ISSN 2658-5014 (Print)
2025. Т. 7, № 1 (20). С. 3–12

СИИТ
СИСТЕМНАЯ ИНЖЕНЕРИЯ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
Научная статья

ISSN 2686-7044 (Online)
http://siit.ugatu.su

УДК 004.89
DOI 10.54708/2658-5014-SIIT-2025-no1-p3
EDN HWZXTV

Система обнаружения аномалий в журналах регистрации событий средств защиты информации

Е. А. Атарская • А. М. Вульфин • А. Д. Кириллова

Аннотация. В работе предложены модели и алгоритмы обнаружения аномалий в функционировании информационной системы на основе анализа журналов регистрации событий средств защиты информации с помощью нейросетевых моделей. Целью исследования является совершенствование моделей и алгоритмов анализа текстовых журналов регистрации событий для повышения оперативности выявления аномальных состояний системы. Научная новизна предлагаемых алгоритма предобработки, модели анализа и алгоритма обнаружения аномалий основана на применении методов и моделей глубокого машинного обучения, отличительной особенностью является совокупность этапов обработки и извлечения признаков, а также применение комитета нейросетевых моделей — автоэнкодера с долгой краткосрочной памятью в сочетании с моделями машинного обучения обнаружения выбросов для адаптивного выделения аномальных цепочек событий. Это позволяет повысить эффективность анализа текстовых журналов работы системы и, следовательно, повысить оперативность выявления аномальных состояний компонент информационной системы.

Ключевые слова: анализ журналов событий; глубокое обучение; нейросетевой автоэнкодер; средства защиты информации; обнаружение аномалий и выбросов.

Введение

Значительные объемы накапливаемых данных мониторинга состояния компонентов информационный системы (ИС) в виде текстовых журналов работы системы (логов) либо не анализируются вовсе, либо подвергаются анализу с помощью ограниченного набора правил. Более глубокий анализ ограничен возможностями систем разбора (парсеров) и последующего сопряжения с SIEM (Security Information and Event Management) системой с помощью настраиваемых индивидуально программных коннекторов. Дальнейший анализ ориентирован, в первую очередь, на выявление однозначно интерпретируемых записей, вызванных событиями информационной безопасности (ИБ), помеченных как потенциально опасные для ИС.

Существенное количество записей журнала регистрации событий (ЖРС) не имеют разметки принадлежности к потенциально опасным маркерам, характеризующим действия возможного злоумышленника. Анализ ЖРС позволяет выявлять основные типы происходящих событий, группировать их по степени сходства, отслеживать изменения (динамику) смены, характерные для системы типов событий во времени, а также выделять нетипичные (аномальные – новые) события для заданного временного окна анализа. Выделенные в ходе анализа нетипичные события могут быть вызваны как внутрисистемными процессами, характеризующими штатный режим работы, так и сбоями и нарушениями нормального функционирования компонентов ИС ввиду ошибок конфигурации, обновления программного обеспечения или внешними причинами. Количество подобных событий относительно общего объема, находящих отражение в ЖРС, невелико, однако их обнаружение требует значительных усилий по анализу накапливаемых данных мониторинга. Следовательно, повышение эффективности анализа за счет автоматизации разбора, формализации и интеллектуального анализа данных

Рекомендовано к публикации программным комитетом X Международной научной конференции ITIDS’2024 «Информационные технологии интеллектуальной поддержки принятия решений», Уфа, 12–14 ноября 2024 г.

является актуальной задачей, направленной на повышение оперативности выявления аномальных состояний компонент информационной системы.

Цель работы – совершенствование моделей и алгоритмов анализа журналов регистрации событий средств защиты информации для повышения оперативности выявления аномальных состояний и/или событий в информационной системе.

1. Обзор известных подходов к анализу ЖРС на основе технологий искусственного интеллекта

Методы и технологии обнаружения аномалий на основе анализа ЖРС широко используются для диагностики и устранения неисправностей в сложных программно-интенсивных системах. ЖРС представляет собой полуструктурированный текст, который содержит различные поля. Одним из важнейших этапов является разбор журнала для автоматического преобразования каждого сообщения в определенный шаблон события – набора пар «key» (постоянная часть) и «value» (переменная часть).

В табл. 1 приводится сравнение известных моделей анализа ЖРС и обнаружений аномалий в них. Вследствие использования различных наборов данных и метрик для оценки невозможно выделить наиболее эффективный метод или алгоритм.

