УДК 004.056
DOI: 10.17586/0021-3454-2022-65-11-833-841

МОДЕЛЬ КОМБИНИРОВАННОГО ПРИМЕНЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ МЕТОДОВ КОРРЕЛЯЦИИ СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Д. А. Левшун

Санкт-Петербургский федеральный исследовательский центр Российской академии наук,
Санкт-Петербург, Россия
gaifulina@comsec.spb.ru

Аннотация. Для решения задачи корреляции событий информационной безопасности предложена модель комбинированного применения интеллектуальных методов корреляции. Интеллектуальные методы корреляции событий безопасности способны анализировать как объемные исторические данные, так и события в реальном времени, а также автоматически обнаруживать изменяющиеся пороговые значения. Предложенная модель содержит два уровня обработки данных: уровень представления знаний и уровень корреляции событий безопасности. На уровне представления знаний осуществляется структурный и семантический анализ событий. На уровне корреляции для обработки событий применяются оценка подобия элементов векторов событий безопасности, нейросетевой графо-ориентированный метод и анализ данных при помощи рекуррентных нейронных сетей. Результатами работы модели являются последовательность взаимосвязанных событий безопасности, тип текущего состояния безопасности системы и прогнозируемые состояния. Эффективность подхода на основе предложенной модели проиллюстрирована результатами эксперимента по прогнозированию событий безопасности системы, показывающими низкие значения показателя ошибок.

Ключевые слова: *корреляция событий, информационная безопасность, мониторинг состояния безопасности, интеллектуальный анализ данных*

Благодарности: работа выполнена при финансовой поддержке бюджетной темы FFZF-2022-0007.

Ссылка для цитирования: *Левшун Д. А.* Модель комбинированного применения интеллектуальных методов корреляции событий информационной безопасности // Изв. вузов. Приборостроение. 2022. Т. 65, № 11. С. 833—841. DOI: 10.17586/0021-3454-2022-65-11-833-841.

MODEL OF COMBINED APPLICATION OF INTELLIGENT METHODS OF INFORMATION SECURITY EVENTS CORRELATION

D. A. Levshun

St. Petersburg Federal Research Center of the RAS,
St. Petersburg, Russia
gaifulina@comsec.spb.ru

Abstract. To solve the problem of information security event correlation, a model for the combined use of intelligent correlation methods is proposed. Intelligent security event correlation methods are able to analyze both historical data and real-time events and automatically detect changing thresholds. The proposed model contains two levels of data processing: the level of knowledge representation and the level of security event correlation. At the level of knowledge representation, structural and semantic analysis of events is carried out. At the correlation level, the similarity assessment of elements of security event vectors, a graph-oriented neural network method and data analysis using recurrent neural networks are used for event processing. The results of the model are the sequence of interrelated security events, the type of the current security state of the system and the predicted states. The performance of the approach based on the proposed model is illustrated by results of an experiment on predicting system security events, showing low values of the error indicator.

Keywords: event correlation, information security, security monitoring, data mining

Acknowledgments: the work was carried out with the financial support of the budget theme FFZF-2022-0007.

For citation: Levshun D. A. Model of combined application of intelligent methods of information security events correlation. *Journal of Instrument Engineering.* 2022. Vol. 65, N 11. P. 833—841 (in Russian). DOI: 10.17586/0021-3454-2022-65-11-833-841.

Введение. Современные вычислительные системы и сети в больших объемах генерируют данные для аналитики информационной безопасности (ИБ). В свою очередь, инструменты безопасности должны оперативно обрабатывать входящий поток информации, отслеживать состояние безопасности целевой системы или сети и предупреждать о наличии потенциальных угроз. Как правило, к таким инструментам относят системы обнаружения и предупреждения вторжений, системы контроля доступа, сканеры уязвимостей, антивирусы и другие средства. Данные о состоянии безопасности могут быть собраны из разнородных источников и обработаны в едином интерфейсе SIEM-системы (Security Information and Event Management), что облегчает изучение характерных особенностей событий безопасности.

Под событием информационной безопасности понимается выявленное состояние системы или сети, указывающее на возможное нарушение политики или мер обеспечения ИБ, либо ранее неизвестная ситуация, которая может иметь отношение к вопросам безопасности. Событие, которое с высокой степенью вероятности может создать угрозы, называется инцидентом ИБ (ГОСТ Р ИСО/МЭК 27000-2021). К подобного рода инцидентам можно отнести атаки и аномалии, характеризуемые разной продолжительностью и определенными сценариями поведения [1]. Так, сценарий атаки может быть описан с помощью определения цели и источника атаки, а также промежуточных шагов атакующего, связанных между собой некоторыми логическими отношениями. С учетом данных факторов перспективным направлением обработки данных о состоянии безопасности является применение интеллектуальных методов корреляции событий ИБ.

