной оценки дефицита логирования

Исходные данные для расчета количественной оценки дефицита логирования. В соответствии с введёнными обозначениями рассматриваются: окно анализа T, индекс часа недели h, множества $J_u \subseteq J_{norm}$ и веса $w_{u,j}$, а также величины $N_{h,j}$, $[L_{h,j}, U_{h,j}], MU_{h,j}$, допуск Δ, профиль $P_{s,j}$ и показатели $A_j(T)$, $K_j(T)$.

Расчет выполняется на каждом окне T и включает три подготовительных шага (задаются один раз):

- сценарная настройка $J_u, w_{u,j}$ – задать сценарий u, множество классов J_u и веса $w_{u,j}$;
- настройка критериев качества события $(\Delta, P_{s,j})$ – задать допуск Δ и режим валидности времени (жёсткий/мягкий), задать минимально достаточные профили атрибутов $P_{s,j}$ для пар «класс источника s/ класс события j»;
- построение опорного профиля интенсивности $([L_{h,j}, U_{h,j}], MU_{h,j})$ для каждой пары h, j.

После подготовки на каждом окне T выполняется расчёт компонент $A_j(T)$, $C_{prob}(h,j)$, $K_j(T)$ и их агрегация в $DL_{new}(T)$и $DL_{old}(T)$.

1. Формирование множества классов J_u. Множество J_u формируется как множество классов событий, необходимых для выполнения условий корреляции, а также классов, обеспечивающих контекст и подтверждение инцидента в рамках анализируемого сценария.

2. Формирование риск-весов $w_{u,j}$. В рамках настоящей работы под риск-весом $w_{u,j}$ понимается **сценарный вес значимости** класса событий для детектирования, вычисляемый по структуре детект-контента через показатели использования $L_{u,j}$ и обязательности $I_{u,j}$.

2.1 Задание детект-контента сценария. Сценарий u фиксируется как конечное множество правил корреляции SIEM (2):

$$R_u = \{r_1, \dots, r_M\}, M = |R_u|, \quad (2)$$

где под правилом r понимается формализованное условие выявления признака сценария (корреляционное правило, запрос-детект, алерт-правило), использующее определённые классы событий и их атрибуты.

2.2 Извлечение классов событий, используемых правилами. Для каждого правила $r \in R_u$ определяется множество классов событий, используемых в его логике:

$$J_{(r)} \subseteq J_{norm}.$$

Класс j считается используемым правилом r, если события данного класса участвуют в условиях правила (предикатах), ключах связывания/агрегации (группировке), либо обязательных проверках, требуемых для фиксации детектируемого признака.

На основе $J(r)$вводится индикатор использования (3):

$$M_{r,j} = \begin{cases} 1, j \in J(r), \\ 0, j \notin J(r). \end{cases} \quad (3)$$

Множество классов сценария J_u при необходимости может быть согласовано с детект-контентом:

$$J_u \Leftarrow \bigcup_{r \in R_u} J_{(r)}$$

либо (если J_u задано ранее по декомпозиции сценария) используется пересечение $J_u \cap \bigcup_{r \in R_u} J(r)$ для обеспечения реализуемости.

2.3. Определение обязательности классов для правил. Для вычисления показателя влияния класса j на детектируемость сценария вводится индикатор обязательности (жёсткой зависимости) класса для правила (4):

$$H_{r,j} = \begin{cases} 1, j \in J_{req}(r), \\ 0, j \notin J_{req}(r). \end{cases} \quad (4)$$

где $J_{\mathrm{req}}(r) \subseteq J(r)$ – подмножество классов, необходимых для вычислимости/проверяемости логики правила r. Практически $j \in J_{\mathrm{req}}(r)$, если без событий класса j правило не может быть проверено по данным окна T (например, класс участвует в ключевом предикате срабатывания или в обязательной корреляционной связке). Классы, используемые только для обогащения, контекстирования или подтверждающих условий, относятся к $J(r) \setminus J_{\mathrm{req}}(r)$.

2.4. Расчёт показателей использования и влияния. На основе $M_{r,j}$ и $H_{r,j}$ вводятся два вычислимых показателя для каждого $j \in J_u$. Показатель использования (usage) (5) – доля правил сценария, в которых используется класс j:

$$L_{u,j} = \frac{1}{M} \sum_{r \in R_u} M_{r,j}, 0 \le L_{u,j} \le 1. \quad (5)$$

Показатель влияния (impact) (6) – доля правил сценария, для которых класс j является обязательным:

$$I_{u,j} = \frac{1}{M} \sum_{r \in R_u} H_{r,j}, 0 \le I_{u,j} \le 1. \quad (6)$$

Интерпретация показателей фиксируется процедурно: $L_{u,j}$ характеризует степень вовлечённости класса в детект-контент сценария, а $I_{u,j}$ – степень структурной необходимости класса для проверки условий детектирования.

2.5. Формирование и нормировка риск-весов. Вес класса определяется как сумма показателей:

$$\widetilde{w}_{u,j} = L_{u,j} + I_{u,j}.$$

Далее выполняется нормировка по множеству J_u (7):

$$w_{u,j} = \frac{\widetilde{w}_{u,j}}{\sum_{k \in J_u} \widetilde{w}_{u,k}}, \sum_{j \in J_u} w_{u,j} = 1. \quad (7)$$

Формула $\widetilde{w}_{u,j} = L_{u,j} + I_{u,j}$ обеспечивает ненулевой вес для каждого класса из J_u. Обязательный класс, используемый во всех правилах сценария ($L_{u,j} = 1$, $I_{u,j} = 1$), получает максимальный балл $\widetilde{w}_{u,j} = 2$: за участие в детект-контенте и за обязательность для срабатывания. Контекстный класс, используемый во всех правилах, но не обязательный ($L_{u,j} = 1$, $I_{u,j} = 0$), получает $\widetilde{w}_{u,j} = 1$. В общем случае оба показателя принимают дробные значения в зависимости от доли правил, в которых участвует класс. Тем самым деградация любого класса, входящего в J_u, вносит измеримый вклад в итоговый показатель дефицита логирования. Таким образом, в данной постановке риск-взвешивание реализуется как приоритизация классов событий по их значимости для обнаружения сценария, выраженной через структуру детект-контента.