Таблица 1

Обзор моделей и алгоритмов обнаружения аномалий на основе анализа журналов событий

Технология, год	Особенности	Методы ML	Datasets	Оценка
log2vec [1], 2019	Модульный метод, основанный на встраивании графов	GNN, алгоритмы кластеризации	—	0.93 AUC
CFDet [2], 2022	Модель на основе анализа распределений; не требует размеченного набора данных для обучения	Deep SVDD, One Class SVM, iForest	BGL, Thunderbird, CERT	0.778–0.984 F1
DeepSyslog [3], 2022	Встраивание предложений для извлечения семантической и контекстной информации	LSTM	BGL, HDFS	0.91–0.98 F1
ADLILog [4], 2022	Построение системы обнаружения аномалий на основе логов инструкций; не требует размеченного набора данных для обучения	Трансформеры	BGL, HDFS	0.61–0.98 F1
LogLR [5], 2022	Метод на основе логических рассуждений	LTN, LSTM	—	—
NeuralLog [6], 2021	Извлекает семантический смысл записей журналов событий с помощью BERT-кодера; классификация на основе трансформеров; не требует разбора журнала событий	BERT, Трансформеры	BGL, Thunderbird, HDFS,	0.96–0.98 F1
Log2Vec [7], 2020	Встраивание слов OOV в векторы во время выполнения	—	Spark, HDFS, Windows, Hadoop	0.925–0.959 F1
LightLog [8], 2022	Устройства IoT; низкоразмерное семантическое векторное пространство на основе word2vec; облегченная TCN для обнаружения	TCN	BGL, HDFS	0.970–0.972 F1
AutoLog [9], 2021	Отбор образцов журналов событий через равномерные интервалы времени и вычисление оценки для классификации; не требует размеченного набора данных для обучения	Глубокий автоэнкодер	BG/L, Hadoop	0.95–0.97 F1
LogFormer [10], 2023	Содержит этап предварительного обучения и этап настройки на основе адаптера	—	HDFS, BGL, Thunderbird	0.99–0.98 F1
LayerLog [11], 2023	Обнаружение аномалий последовательности записей в журнале событий, основанное на иерархической семантике данных ЖРС	—	HDFS, BGL	—

Однако в исследованиях появляется тенденция применения моделей трансформеров, использующих механизмы внимания, для улучшения классификации и обнаружения аномалий.

Существующие методы обнаружения аномалий на основе анализа ЖРС сталкиваются со следующими ограничениями:

- При извлечении последовательности записей событий из журнала в виде временных векторов сохраняется информация только о времени между событиями. Затруднено отслеживание причинно-следственных связей между записями ЖРС.
- Не учитывается контекст каждого события и метаданные события в журнале.
- Не размечены аномальные записи/последовательности для обучения модели обнаружения аномалий.
- В ходе разбора ЖРС возникают ошибки из-за слов, не входящих в словарный запас (OOV – out-of-vocabulary), и семантических несоответствий ввиду специфичности конкретной области.
- Высокая размерность получаемого признакового пространства.

Следовательно, общая схема применения глубоких нейросетевых моделей для обнаружения аномалий на основе анализа ЖРС состоит из четырех этапов: разбор журнала; группировка событий журнала; представление журнала в векторном пространстве признаков; обнаружение аномалий (выбросов, новых событий) с помощью нейросетевых моделей.

Таким образом, перспективным направлением построения моделей обнаружения аномалий на основе анализа слабоструктурированных данных ЖРС является применение технологий нейросетевой обработки с помощью, в том числе, моделей трансформеров.

2. Структурно-функциональная организация системы обнаружения аномалий в ЖРС средств защиты информации

На рис. 1 представлена схема предлагаемого алгоритма обработки потока текстовых событий, включающая следующие основные шаги:

1) Исходный поток событий s_1, s_2, …, s_p разбивается на группы с помощью скользящего окна W_n с шагом W_s.

2) Каждое s из попавших в текущее окно W_n анализа событий с помощью набора регулярных выражений, предварительно подготовленных в ходе разбора формата представления записей и построения грамматики разбора, преобразуется к виду (k_i, v_i), где k_i — токен постоянной части записи; v_i – токен переменной части соответствующей записи. Для каждой записи выделяется переменное количество пар (k, v). В текстовой записи каждого события присутствует набор обязательных пар (k, v).

3) Для каждого события в окне анализа с помощью предварительно построенного словаря K всех возможных вариантов постоянной части (за весь период наблюдений) выполняется кодирование k_i в виде бинарного вектора H.

4) С помощью предварительно построенной нейросетевой модели CBOW Word2Vec для постоянных частей записей событий в окне анализа выполняется построение векторов вложений E. Для каждого события строится собственный вектор вложений. Результирующий вектора для всех событий окна анализа находится как взвешенная сумма.

5) В текстовой записи каждого события присутствует набор обязательных пар (k, v), среди которых выделяются 10 ключевых пар, переменные части которых рассматриваются как вектор признаков V_c. Все признаки разделяются на три группы: количественные признаки (используется RobustScaler – RS, который при масштабировании вычитает из данных медиану и делит результат на интерквартильный размах), порядковые категориальные признаки (для кодирования используется OrdinarEncoder – OE), номинальные признаки (для кодирования данных без внутренней иерархии используется OneHotEncoder – OhE).

6) В результате формируются несколько векторов признаков: $\{F^E\}_{W1}$ – векторы вложений для каждого события в текущем окне анализа; $\{F^V\}_{W1}$ – вектор признаков из обязательной части текстовой записи события; $F^E{}_{W1}$ – взвешенная сумма векторов вложений отдельных событий.