В широком смысле под корреляцией понимается определение причинно-следственных взаимосвязей событий безопасности, что позволяет как идентифицировать текущее состояние безопасности, так и прогнозировать вероятностные состояния [2, 3]. Корреляция событий ИБ способствует лучшему пониманию развития атаки, определению источника и цели атаки, выявлению наиболее значимых событий. В задачах прогнозирования интеллектуальные методы корреляции событий безопасности способны анализировать как объемные исторические данные, так и события в реальном времени, а также автоматически обнаруживать изменяющиеся пороговые значения характеристик событий [4]. Это позволяет выявлять аномальные события и предотвращать кибератаки на ранних стадиях.

Постановка задачи. Цель настоящего исследования — повышение защищенности целевых систем и сетей за счет повышения уровня ситуационной осведомленности об ИБ путем разработки усовершенствованной модели корреляции событий безопасности. Входными данными модели корреляции событий безопасности являются данные о событиях безопасности в процессе функционирования системы, а также данные о возможных состояниях безопасности; выходные данные модели — 1) последовательность взаимосвязанных событий безопасности; 2) тип текущего состояния безопасности системы, 3) прогнозируемое состояние безопасности системы.

Поток регистрируемых событий безопасности обозначим как множество $E = \{e_n\}$, $n = 1,\ldots,N$, где N — размер множества, а возможные классы (типы) состояний безопасности системы — как $Y = \{y_к\}$, к= 0,…, K, где K — количество классов. В качестве классов состояний безопасности системы могут применяться как бинарное множество Y = {normal, anomaly}, где normal — штатное состояние системы, anomaly — состояние системы, отличное от штатного (аномалия), так и множество $Y = \{ns_1,\ldots,ns_d, as_1,\ldots,as_k\}$, где $ns_1,\ldots,ns_d$ — подмножество штатных состояний системы, $as_1,\ldots,as_k$ — подмножество, состоящее из атак [5]. Таким образом, входные данные модели являются парой (E, Y).

Последовательность взаимосвязанных событий безопасности eS = (**E**, R), где **E** — вектор событий безопасности, а $R = \{R_h\}$, $h = 1,\ldots,H$, — множество отношений между событиями безопасности, H — размер множества R. Каждый элемент множества R описывает некоторые условия отношения пары событий $e_i \in \mathbf{E}$ и $e_j \in \mathbf{E}$ как отображение r: $e_i \rightarrow e_j$, $r \in R$, где символ

„→“ обозначает функциональную зависимость между событиями. Последовательность событий безопасности можно охарактеризовать классом текущего состояния системы y^E, где $y^E \in Y$, а также классом прогнозируемого состояния безопасности $\hat{y}^E \in Y$. Таким образом, получение выходных данных модели корреляции событий безопасности можно представить как отображение M_γ: $(E, Y) \rightarrow (eS, y^E, \hat{y}^E)$.

Комбинированная модель корреляции. В научной литературе представлено большое количество моделей корреляции событий безопасности. На основе проведенного аналитического обзора [6] можно выделить три основные категории методов корреляции событий безопасности:

— корреляция на основе сходства — определение взаимосвязей на основе подобия атрибутов или векторов событий;

— причинно-следственная корреляция — определение взаимосвязей пошаговым образом на основе предпосылок и последствий;

— корреляция на основе интеллектуального анализа данных — обнаружение существенных закономерностей и шаблонов в потоках событий на основе методов вычислительного интеллекта.

Также предложена классификация моделей представления знаний о событиях на основе классических методов: правила, логическое представление, семантические сети и фреймы [7, 8]. В качестве логического представления знаний о событиях безопасности определены семантические модели [9], а в качестве семантических сетей — графовые [10, 11]. Фреймы рассмотрены как набор функций и меток, используемых при машинном обучении. Таким образом, выделены следующие модели представления знаний о событиях: модели на основе правил, семантические модели, графовые модели и модели машинного обучения.

В данном исследовании представлена разработанная комбинированная модель корреляции событий безопасности:

$$\mathrm{M}_\gamma = < (\mathrm{MR}_\gamma, \mathrm{MX}_\gamma), (\mathrm{MSB}_\gamma, \mathrm{MGB}_\gamma, \mathrm{MDM}_\gamma) >,$$

где $(\mathrm{MR}_\gamma, \mathrm{MX}_\gamma)$ — уровень представления событий безопасности: MR_γ — модель представления событий безопасности на основе структурного анализа, MX_γ — модель представления событий безопасности на основе семантического анализа; $(\mathrm{MSB}_\gamma, \mathrm{MGB}_\gamma, \mathrm{MDM}_\gamma)$ — уровень корреляции событий безопасности: MSB_γ — модель корреляции событий безопасности на основе сходства, MGB_γ — модель причинно-следственной корреляции событий безопасности на основе графо-ориентированного подхода, MDM_γ — модель корреляции на основе интеллектуального анализа данных.

Уровень представления событий безопасности. Под структурным анализом понимается определение вектора события безопасности в виде ряда признаков, характеризующих событие. Такой подход, как правило, применяется для построения моделей корреляции на основе правил, графов или моделей машинного обучения [12]. Под семантическим анализом понимается определение вектора события с использованием методов контекстно-ориентированного анализа естественного языка для построения семантических моделей корреляции событий [13].