3. Формирование минимально достаточных профилей атрибутов $P_{s,j}$. Для каждой пары «класс источника s/ класс события j» задаётся минимально достаточный профиль корреляционно-значимых атрибутов $P_{s,j}$. Профиль формируется на основе детект-контента сценария R_u и определяет набор полей, необходимых для вычислимости условий корреляции и формирования ключей связывания/агрегации (группировки) в пределах окна T, а также для минимальной интерпретации результата события.

Процедура построения $P_{s,j}$ включает:

1) извлечение атрибутов, используемых правилами $r \in R_u$ для класса j в предикатах (условиях), ключах агрегации/связывания и обязательных проверках;

2) добавление минимального набора идентификаторов субъект–объект–контекст (например, host, user, src/dst, object/process), если они участвуют в связывании или необходимы для интерпретации инцидента;

3) исключение избыточных атрибутов, не влияющих на вычислимость, связывание и интерпретацию.

Профиль фиксируется как конечное множество полей (8):

$$F_{s,j} = P_{s,j} = \{f_1, \dots, f_{|F_{s,j}|}\}. \quad (8)$$

Для каждого атрибута $f \in F_{s,j}$ задаётся правило валидности $\mathrm{valid}_f(\cdot)$(проверка формата/типа/домена значений), необходимое для расчёта пригодности записи (разд. 5.3).

4. Построение опорного профиля интенсивности. В настоящей работе baseline строится на уровне класса события j, а не пары (s, j), исходя из предположения, что нормализация приводит события одного класса j к единой схеме независимо от источника s. Если в конкретном контуре мониторинга классы одного типа существенно различаются по интенсивности в зависимости от источника, baseline следует строить на уровне пары (s, j). Опорный профиль интенсивности строится по историческому интервалу (baseline-period) путём агрегации событий в окна длительности T с группировкой по индексу часа недели h и классу события j. Для каждой пары (h, j) формируется выборка:

$$\left\{N_{h,j}^{(k)}\right\}, k = 1, \dots, n_{h,j}, \quad (9)$$

где $N_{h,j}^{(k)}$ – число событий класса j в k-м окне, относящемся к h.

Далее вычисляются параметры профиля:

♦ типичный уровень интенсивности задаётся робастной оценкой центральной тенденции: (10):

$$MU_{h,j} = \mathrm{median}\{N_{h,j}^{(k)}\}, \quad (10)$$

♦ интервальные границы $[L_{h,j}, U_{h,j}]$ (на основе квантилей при фиксированном α) (11):

$$L_{h,j} = Q_{\alpha/2}\left\{N_{h,j}^{(k)}\right\}, U_{h,j} = Q_{1-\frac{\alpha}{2}}\left\{N_{h,j}^{(k)}\right\}. \quad (11)$$

Здесь Q_q– q-квантиль: значение, ниже которого находится доля q наблюдений (и доля 1-q – выше), что задаёт центральный интервал покрытия $1 - \alpha$ для популяционных квантилей и приближённое покрытие для эмпирических квантилей при конечной выборке [14, 15].

Параметр $\alpha \in (0,1)$ задаёт уровень покрытия интервальной оценки опорного профиля: интервал $[L_{h,j}, U_{h,j}]$ соответствует центральной части распределения и исключает хвосты суммарной массы порядка α (по $\alpha/2$ с каждой стороны). Уменьшение α расширяет интервал нормы и снижает чувствительность к редким выбросам; увеличение α сужает интервал и повышает чувствительность к отклонениям интенсивности (under/overlogging). В работе используется фиксированное $\alpha = 0.05$ (95%-покрытие) как общепринятый уровень в статистической практике для интервальных оценок и доверительных процедур, обеспечивающий баланс между устойчивостью и чувствительностью; значение α фиксируется и не изменяется между сценариями эксперимента [14, 15].

5. Расчёт компонент качества на окне T. Пусть $E_{j,T}$ – множество событий класса j, наблюдаемых в окне T, и $N_{h,j} = |E_{j,T}|$. Для каждого $j \in J_u$ вычисляются компоненты валидности времени $A_j(T)$, соответствия интенсивности $C_{prob}(h,j)$ (и $C_{old}(h,j)$) и структурной пригодности $K_j(T)$. Принадлежность записи к окну T определяется по времени поступления в индексатор $t_{ingest}(e) \in [T_{start}, T_{end}]$.

5.1. Валидность времени. Для события e используются метки $t_{event}(e)$ и $t_{ingest}(e)$. В мягком режиме валидность определяется допуском Δ. Здесь $parse$ – операция преобразования строкового представления временной метки в числовое значение (unix timestamp, секунды). Формула валидности события Δ (12):

$$A_e = \begin{cases} 1, & \left|parse\left(t_{event}(e)\right) - parse\left(t_{ingest}(e)\right)\right| \le \Delta, \\ 0, & \text{иначе.} \end{cases} \quad (12)$$

В жёстком режиме дополнительно проверяется принадлежность времени события анализируемому окну: запись считается валидной только если $t_{event} \in [T_{start}, T_{end}]$ и выполняется условие мягкого режима (12). Таким образом, для жёсткого режима (12а):

$$A_e{}^{hard} = \begin{cases} 1, & t_{event} \in [T_{start}, T_{end}] \wedge |t_{event} - t_{ingest}| \le \Delta, \\ 0, & \text{иначе.} \end{cases} \quad (12а)$$

Агрегирование по классу (13):

$$A_j(T) = \begin{cases} \frac{1}{|E_{j,T}|}\sum_{e \in E_{j,T}} A_e\,, |E_{j,T}| > 0, \\ 1, \quad |E_{j,T}| = 0. \end{cases} \quad (13)$$

Соглашение $A_j(T) = 1$ при $|E_{j,T}| = 0$ является нейтральным значением и обеспечивает корректность диагностической декомпозиции: при отсутствии событий весь дефицит относится к компоненте интенсивности ($C_{\text{prob}} = 0$), а не к временной валидности. Итоговое значение вклада $Q_{\text{new},j} = 1 \cdot 0 \cdot 1 = 0$ при этом не изменяется.