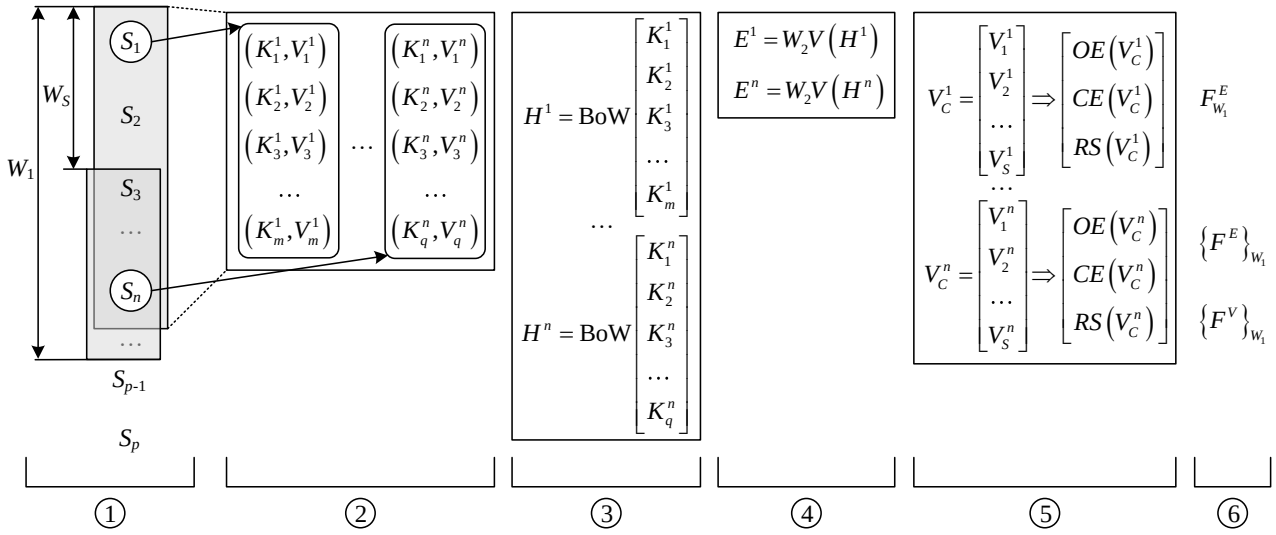


Рис. 1 Схема предлагаемого алгоритма обработки потока текстовых событий.

Предлагаемая структурная схема системы обнаружения аномалий на основе анализа ЖРС, основанная на композиции моделей машинного обучения и нейросетевых моделей извлечения и формализации признаков, приведена на рис. 2.

Подсистема I выполняет сбор слабоструктурированных данных с помощью типовых протоколов (например, протокола SYSLOG) с анализируемых средств защиты. Модуль *1* выполняет предварительную обработку текстовых строк журнала. Модуль *2* с помощью набора регулярных выражений осуществляет разбор слабоструктурированных текстовых записей ЖРС в структурированный формат «key-value» и обеспечивает эффективную схему хранения данных в колоночной БД как для исходного текстового представления, так и для словаря разбора. С помощью консоли *3* специалист *1* выполняет настройку и последующий контроль процесса сбора и хранения первичных данных ЖРС.

Подсистема II выполняет задачи извлечения признаков из структурированных записей ЖРС и построения моделей обнаружения аномалий для последовательности событий в окне анализа. Модуль *4* выполняет отбор признаков постоянной (key) и переменной (value) части, а также выделяет среди (*k, v*) пар группу обязательных (присутствующих в каждой текстовой записи ЖРС). Модуль *6* с помощью модели вложений на основе Word2Vec (модуль *7* – построение словаря для модели) выполняет преобразование закодированных представлений постоянной части каждого события в вектор вложений. Модуль *5* выполняет кодирование признаков для категориальных, временных и непрерывных признаков (value) обязательной группы пар (*k, v*). Модуль *9* объединяет результаты работы модулей *5* и *6* в кортеж множеств признаков, которые могут быть использованы различными моделями анализа.

Модуль *10* с помощью метода скользящего окна выделяет последовательность записей. Скользящее окно движется по потоку событий с заданным шагом. События окна рассматриваются либо независимо, либо для всего окна строится общий вектор признаков. Например, векторы вложений объединяются для формирования вектора признаков как одного события, так и всех событий в окне анализа.

Модуль *8* позволяет строить модели обнаружения аномалий (новизны) для отдельных событий. В качестве моделей могут быть использованы: Isolation Forest, KNN, LCP, AE и т. п.

В качестве признаков событий используются закодированные представления VC обязательной части либо векторы вложений E. Несколько моделей объединяются в комитет, принимающий решение о наличии аномалии путем голосования.

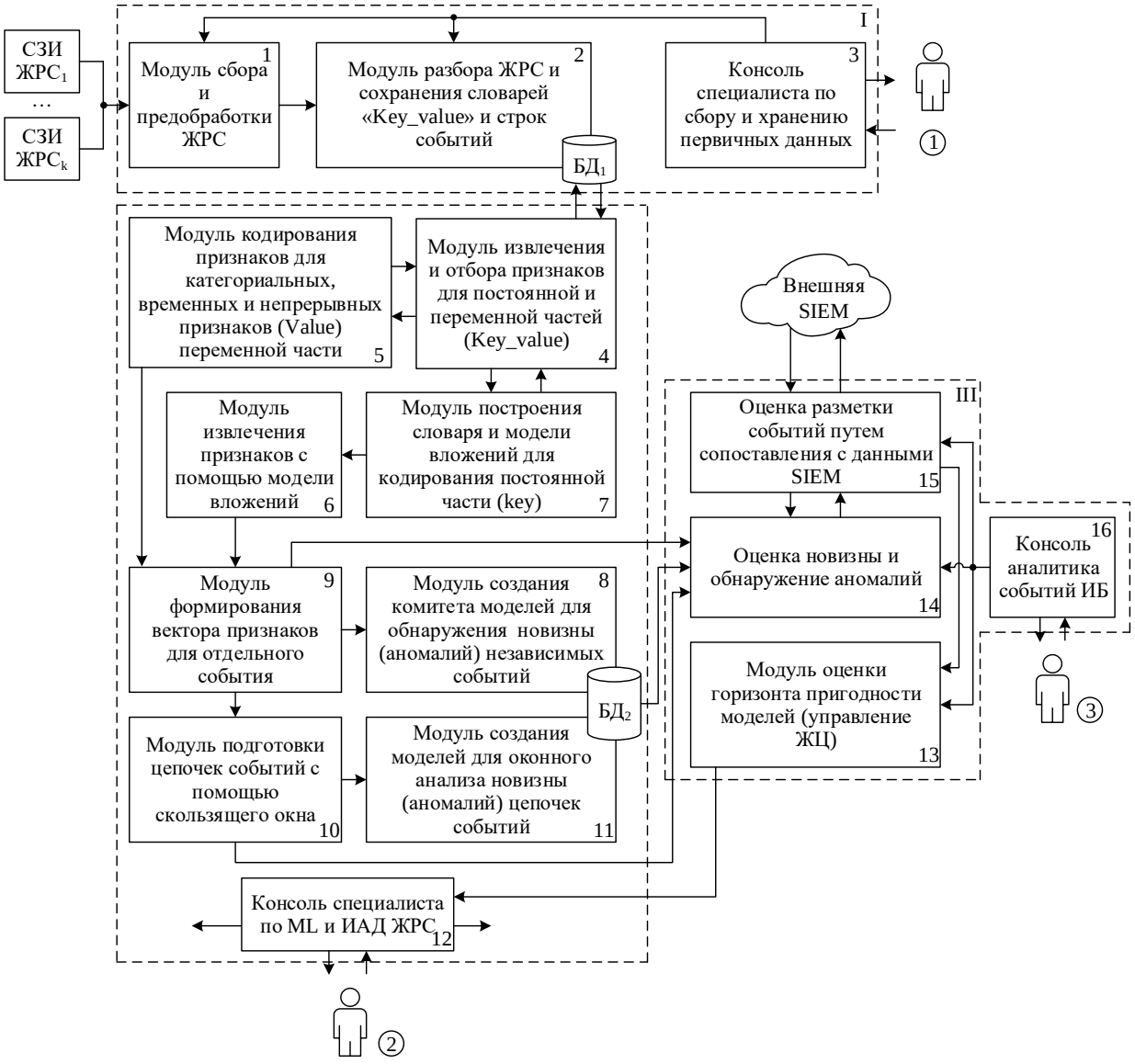


Рис. 2 Структурная схема системы обнаружения аномалий на основе анализа ЖРС.

Модуль *11* позволяет строить модели для оконного анализа, что позволяет уменьшить временной масштаб (агрегируя признаки событий в окне) и использовать более ресурсоемкие модели, не способные справиться с анализом отдельных событий в режиме реального времени.

Специалист (2) выполняет контроль и управление процессом конвейера анализа данных.

Подсистема III обеспечивает работу (модуль *14*) построенных моделей обнаружения аномалий (новизны) и взаимодействие с SIEM для передачи сведений об выявленных аномалиях, а также получения обратной связи о ложных срабатываниях (модуль *15*). Модуль *13* позволяет специалисту *3* управлять жизненным циклом моделей для контроля дрейфа данных.

3. Вычислительный эксперимент по анализу ЖРС Web Application Firewall

В качестве исходных данных были использованы журналы событий Advanced Web Application Firewall F5, развернутой в корпоративной сети с более чем 1000 устройств. Данное

средство защиты информации передает поток событий в SIEM по протоколу SYSLOG в формате Common Event Format (CEF). CEF использует формат «ключ–значение», включает заголовок и расширение, общие для всех категорий по конкретной категории событий [12].