5.2. Соответствие интенсивности опорному профилю. Используются параметры $[L_{h,j}, U_{h,j}]$, $MU_{h,j}$ для соответствующего h. Коэффициент соответствия интенсивности расширенной методики (14):

$$C_{prob}(h,j) = \begin{cases} \frac{N_{h,j}}{L_{h,j}}, N_{h,j} < L_{h,j}, \\ 1, L_{h,j} \le N_{h,j} \le U_{h,j}\,, 0 \le C_{prob}(h,j) \le 1. \\ \frac{U_{h,j}}{N_{h,j}}, N_{h,j} > U_{h,j}, \end{cases} \quad (14)$$

Коэффициент базового показателя (15):

$$C_{old}(h,j) = \min\left(\frac{N_{h,j}}{MU_{h,j}}, 1\right) \quad (15)$$

При $MU_{h,j} = 0$ в базовом показателе используется фиксированное правило обработки: полагается $C_{old}(h,j) = 1$ при $N_{h,j} = 0$ и $C_{old}(h,j) = 0$ при $N_{h,j} > 0$. Формула (14) применяется при $L_{h,j} > 0$. Граничные случаи обрабатываются отдельно. Если $L_{h,j} = U_{h,j} = 0$, то полагается $C_{prob}(h,j) = 1$ при $N_{h,j} = 0$ и $C_{prob}(h,j) = 0$ при $N_{h,j} > 0$. Если $L_{h,j} = 0$, но $U_{h,j} > 0$, то значения $0 \le N_{h,j} \le U_{h,j}$считаются соответствующими опорному профилю, а при $N_{h,j} > U_{h,j}$ используется $C_{prob}(h,j) = U_{h,j}/N_{h,j}$. Таким образом, коэффициент $C_{prob}(h,j)$ корректно учитывает как недологирование, так и перелогирование, не штрафуя допустимые значения интенсивности при нулевой нижней границе интервального baseline.

5.3. Структурная пригодность записи к корреляции. Для пары (s,j) используется профиль $F_{s,j} = P_{s,j}$. Для события e показатель пригодности определяется как доля атрибутов профиля (16), которые присутствуют и валидны.

$$K_e = \frac{1}{|F_{s,j}|}\sum_{f \in F_{s,j}} 1\{present_f(e) \wedge valid_f(e)\}. \quad (16)$$

Здесь $present_f(e) = 1$, если атрибут f присутствует в записи e, и $present_f(e) = 0$ иначе. $valid_f(e) = 1$, если значение присутствующего атрибута удовлетворяет заданному правилу валидности $valid_f(\cdot)$, и $valid_f(e) = 0$ иначе.

Агрегирование по классу (17):

$$K_j(T) = \begin{cases} \frac{1}{|E_{j,T}|}\sum_{e \in E_{j,T}} K_e, \ |E_{j,T}| > 0, \\ 1, \quad |E_{j,T}| = 0. \end{cases} \quad (17)$$

Аналогично формуле (13), соглашение $K_j(T) = 1$ при $|E_{j,T}| = 0$ является нейтральным: при отсутствии событий нет материала для оценки структурной пригодности, а единственной диагностически значимой причиной дефицита является отсутствие потока ($C_{\text{prob}} = 0$). Итоговое значение вклада $Q_{\text{new},j} = 0$ при этом не изменяется.

6. Итоговые показатели $DL_{new}(T), DL_{old}(T)$. *6.1. Вклады качества классов.* Для каждого $j \in J_u$ вводятся вклады:

Агрегация компонент в вклад качества класса строится по мультипликативной схеме, поскольку все три свойства – соответствие интенсивности, временна́я валидность и структурная пригодность – являются необходимыми условиями пригодности события к корреляции [16]. Низкое качество по одному компоненту не должно компенсироваться высокими значениями по другим: событие с корректными полями, но невалидным timestamp, непригодно для корреляции так же, как и событие с валидным временем, но потерявшее ключевые атрибуты. Мультипликативная схема обеспечивает это свойство: $Q_{new,j} = 0$при нулевом значении любой из компонент.

$$Q_{new,j}(T) = A_j(T) \cdot C_{prob}(h,j) \cdot K_j(T). \quad (18)$$

$$Q_{old,j}(T) = A_j(T) \cdot C_{old}(h,j). \quad (19)$$

6.2. Сценарно- и риск-взвешенная агрегация. Итоговые показатели дефицита логирования:

$$DL_{new}(T) = 1 - \sum_{j \in J_u} w_{u,j} \cdot Q_{new,j}(T). \quad (20)$$

$$DL_{old}(T) = 1 - \sum_{j \in J_u} w_{u,j} \cdot Q_{old,j}(T). \quad (21)$$

Дополнительно вводится сценарный показатель дефицита логирования $DL_{scn}(T)$, вычисляемый с равномерными весами (без риск-взвешивания) по компонентам расширенного показателя:

$$DL_{scn}(T) = 1 - \frac{1}{|J_u|} \sum_{j \in J_u} Q_{new,j}(T). \quad (22)$$

Показатель $DL_{scn}(T)$ использует усреднение без риск-взвешивания и служит контрольной точкой сравнения: отклонение DL_{new} от DL_{scn} свидетельствует о влиянии риск-взвешивания на итоговую оценку дефицита логирования.