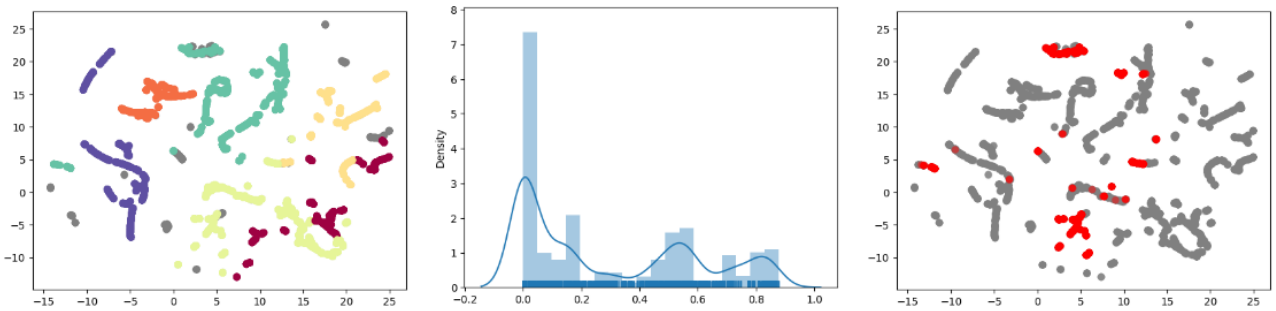
События в журнале содержат до 45 пар (*k*, *v*) признаков (основные и расширенные). Из набора данных были исключены признаки, содержащие нулевые значения или единственное уникальное. Оставшийся набор признаков включает номинальные (порт источника и назначений, сетевые адреса источника и назначения), категориальные (в том числе текстовые) признаки (символьное имя хоста, веб-приложение, метод запроса, тип действия) и временные (метка времени) (рис. 3). Общее количество доступных данных составило более 5 млрд событий.

Out[ ]:

	dvchost	dvc	act	src	spt	dst	dpt	requestMethod	app	rt
0	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	217.15.128.149	3177	10.159.253.98	443	GET	HTTPS	2023-03-03 11:45:03
1	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	87.118.253.190	47202	10.159.253.44	443	GET	HTTPS	2023-03-03 11:45:04
2	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	178.176.83.62	6854	10.159.253.98	443	GET	HTTPS	2023-03-03 11:45:04
3	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	178.214.247.34	14203	10.159.253.98	443	GET	HTTPS	2023-03-03 11:45:04
4	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	136.169.168.53	24276	10.159.253.98	443	GET	HTTPS	2023-03-03 11:45:04
...	...	...	...	...	...	...	...	...	...	...
59300	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	213.59.137.87	59955	10.159.253.44	443	POST	HTTPS	2023-03-03 11:48:00
59301	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	85.249.21.215	8027	10.159.253.44	443	GET	HTTPS	2023-03-03 11:48:00
59302	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	81.211.82.2	58899	10.159.253.44	443	GET	HTTPS	2023-03-03 11:48:00
59303	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	91.215.226.211	54500	10.159.253.44	443	GET	HTTPS	2023-03-03 11:48:00
59304	msk1-waf01.fc.uralsibbank.ru	10.159.249.148	passed	85.140.22.109	51404	10.159.253.98	443	GET	HTTPS	2023-03-03 11:48:00

Рис. 3 Пример структурирования записи ЖРС.

Подвыборка для разведочного анализа содержит 10000 строк по 10 признаков обязательной части текстовых записей. Для повышения производительности алгоритмов кластеризации использован UMAP как эффективный шаг предварительной обработки на основе плотности распределения с целью понижения размерности пространства признаков. Результаты кластеризации с помощью алгоритма HDBSCAN приведены на рис. 4. Распределение оценок внутрикластерных и межкластерных отношений позволяет установить порог определения выбросов (аномальности), равный 0.9 (рис. 4).



а *б* *в*

Рис. 4 Кластеризация набора данных с помощью HDBSCAN (*а*);
распределение расстояний (*б*); визуализация выбросов после установления порога (*в*).

Результаты разведочного анализа показали, что сформированное признаковое пространство позволяет оценить наличие структуры распределения событий, в том числе наличие выбросов.

Предварительно размеченные моделями выбросы были соотнесены с дополнительной информацией из SIEM и других источников мониторинга ИБ (контекстом), проведена ручная

разметка групп событий и выбросов на основе контекста (табл. 2). Результаты работы моделей приведены в табл. 3.

Таблица 2

Используемые модели обнаружения аномалий и их параметры

№	Модель	Независимые события или окно анализа	Тип признаков	Параметры модели	
1	Isolation Forest, IFO	события	V_C	Количество моделей	128
				contamination	auto
2	Local Outlier Factor, LOF	события	V_C	Количество соседей	16
3	Автоэнкодер $LSTM_1$	окно	V_C	Слой: Input; LSTM (L1); LSTM (L2); RepeatVector (L3); LSTM (L4); LSTM (L5); Output Функция активации: —; 'relu'; 'relu'; —; 'relu'; 'relu'; — Размеры слоя: X, n_features; W_S, 128; 64; W_S, 64; W_S, 64; W_S, 128; W_S, n_features	
4	Автоэнкодер $LSTM_1$	события	V_C		
5	Автоэнкодер $LSTM_1$	события	E		
6	Комитет IFO + $LSTM_2$ + $LSTM_3$ + $LSTM_1$	события + окно	$V_C + E$	—	—

Таблица 3

Результаты работы моделей обнаружения аномалий на размеченной выборке

№	Модель	True Positive	True Negative	False Negative	False Positive	Включены в мониторинг (из FP)	Новые события (из FP)
1	IFO	972323	589	1347	25741	14	1783
2	LOF	976595	707	1229	21469	4	1987
3	$LSTM_1$	997644	1147	789	420	175	227
4	$LSTM_2$	995491	953	983	2573	106	697
5	$LSTM_3$	997552	1278	658	512	181	204
6	Комитет	997775	1520	416	289	201	88

Согласованность экспертной разметки выявленных цепочек событий, отнесенных к выбросам, с контекстом процессов в информационной системе составила 78 % (1520 событий) (обнаруженные цепочки событий, попадающие во временное окно анализа, соответствуют событиям, выделенным SIEM) для комитета моделей. Для 21 % (416 событий) выявленных событий в SIEM отсутствовала информация из прочих источников. Причем для 201 события подтверждена необходимость их маркировки, как вредоносной. Оставшиеся 88 событий вызваны сменой типа квазистационарного состояния информационной системы и обусловлены их «новизной».

Использование моделей (1–2) для анализа отдельных событий приводит к существенному количеству ложноположительных срабатываний. Модели на основе автоэнкодеров показали лучший результат для признаков на основе вложений. Использование комитета моделей (взвешенное голосование по каждому событию в окне анализа и оценка всего окна для смешанного набора признаков) позволило существенно уменьшить количество ложноположительных

меток. Однако использование подобного комитета в текущей системе в режиме реального времени для анализа 15 тыс. событий в секунду вычислительно невозможно. Компромиссным решением стало применение одиночного автоэнкодера для оконного анализа.

4. Анализ результатов и оценка эффективности предложенного решения

Практическое применение подобной системы возможно в составе комплекса средств защиты корпоративной сети, выступающих в качестве источников событий безопасности для системы SIEM. Дальнейший анализ больших массивов гетерогенных данных о событиях безопасности и обнаружения инцидентов и угроз безопасности является основной, наиболее ресурсоемкой операцией, которая выявляет причинно-следственные связи между поступающими на обработку событиями. Операция корреляции позволяет выявлять вредоносную и аномальную активности, определять источник и цель атаки на основе анализа комплекса показателей, но не единичных инструментов и детекторов, и отсеивать значительное количество ложноположительных срабатываний за счет сопоставления признаков атаки из разных источников.

Таким образом, особенностями предложенного подхода к построению детекторов аномалий по сравнению с аналогичными исследованиями является использование моделей, не требующих разметки на классы нормального и аномального функционирования, что потенциально позволяет обнаруживать аномалии, вызванные новыми типами атак злоумышленника.

Перспективной является гетерогенная архитектура нейросетевого автоэнкодера, обрабатывающего как признаки на уровне моделей вложений, так и выделенные количественные и качественные признаки постоянной и переменной частей шаблона события в сочетании с моделями машинного обучения для обнаружения аномалий (LOF, IFO и т.п.), а также добавление еще одного выходного фильтра, позволяющего учитывать временные особенности реализации атак и длительность вызванных изменений в наблюдаемых параметрах.

Заключение

Для выявления сложных атак злоумышленников на информационные системы необходимо применение методов и инструментов расширенной аналитики данных, позволяющих выполнять оперативный анализ и выявление скрытых признаков злонамеренной активности на основе анализа журналов работы [13, 14].

Предложена структурная схема системы обнаружения аномалий, основанная на применении методов анализа собираемых событий и позволяющая выявить воздействия злоумышленника, получившего доступ в корпоративную сеть.

Разработан алгоритм обнаружения аномалий на основе анализа формализованных описаний цепочек событий и применения нейросетевых моделей автоэнкодеров с долгой краткосрочной памятью. Применение модели обнаружения аномалий позволяет с помощью адаптивной настройки порога фильтрации выделять одиночные события и цепочки событий, отличающиеся от текущего нормального состояния объектов в составе корпоративной информационной системы. Выделение аномалий варьируется в количественном диапазоне от 0.1 до 10 %, и является дополнительным источником данных для SIEM системы.

Особенностями предложенного подхода к построению детекторов аномалий по сравнению с аналогичными исследованиями является использование моделей, не требующих разметки на классы нормального и аномального функционирования, что потенциально позволяет обнаруживать аномалии, вызванные новыми типами атак злоумышленника.

Научная новизна предлагаемых алгоритма предобработки, модели анализа и алгоритма обнаружения аномалий основана на применении методов и моделей глубокого и машинного обучения, отличительной особенностью является совокупность этапов обработки и извлечения признаков, а также применение комитета нейросетевых моделей автоэнкодера с долгой

краткосрочной памятью в сочетании с моделями машинного обучения IFO и LOF для адаптивного выделения аномальных цепочек событий. Это позволяет повысить эффективность анализа текстовых ЖРС и, следовательно, повысить оперативность выявления аномальных состояний компонент информационной системы.