6.3. Диагностическая декомпозиция причин ухудшения качества логирования. Для диагностики причин ухудшения интегрального показателя $DL_{\text{new}}(T)$ вводятся три парциальных показателя дефицита по отдельным компонентам качества. Для каждого класса $j \in J_u$ определяются парциальные дефициты по компонентам интенсивности (23), временной валидности (24) и структурной пригодности (25):

$$d_{\text{C}}(j) = 1 - C_{prob}(h,j) \quad (23)$$

$$d_{\text{A}}(j) = 1 - A_j(T) \quad (24)$$

$$d_{\text{K}}(j) = 1 - K_j(T) \quad (25)$$

На основе парциальных дефицитов (23)–(25) строятся сценарно-взвешенные агрегированные показатели по каждой компоненте:

$$DL_C(T) = \sum_{j \in J_u} w_{u,j} \cdot d_{\text{C}}(j) \quad (26)$$

$$DL_A(T) = \sum_{j \in J_u} w_{u,j} \cdot d_{\text{A}}(j) \quad (27)$$

$$DL_K(T) = \sum_{j \in J_u} w_{u,j} \cdot d_{\text{K}}(j) \quad (28)$$

Показатели $DL_C(T)$, $DL_A(T)$, $DL_K(T)$ являются диагностическими и не суммируются в $DL_{new}(T)$. Это обусловлено мультипликативной природой формулы (18): при одновременной деградации нескольких компонент итоговый дефицит меньше суммы парциальных показателей. Назначение декомпозиции – выделение доминирующего компонента деградации по принятому диагностическому правилу. Компонента X рассматривается как доминирующий диагностический компонент ухудшения DL_{new}, если:

$$DL_X(T) = \max(DL_C, DL_A, DL_K).$$

Класс событий j признаётся основным источником дефицита по компоненте X, если взвешенный вклад $w_{u,j} \cdot d_X(j)$ является наибольшим среди всех классов сценария. Декомпозиция (23)–(28) позволяет количественно ответить на вопрос: какая именно компонента качества является приоритетным источником ухудшения $DL_{\text{new}}(T)$ и в каком классе событий это ухудшение максимально, что обеспечивает обоснованную приоритизацию корректирующих мер [17, 18].

7. Практическая апробация разработанной методики. Практическая апробация выполнялась на стенде на базе SIEM-системы Wazuh (версия 4.x) с хранением событий в OpenSearch [19]. Генерация тестовых потоков осуществлялась программным генератором событий, записывающим нормализованные JSON-записи в файл, доступный агенту Wazuh. Далее события проходили полный конвейер обработки: сбор агентом, нормализацию декодером, индексацию в OpenSearch по индексу wazuh-alerts-4.x-*. Построение *baseline* и расчёт показателей DL выполнялись непосредственно по данным OpenSearch через API агрегаций. Цель эксперимента состояла в проверке того, что предложенные показатели $DL_{new}(T)$ и $DL_{scn}(T)$ адекватно отражают изменение качества логирования при различных типах нарушений и позволяют оценивать не только полноту поступления событий, но и их пригодность к дальнейшей корреляции.

Эксперимент был построен в соответствии со структурой предложенной методики. На первом этапе задавалось сценарное множество контролируемых классов событий:

$$J_u = \{AUTH\_OK,\ AUTH\_FAIL,\ NET\_CONN,\ PRIV\_CHANGE\}.$$

Выбор именно этих классов обусловлен тем, что они образуют минимально достаточный набор событий для описания типового сценария анализа действий субъекта в информационной системе. В указанное множество входят события успешной и неуспешной аутентификации, события сетевого взаимодействия, а также события, связанные с изменением привилегий. Отсутствие либо искажение этих классов непосредственно снижает возможность последующей корреляции и интерпретации событий в SIEM-системе. Риск-веса классов в эксперименте определялись по процедуре раздела 2 на основе единственного корреляционного правила, использующего все четыре класса, при этом классы AUTH_FAIL и PRIV_CHANGE признавались обязательными (H=1), а AUTH_OK и NET_CONN – контекстными (H=0). Итоговые нормированные веса составили:

$$w(AUTH\_FAIL) = 0{,}333, w(PRIV\_CHANGE) = 0{,}333, w(AUTH\_OK) = 0{,}167,$$
$$w(NET\_CONN) = 0{,}167.$$

На втором этапе формировался базовый режим наблюдения, относительно которого далее оценивались отклонения. Базовый режим соответствовал штатному состоянию потока событий и использовался как опорная точка для сравнения с режимами управляемого нарушения качества логирования. В эксперименте были зафиксированы следующие параметры методики: $T = 5$ мин; $\Delta = 300$ с (5 мин); режим временной валидности – мягкий (формула 12); $\alpha = 0{,}05$ (95%-покрытие); baseline-период – 240 мин (48 окон по 5 мин). Индекс часа недели h не применялся, поскольку поток был однородным и не имел суточно-недельной цикличности; поэтому для всех окон использовался единый профиль $[L_j, U_j, MU_j]$.

Таблица 1

Минимально достаточные профили атрибутов $P_{s,j}$ и параметры сценария *«field_loss»*

Класс *j*	**Профиль $F_{s,j}$ (обязательные поля)**	**Доля потери в *field_loss***
AUTH_OK	event_time, host, user, src_ip, outcome	~50%
AUTH_FAIL	event_time, host, user, src_ip, outcome	~75%
NET_CONN	event_time, host, src_ip, dst_ip, dst_port, proto, direction	~60%
PRIV_CHANGE	event_time, host, user, action	~25%

В данной апробации для всех атрибутов профиля использовалось упрощённое правило валидности: значение считается валидным, если оно присутствует, не равно null и не является пустой строкой.

На третьем этапе были последовательно сформированы четыре типа нарушений качества логирования. Сценарий *«underlogging»* моделировал частичную утрату событийного потока, при которой часть значимых сообщений не поступает в систему анализа. Сценарий

«timestamp_drift» имитировал нарушение временной валидности, когда события перестают корректно соотноситься с анализируемым окном и становятся непригодными для надёжной временной корреляции. Сценарий *«field_loss»* отражал потерю обязательных атрибутов записи, необходимых для идентификации субъекта, объекта, процесса или контекста события (состав профилей и доля потери полей приведены в табл. 1). Сценарий *«duplicates»* задавал избыточную генерацию и дублирование записей, что позволяло проверить устойчивость предложенной методики не только к дефициту потока, но и к его переизбыточности.

Для интерпретации получаемых значений DL в практической эксплуатации предлагается ориентировочная шкала: DL $\in$ [0, 0,2] – поток пригоден для корреляции, отклонения незначительны; DL $\in$ (0,2, 0,5] – умеренная деградация, требуется анализ компонент; DL $\in$ (0,5, 0,8] – существенная деградация, корреляция по сценарию ненадёжна; DL $\in$ (0,8, 1,0] – критическая непригодность потока. Указанные пороги носят эвристический характер и не имеют строгого статистического обоснования. Они предложены как отправная точка для практической эксплуатации и подлежат уточнению на основе характеристик конкретного SOC, применяемого детект-контента и накопленной статистики прогонов. Результаты экспериментальных прогонов представлены в табл. 2.

Таблица 2

Результаты экспериментальной проверки разработанной методики

Режим эксперимента	**Характер моделируемого нарушения**	DL_{old}	DL_{new}	DL_{scn}	DL_C	DL_A	DL_K	**Интерпретация результата**
Baseline	Штатный поток событий	0,297	0,000	0,000	0,000	0,000	0,000	Новая методика не даёт ложных срабатываний на нормальный поток
Underlogging	Частичная потеря значимых событий	0,915	0,858	0,870	0,858	0,000	0,000	Доминант: интенсивность; все 4 класса зафиксировали дефицит
Timestamp drift	Нарушение временной валидностисобытий	1,000	1,000	1,000	0,000	1,000	0,000	Поток становится практически непригодным для корреляции
Field loss	Потеря обязательных атрибутов записи	0,539	0,458	0,512	0,125	0,000	0,394	Доминант: структурная пригодность
Duplicates	Дублирование и избыточная генерация записей	0,000	0,073	0,084	0,073	0,000	0,000	Новая методика учитывает не только дефицит, но и перелогирование

Для базового (опорного) режима получены значения $DL_{old} = 0{,}297$, $DL_{new} = 0{,}000$, $DL_{scn} = 0{,}000$. Нулевые значения DL_{new} и DL_{scn} при штатном потоке подтверждают, что предложенный показатель не даёт ложных срабатываний на нормальные потоки интенсивности: все четыре класса J_u присутствовали в окне наблюдения, а наблюдаемые значения N попали в интервал $[L, U]$ опорного профиля. Ненулевое значение $DL_{old} = 0{,}297$ объясняется тем, что базовый показатель нормирует на медиану MU, штрафуя за любое отклонение от типичного уровня вниз, тогда как новый показатель считает поток нормальным при $N \in [L, U]$.

В режиме «*underlogging»* все три показателя увеличиваются: $DL_{old} = 0{,}915$, $DL_{new} = 0{,}858$, $DL_{scn} = 0{,}870$. Основной вклад в ухудшение даёт интенсивностный компонент $DL_C = 0{,}858$, тогда как $DL_A = 0$ и $DL_K = 0$. Следовательно, в данном режиме снижение качества связано с потерей части событийного потока, а не с искажением временных меток или структуры записей. При этом $DL_{old} > DL_{new}$, что объясняется различием в формулах интенсивностного компонента: в базовом показателе используется нормировка на медиану MU, а в новой – на нижнюю границу L. Поскольку $L \leq MU$, при $N < L$ новый показатель даёт более сдержанную оценку, не трактуя каждое снижение относительно медианы как дефект логирования.

В режиме «*timestamp_drift*» значения DL_{old}, DL_{new}и DL_{scn}достигают 1,000. Это соответствует полной непригодности потока к корреляции в пределах анализируемого окна: даже при сохранении части записей потеря корректной временной привязки делает их ненадёжными для сопоставления.

В режиме «*field_loss*» получены значения $DL_{old} = 0{,}539$, $DL_{new} = 0{,}458$, $DL_{scn} = 0{,}512$. Наибольший вклад здесь вносит структурная компонента: $DL_K = 0{,}394$ при $DL_C = 0{,}125$ и $DL_A = 0{,}000$. Это указывает на то, что основной источник ухудшения – неполнота корреляционно-значимых атрибутов записи. Базовый показатель фиксирует общее ухудшение, но не позволяет различить его причину. В этом режиме преимущество нового показателя проявляется наиболее явно: снижение качества логирования выявляется даже тогда, когда проблема связана не с числом событий, а с потерей корреляционно-значимой информации.

В режиме «*duplicates*» получены значения $DL_{old} = 0{,}000$, $DL_{new} = 0{,}073$, $DL_{scn} = 0{,}084$. Для базового показателя этот режим остаётся «нормальным», поскольку он ориентирован на дефицит потока и не рассматривает дублирование как самостоятельный дефект. В новом показателе превышение верхней границы профиля ($N > U$)отражается в ненулевом значении $DL_C = 0{,}073$, тогда как $DL_A = 0{,}000$ и $DL_K = 0{,}000$. Тем самым показатель фиксирует именно интенсивностное искажение потока, обусловленное перелогированием.