Следует также отметить работы по смежной тематике [15–17], оказавшие влияние на данное исследование.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

[1] Liu F. et al. Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise // 2019 ACM SIGSAC conference on computer and communications security. London, United Kingdom, 2019. Pp. 1777–1794. DOI: 10.1145/3319535.3363224.

[2] Cheng H. et al. Fine-grained Anomaly Detection in Sequential Data via Counterfactual Explanations // arXiv preprint arXiv: 2210.04145, 2w022. [Online]. Available: https://arxiv.org/abs/2210.04145. [Accessed July. 14, 2024].

[3] Zhou J. et al. DeepSyslog: Deep Anomaly Detection on Syslog Using Sentence Embedding and Metadata // IEEE Transactions on Information Forensics and Security. 2022. 17. Pp. 3051–3061. DOI: 10.1109/tifs.2022.3201379 EDN: VHBSTQ.

[4] Bogatinovski J. Leveraging Log Instructions in Log-based Anomaly Detection // 2022 IEEE International Conference on Services Computing (SCC). Barcelona, Spain, 2022. Pp. 321–326. DOI: 10.1109/SCC55611.2022.00053.

[5] Zhang K. A Log Anomaly Detection Method Based on Logical Reasoning // International Conference on Wireless Algorithms, Systems, and Applications. Dalian, China, 2022. Pp. 489–500. DOI: 10.1007/978-3-031-19214-2_41.

[6] Le V., Zhang H. Log-based anomaly detection without log parsing // 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE). Melbourne, Australia, 2021. Pp. 492–504.

[7] Meng W. et al. A semantic-aware representation framework for online log analysis // 2020 29th International Conference on Computer Communications and Networks (ICCCN). Honolulu, HI, USA, 2020. Pp. 1–7. DOI: 10.1109/ICCCN49398.2020.9209707

[8] Wang Z. et al. LightLog: A lightweight temporal convolutional network for log anomaly detection on the edge // Computer Networks. 2022. 203. Pp. 108616–108626. DOI: 10.1016/j.comnet.2021.108616. EDN: TYSFWZ.

[9] Catillo M. Pecchia A., Villano U. AutoLog: Anomaly detection by deep autoencoding of system logs // Expert Systems with Applications. 2022. 191. Pp. 116263–116273. DOI: 10.1016/j.eswa.2021.116263. EDN: JRVCVG.

[10] Guo H. et al. Logformer: A pre-train and tuning pipeline for log anomaly detection // AAAI Conference on Artificial Intelligence, Huawei. Beijing, China, 2024. Pp. 135–143. DOI: 10.1609/aaai.v38i1.27764. EDN: PECHSS.

[11] Zhang C. et al. LayerLog: Log sequence anomaly detection based on hierarchical semantics // Applied Soft Computing. 2023. 132. Pp. 109860. DOI: 10.1016/j.asoc.2022.109860. EDN: HWQPXW.

[12] Get CEF-formatted logs from your device or appliance into Microsoft Sentinel [Online]. Available: https://learn.microsoft.com/en-gb/azure/sentinel/ connect-common-event-format. [Accessed July. 14, 2024].

[13] Вульфин А. М. Модели и методы комплексной оценки рисков безопасности объектов критической информационной инфраструктуры на основе интеллектуального анализа данных // СИИТ. 2023. Т. 5. № 4(13). С. 50–76. EDN FJPFKC. [[Vulfin A. M. "Models and methods for comprehensive assessment of security risks of critical information infrastructure facilities based on intelligent data analysis" // SIIT. 2023. Vol. 5, No. 4(13), pp. 50–76. EDN FJPFKC. (In Russian).]]

[14] Васильев В. И., Картак В. М. Применение методов искусственного интеллекта в задачах защиты информации (по материалам научной школы УГАТУ) // СИИТ. 2020. Т. 2. № 2(4). С. 43–50. EDN ZTQFCW. [[Vasiliev V. I., Kartak V. M. "Application of artificial intelligence methods in information security problems (based on the materials of the scientific school of Ufa State Aviation Technical University)" // SIIT. 2020. Vol. 2, No. 2(4), pp. 43–50. EDN ZTQFCW. (In Russian).]]

[15] Аникин И. В. Методы и алгоритмы количественной оценки и управления рисками безопасности в корпоративных информационных сетях на основе нечеткой логики // СИИТ. 2023. Т. 5. № 3(12). С. 93–113. EDN KUKQGP. [[Anikin I. V. "Methods and algorithms for quantitative assessment and management of security risks in corporate information networks based on fuzzy logic" // SIIT. 2023. Vol. 5, No. 3(12), pp. 93–113. EDN KUKQGP. (In Russian).]]