В целом проведённая апробация показывает, что методика различает четыре типа нарушений качества логирования: потерю событий, нарушение временной валидности, потерю обязательных атрибутов и дублирование. Для каждого режима диагностическая декомпозиция выделяет компонент, вносящий основной вклад в итоговый дефицит. Кроме того, новый показатель обнаруживает дефекты, которые не отражаются базовым показателем: в режиме «*duplicates*» $DL_{old} = 0$ при $DL_{new} > 0$, а в baseline-режиме, наоборот, $DL_{new} = 0$ при $DL_{old} > 0$. Это позволяет использовать DL_{new} и DL_{scn} как более адекватные показатели пригодности событийного потока к корреляции в условиях практической эксплуатации SIEM.

Рис. 1. Сравнение DL-показателей по режимам эксперимента

Результаты, представленные на рис. 1, показывают различия между показателями для разных типов нарушений. В baseline-режиме ненулевое значение принимает только базовый показатель: $DL_{old} = 0{,}297$, тогда как $DL_{new} = 0{,}000$ и $DL_{scn} = 0{,}000$. Это указывает на отсутствие ложных срабатываний нового показателя на штатный поток. В режимах «*underlogging*» и «*timestamp_drift*» значения всех трёх показателей возрастают; при «*timestamp_drift*» они достигают 1,000, что соответствует полной непригодности потока к корреляции событий ИБ. Для режима «*field_loss*» значения $DL_{new} = 0{,}458$ и $DL_{scn} = 0{,}512$отражают структурную деградацию записи при умеренном значении $DL_{old} = 0{,}539$. В режиме «*duplicates*» значение DL_{old} остаётся нулевым, тогда как DL_{new} и DL_{scn} принимают малые, но ненулевые значения. Это показывает нечувствительность базового показателя к перелогированию и способность предложенной методики фиксировать превышение верхней границы опорного профиля.

Заключение. В работе разработана и апробирована методика количественной оценки дефицита логирования в SIEM, ориентированная на оценку пригодности событийного потока к последующей корреляции. Предложенная методика учитывает сценарный состав контролируемых классов событий, их относительную значимость, отклонение интенсивности потока от опорного профиля, временную валидность событий и полноту корреляционно-значимых атрибутов записи [20].

Проверка методики на тестовых режимах показала, что предложенная методика позволяет выявлять различные типы деградации логирования, включая недологирование, нарушение временной валидности, потерю корреляционно-значимых атрибутов и дублирование событий. Полученные результаты подтвердили, что оценка качества логирования должна учитывать не только наличие событий, но и их пригодность к аналитической обработке и корреляции в рамках выбранного сценария детектирования. В частности, в сценарии дублирования $DL_{old} = 0{,}000$ при $DL_{new} = 0{,}073$, что количественно подтверждает слепоту базового показателя к перелогированию; в baseline-режиме $DL_{new} = 0{,}000$ при $DL_{old} = 0{,}297$, что исключает ложные срабатывания на нормальный поток.

Тем самым поставленная цель работы достигнута в пределах проведённой экспериментальной апробации, а предложенная методика может рассматриваться как инструмент для обоснованной оценки качества логирования и приоритизации доработок источников событий, процедур нормализации и правил корреляции в SIEM-системах. Полученные результаты относятся к условиям синтетического стенда, фиксированного набора классов событий и заданных параметров Δ, α и T; дальнейшее развитие работы связано с проверкой методики на нескольких сценариях и на более разнородных реальных потоках SIEM. Научная новизна состоит в формализации дефицита логирования как интегральной меры пригодности событийного потока к корреляции и введении диагностической декомпозиции (DL_C, DL_A, DL_K), позволяющей выделять доминирующий диагностический компонент деградации. В отличие от ранее предложенного показателя DL [10], ориентированного на интегральную оценку полноты и корректности логирования, в настоящей работе вводится компонентное представление качества потока, включающее соответствие интенсивности опорному профилю, валидность времени и структурную пригодность записи к корреляции. Это позволяет выявлять качественно различные типы деградации потока, в том числе перелогирование/дублирование и потерю корреляционно-значимых атрибутов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. NIST SP 800-92. Руководство по управлению журналами событий компьютерной безопасности. Национальный институт стандартов и технологий США, 2006. – URL: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-92.pdf.
2. *Гонсалес-Гранадильо Г., Гонсалес-Сарсоса С., Диас Р.* Системы управления информацией и событиями безопасности (SIEM): анализ, тенденции и применение в критических инфраструктурах // Sensors. – 2021. – Т. 21, № 14. – Ст. 4759. – DOI: 10.3390/s21144759.
3. CIS Controls v8.1. Контроль 8: Управление журналами аудита. Центр интернет-безопасности, 2021. – URL: https://cas.docs.cisecurity.org/en/latest/source/Controls8/.
4. Elastic Filebeat Documentation: Дедупликация данных (доставка типа at-least-once). Elastic N.V., 2024. – URL: https://www.elastic.co/guide/en/beats/filebeat/8.19/filebeat-deduplication.html.
5. Elastic Filebeat Documentation: Ротация журналов приводит к потере или дублированию событий. Elastic N.V., 2024. – URL: https://www.elastic.co/guide/en/beats/filebeat/8.19/file-log-rotation.html.
6. *Студиаван Х., Сохел Ф., Пэйн К.* Обзор методов криминалистического исследования журналов операционных систем // Digital Investigation. – 2019. – Т. 29. – С. 1-20. – DOI: 10.1016/j.diin.2019.02.005.
7. *Копполино Л., Д'Антонио С., Формикола В., Романо Л.* Совершенствование технологии SIEM для защиты критических инфраструктур // Critical Information Infrastructures Security (CRITIS 2011). LNCS. Т. 6983. – Springer, 2013. – С. 10-21. – DOI: 10.1007/978-3-642-41476-3_2.
8. *Котенко И.В., Гайфулина Д.А., Зеличенок И.И.* Систематический обзор методов корреляции событий безопасности // IEEE Access. – 2022. – Т. 10. – С. 43387-43420. – DOI: 10.1109/ACCESS.2022.3168976.
9. *Чуа Э., Калутараж Х., Тасдемир К., Абрахам А., Мэйпл К.* Систематический обзор инструментов лог-корреляции для обнаружения и прогнозирования кибератак в крупных сетях // Journal of Information Security and Applications. – 2025. – Т. 92. – Ст. 104096. – DOI: 10.1016/j.jisa.2025.104096.
10. *Иванов А.В., Киселев М.А.* Методика оценки качества логирования событий информационной безопасности в итерационно-управленческом цикле с применением метрики дефицита логирования // Доклады ТУСУРа. – 2026. – Т. 29, № 1. – С. 114-123. – DOI: 10.21293/1818-0442-2026-29-1-114-123.
11. *Герхардс Р.* Протокол Syslog. RFC 5424 // IETF. – 2009. – URL: https://datatracker.ietf.org/doc/html/rfc5424.
12. *Фишер Д., Гоэл К., Эндрюс Р. и др.* Повышение качества журналов событий: обнаружение и количественная оценка дефектов временных меток // Business Process Management (BPM 2020). LNCS. Т. 12168. – Springer, 2020. – С. 309-326. – DOI: 10.1007/978-3-030-58666-9_18.