[16] Кириллова А. Д. Оценка рисков информационной безопасности АСУ ТП промышленных объектов методами когнитивного моделирования // СИИТ. 2023. Т. 5. № 4(13). С. 77–93. EDN CUEUUP. [[Kirillova A. D. "Assessment of information security risks of industrial control systems using cognitive modeling methods" // SIIT. 2023. Vol. 5, No. 4(13), pp. 77-93. EDN CUEUUP. (In Russian).]]

[17] Шамсутдинов Р. Р., Васильев В. И., Вульфин А. М. Интеллектуальная система мониторинга информационной безопасности промышленного интернета вещей с использованием механизмов искусственных иммунных систем // СИИТ. 2024. Т. 6. № 4(19). С. 14–31. EDN LTXBSG. [[Shamsutdinov R. R., Vasiliev V. I., Vulfin A. M. "Intelligent system for monitoring information security of the industrial Internet of things using mechanisms of artificial immune systems" // SIIT. 2024. Vol. 6, No. 4(19), pp. 14–31. EDN LTXBSG. (In Russian).]]

Поступила в редакцию 28 ноября 2024 г.

МЕТАДАННЫЕ / METADATA

Title: System for detecting anomalies in the event logs of means of information security.

Abstract: The paper proposes models and algorithms for detecting anomalies in the functioning of information system based on analyzing the event logs of information protection means by using neural network models. The aim of the research is to improve the models and algorithms for the analysis of text logs to increase the efficiency of detecting abnormal states of the system. Scientific novelty of the proposed preprocessing algorithm, analysis model and anomaly detection algorithm is based on the application of deep machine learning methods and models. The distinctive feature is a combination of processing and feature extraction stages, as well as the application of a committee of neural network models - autoencoder with long short-term memory combined with IFO and LOF for adaptive selection of anomalous chains of events. This allows to increase the efficiency of analysis of text logs of the system operation and increase the efficiency of detection of abnormal states of the information system components.

Key words: event log analysis; deep learning; neural network autoencoder; information protection; anomaly and outlier detection

Язык статьи / Language: Русский / Russian.

Об авторах / About the authors:

АТАРСКАЯ Елена Андреевна
Уфимский университет науки и технологий, Россия.
Асп. каф. вычислительной техники и защиты информации. Дипл. магистр по направлению «информатика и вычислительная техника» (Уфимск. ун-т науки и технологии, 2023). Иссл. в обл. интелл. систем обнаружения аномалий.
E-mail: atarskaya.ea@ugatu.su
ORCID: https://orcid.org/0009-0005-2455-4387

ВУЛЬФИН Алексей Михайлович
Уфимский университет науки и технологий, Россия.
Профессор каф. вычислительной техники и защиты информации. Дипл. инж.-программист (Уфимск. гос. нефтян. техн. ун-т, 2008). Д-р техн. наук по методам и системам защиты информации, инф. безопасности (Уфимск. гос. авиац. техн. ун-т, 2022). Иссл. в обл. информ. безопасности, интеллектуальных систем, нечеткого и нейросетевого моделирования.
E-mail: vulfin.alexey@gmail.com
ORCID: https://orcid.org/0000-0001-5857-2413

КИРИЛЛОВА Анастасия Дмитриевна
Уфимский университет науки и технологий, Россия.
Доцент каф. вычислительной техники и защиты информации. Дипл. преп., преп.-исследователь (Уфимск. гос. авиац. техн. ун-т, 2022). Канд. техн. наук по методам и системам защиты информации, инф. безопасности (Уфимск. ун-т науки и технологии, 2023).
E-mail: kirillova.ad@ugatu.su
ORCID: https://orcid.org/0009-0000-4164-2526

ATARSKAYA Elena Andreevna
Ufa University of Science and Technology, Russia.
Postgrad. Student, Dept. of Computer Engineering and Information Security. Master's Degree of Informatics and Computing (Ufa Univ. of Science and Technology, 2023). Research in the field of intelligent systems for detecting anomalies.
E-mail: atarskaya.ea@ugatu.su
ORCID: https://orcid.org/0009-0005-2455-4387

VULFIN Aleksey Mikhaylovich
Ufa University of Science and Technology, Russia.
Professor of the Department Computer Technology and Information Security. Dipl. software engineer (Ufa State Oil Techn. Univ., 2008). Dr. Tech. Sciences on Methods and Systems of Information Security (Ufa State Aviat. Techn. Univ., 2022). Research in the field of information security, intelligent systems, fuzzy and neural network modeling.
E-mail: vulfin.alexey@gmail.com
ORCID: https://orcid.org/0000-0001-5857-2413

KIRILLOVA Anastasiya Dmitrievna
Ufa University of Science and Technology, Russia.
Senior Lecturer of the department computer technology and information security. Dipl. teacher, teacher-researcher (Ufa State Aviat. Techn. Univ., 2022). Candidate of Technical Sciences in methods and systems of information protection, inf. security (Ufa Univ. of Science and Technology, 2023).
E-mail: kirillova.ad@ugatu.su
ORCID: https://orcid.org/0009-0000-4164-2526