13. *Ван Р.Й., Стронг Д.М.* За пределами точности: что качество данных означает для потребителей // Journal of Management Information Systems. – 1996. – Т. 12, № 4. – С. 5-33. – DOI: 10.1080/07421222.1996.11518099.
14. NIST/SEMATECH. Электронный справочник по статистическим методам. Мерки разброса: межквартильный размах. – URL: https://itl.nist.gov/div898/handbook/eda/section3/eda352.htm.
15. NIST/SEMATECH. Электронный справочник по статистическим методам. Выборочные квантили. – URL: https://itl.nist.gov/div898/software/dataplot/refman2/auxillar/quantile.htm.
16. *Левшун Д.С., Котенко И.В.* Обзор методов искусственного интеллекта для корреляции событий безопасности: модели, вызовы и возможности // Artificial Intelligence Review. – 2023. – Т. 56. – С. 8547-8590. – DOI: 10.1007/s10462-022-10381-4.
17. *Котенко И.В., Федорченко А.В., Дойникова Е.В.* Аналитика данных для управления безопасностью сложных гетерогенных систем: задачи корреляции событий и оценки защищённости // Advances in Cyber Security Analytics and Decision Systems. – Springer, Cham, 2020. – С. 79-116. – DOI: 10.1007/978-3-030-19353-9_5.
18. *Брайант Б.Д., Саедиан Х.* Улучшение агрегации метаданных алертов SIEM с использованием новой классификационной модели на основе kill-chain // Computers & Security. – 2020. – Т. 94. – Ст. 101783. – DOI: 10.1016/j.cose.2020.101783.
19. *Шираз М. и др.* Эффективный мониторинг безопасности с использованием оптимизированной архитектуры SIEM // Human-centric Computing and Information Sciences. – 2023. – Т. 13. – Ст. 17. – DOI: 10.22967/HCIS.2023.13.017.
20. *Эрлингер Л., Восс В.* Обзор инструментов измерения и мониторинга качества данных // Frontiers in Big Data. – 2022. – Т. 5. – Ст. 850611. – DOI: 10.3389/fdata.2022.850611.

REFERENCES

1. NIST SP 800-92. Rukovodstvo po upravleniyu zhurnalami sobytiy komp'yuternoy bezopasnosti. Natsional'nyy institut standartov i tekhnologiy SShA, 2006 [NIST SP 800-92. Guide to Computer Security Log Management. National Institute of Standards and Technology, 2006]. Available at: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-92.pdf.
2. Gonsales-Granadil'o G., Gonsales-Sarsosa S., Dias R. Sistemy upravleniya informatsiey i sobytiyami bezopasnosti (SIEM): analiz, tendentsii i primenenie v kriticheskikh infrastrukturakh [Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures], *Sensors*, 2021, Vol. 21, No. 14. Art. 4759. DOI: 10.3390/s21144759.
3. CIS Controls v8.1. Kontrol' 8: Upravlenie zhurnalami audita. TSentr internet-bezopasnosti, 2021 [CIS Controls v8.1. Control 8: Audit Log Management. Center for Internet Security, 2021]. Available at: https://cas.docs.cisecurity.org/en/latest/source/Controls8/.
4. Elastic Filebeat Documentation: Deduplikatsiya dannykh (dostavka tipa at-least-once). Elastic N.V., 2024 [Elastic Filebeat Documentation: Deduplicate data (at-least-once duplicates). Elastic N.V., 2024]. Available at: https://www.elastic.co/guide/en/beats/filebeat/8.19/filebeat-deduplication.html.
5. Elastic Filebeat Documentation: Rotatsiya zhurnalov privodit k potere ili dublirovaniyu sobytiy. Elastic N.V., 2024 [Elastic Filebeat Documentation: Log rotation results in lost or duplicate events. Elastic N.V., 2024]. Available at: https://www.elastic.co/guide/en/beats/filebeat/8.19/file-log-rotation.html.
6. *Studiavan Kh., Sokhel F., Peyn K.* Obzor metodov kriminalisticheskogo issledovaniya zhurnalov operatsionnykh sistem [A survey on forensic investigation of operating system logs], *Digital Investigation*, 2019, Vol. 29, pp. 1-20. DOI: 10.1016/j.diin.2019.02.005.
7. *Koppolino L., D'Antonio S., Formikola V., Romano L.* Sovershenstvovanie tekhnologii SIEM dlya zashchity kriticheskikh infrastruktur [Enhancing SIEM technology to protect critical infrastructures], *Critical Information Infrastructures Security (CRITIS 2011)*. LNCS. Vol. 6983. Springer, 2013, pp. 10-21. DOI: 10.1007/978-3-642-41476-3_2.
8. *Kotenko I.V., Gayfulina D.A., Zelichenok I.I.* Sistematicheskiy obzor metodov korrelyatsii sobytiy bezopasnosti [Systematic literature review of security event correlation methods], *IEEE Access*, 2022, Vol. 10, pp. 43387-43420. DOI: 10.1109/ACCESS.2022.3168976.
9. *Chua E., Kalutarazh Kh., Tasdemir K., Abrakham A., Meypl K.* Sistematicheskiy obzor instrumentov log-korrelyatsii dlya obnaruzheniya i prognozirovaniya kiberatak v krupnykh setyakh [A systematic literature review of log-correlation tools for cyberattack detection and prediction in large networks], *Journal of Information Security and Applications*, 2025, Vol. 92. Art. 104096. DOI: 10.1016/j.jisa.2025.104096.
10. *Ivanov A.V., Kiselev M.A.* Metodika otsenki kachestva logirovaniya sobytiy informatsionnoy bezopasnosti v iteratsionno-upravlencheskom tsikle s primeneniem metriki defitsita logirovaniya [Methodology for assessing the quality of information security event logging in an iterative management cycle using the logging deficit metric], *Doklady TUSURa* [Proceedings of TUSUR University], 2026, Vol. 29, No. 1, pp. 114-123. DOI: 10.21293/1818-0442-2026-29-1-114-123.
11. *Gerkhards R.* Protokol Syslog. RFC 5424 [The Syslog Protocol. RFC 5424], *IETF*, 2009. Available at: https://datatracker.ietf.org/doc/html/rfc5424.

12. *Fisher D., Goel K., Endryus R. i dr.* Povyshenie kachestva zhurnalov sobytiy: obnaruzhenie i kolichestvennaya otsenka defektov vremennykh metok [Enhancing event log quality: detecting and quantifying timestamp imperfections], *Business Process Management (BPM 2020)*. LNCS. Vol. 12168. Springer, 2020, pp. 309-326. DOI: 10.1007/978-3-030-58666-9_18.
13. *Van R.Y., Strong D.M.* Za predelami tochnosti: chto kachestvo dannykh oznachaet dlya potrebiteley [Beyond accuracy: what data quality means to data consumers], *Journal of Management Information Systems*, 1996, Vol. 12, No. 4, pp. 5-33. DOI: 10.1080/07421222.1996.11518099.
14. NIST/SEMATECH. Elektronnyy spravochnik po statisticheskim metodam. Merki razbrosa: mezhkvartil'nyy razmakh [NIST/SEMATECH e-Handbook of statistical methods. Measures of scale: interquartile range]. Available at: https://itl.nist.gov/div898/handbook/eda/section3/eda352.htm.
15. NIST/SEMATECH. Elektronnyy spravochnik po statisticheskim metodam. Vyborochnye kvantili [NIST/SEMATECH e-Handbook of statistical methods. Sample quantiles]. Available at: https://itl.nist.gov/div898/software/dataplot/refman2/auxillar/quantile.htm.
16. *Levshun D.S., Kotenko I.V.* Obzor metodov iskusstvennogo intellekta dlya korrelyatsii sobytiy bezopasnosti: modeli, vyzovy i vozmozhnosti [A survey on artificial intelligence techniques for security event correlation: models, challenges, and opportunities], *Artificial Intelligence Review*, 2023, Vol. 56, pp. 8547-8590. DOI: 10.1007/s10462-022-10381-4.
17. *Kotenko I.V., Fedorchenko A.V., Doynikova E.V.* Analitika dannykh dlya upravleniya bezopasnost'yu slozhnykh geterogennykh sistem: zadachi korrelyatsii sobytiy i otsenki zashchishchennosti [Data analytics for security management of complex heterogeneous systems: event correlation and security assessment tasks], *Advances in Cyber Security Analytics and Decision Systems*. Springer, Cham, 2020, pp. 79-116. DOI: 10.1007/978-3-030-19353-9_5.
18. *Brayant B.D., Saedian Kh.* Uluchshenie agregatsii metadannykh alertov SIEM s ispol'zovaniem novoy klassifikatsionnoy modeli na osnove kill-chain [Improving SIEM alert metadata aggregation with a novel kill-chain based classification model], *Computers & Security*, 2020, Vol. 94, Art. 101783. DOI: 10.1016/j.cose.2020.101783.
19. *Shiraz M. i dr.* Effektivnyy monitoring bezopasnosti s ispol'zovaniem optimizirovannoy arkhitektury SIEM [Effective security monitoring using efficient SIEM architecture], *Human-centric Computing and Information Sciences*, 2023, Vol. 13, Art. 17. DOI: 10.22967/HCIS.2023.13.017.
20. *Erlinger L., Voss V.* Obzor instrumentov izmereniya i monitoringa kachestva dannykh [A survey of data quality measurement and monitoring tools], *Frontiers in Big Data*, 2022, Vol. 5, Art. 850611. DOI: 10.3389/fdata.2022.850611.

Киселев Максим Алексеевич – Новосибирский государственный технический университет; e-mail: asuszenfone_2@mail.ru; г. Новосибирск, Россия; тел.: +79511613045; аспирант.

Kiselev Maksim Alexeevich – Novosibirsk State Technical University; e-mail: asuszenfone_2@mail.ru; Novosibirsk, Russia; phone: +79511613045; the Department of Information Protection; postgraduate student.