Извлечение признаков событий безопасности путем структурного анализа предполагает выделение ряда значимых характеристик. Модель представления событий безопасности на основе структурного анализа можно описать как

$$\mathrm{MR}_\gamma = (E, F^E, D, \varphi^d),$$

где $F^E = \{f_k\}$ — множество признаков событий безопасности, $k = 1,\ldots,K$, K — размер множества; $D = \{D_k\}$— область допустимых значений признака, так что $D_k = \{d_{kl}\}$— множество

допустимых значений признака f_k, $l = 1,\ldots,L$, L — количество значений; φ^d: $E \to D_k$ — функция извлечения признаков событий безопасности.

Каждое событие безопасности $e \in E$ можно представить в виде вектора признаков $\boldsymbol{\alpha}^e = \{(f_1, d_1), \ldots, (f_k, d_k)\}$, где пара (f_i, d_i) соответствует i-му признаку события ($f_i \in F^E$) со значением $d_i \in D_k$. Результатом структурного анализа множества E является матрица векторов событий безопасности $\boldsymbol{\Phi}$ размером $N \times K$ (N — количество событий в наборе) для описания состояния безопасности системы: $\boldsymbol{\Phi} = \left(\boldsymbol{\alpha}^{e_1}, \ldots, \boldsymbol{\alpha}^{e_n}\right)^T = \left\|(f_{ij}, d_{ij})\right\|_{k \times n}$.

Для обработки семантического контекста событий безопасности, которые содержат данные в текстовом формате (лог-линии, полезная нагрузка и т.д.), предлагается использовать методы обработки естественного языка, такие как встраивания слов. Модель представления событий безопасности на основе семантического анализа можно описать как

$$\mathrm{MX}_\gamma = (E, V^E, \varphi^v),$$

где V^E — векторное пространство событий; φ^v: $E \to V^E$ — функция преобразования события в вектор действительных чисел (эмбеддинг), который определяет координаты события в V^E.

Соответственно событие безопасности $e \in E$ можно представить в виде вектора $\boldsymbol{v}^e = \{v_q\}$, $v_q \in \mathbb{R}$, $q=1,\ldots,Q$, где Q — длина вектора. Аналогично матрица $\boldsymbol{\Phi}$ состоит из эмбеддингов событий и имеет размер $N \times Q$: $\boldsymbol{\Phi} = \left(\boldsymbol{v}^{e_1}, \ldots, \boldsymbol{v}^{e_n}\right)^T = \left\|v_{ij}\right\|_{q \times n}$.

Так, уровень представления событий безопасности позволяет получить матрицу векторов событий, которая является, в свою очередь, входными данными для уровня корреляции. Далее данная матрица обозначается как $\boldsymbol{\Phi} = \left(\mathbf{e}_1, \ldots, \mathbf{e}_n\right)^T$, где $\mathbf{e}_i = (x_{i1}, \ldots, x_{iz})$ — вектор события безопасности $e_i \in E$ длиной z, x_{ij} — элемент i-го вектора. Матрица, в которой события безопасности упорядочены по времени появления, определяется как $\boldsymbol{\Phi}(t) = \left((\mathbf{e}_1, t_1), \ldots, (\mathbf{e}_n, t_h)\right)^T$, где t_i — временная метка события. Для обучающей матрицы событий $\boldsymbol{\Phi}^y$ содержащиеся в ней события относятся некоторому классу $y \in Y$: $\boldsymbol{\Phi}^y = \left((\mathbf{e}_1, y_1), \ldots, (\mathbf{e}_n, y_n)\right)^T$.

Уровень корреляции событий безопасности. Предлагаемая модель корреляции позволяет как идентифицировать схожие события, так и обнаруживать между ними причинно-следственные связи. В основе корреляции событий безопасности на основе сходства лежит идея, что схожие состояния безопасности могут иметь одинаковый класс. Модель корреляции событий безопасности на основе сходства описывается как

$$\mathrm{MSB}_\gamma = (\boldsymbol{\Phi}, Y, c^f, s^e, \omega^k),$$

где c^f: $\boldsymbol{\Phi} \to \mathbf{C}$ — функция корреляции элементов векторов событий, $\mathbf{C}$ — матрица корреляции элементов векторов событий безопасности; s^e: $(\boldsymbol{\Phi}, \mathbf{C}) \to \mathbf{S}$ — функция, определяющая подобие векторов событий, $\mathbf{S}$ — матрица сходства векторов событий безопасности; ω^k — алгоритм определения наиболее схожих событий.

Матрица корреляции $\mathbf{C}$ определяется как $\mathbf{C} = \left\|c^f(\mathbf{x}_i, \mathbf{x}_j)\right\|_{z \times z} = \left\|c_{ij}\right\|_{z \times z}$, где c_{ij} — коэффициент корреляции для событий x_i и x_j. Подобие векторов событий определяется путем вычисления „мягкой“ косинусной меры сходства s^e, которая характеризует угол между векторами, а также учитывает корреляцию между их элементами. В результате преобразования формируется матрица сходства векторов событий безопасности:

$$\mathbf{S} = s^e(\boldsymbol{\Phi}, \mathbf{C}) = \left\|s^e(\mathbf{e}_i, \mathbf{e}_j)\right\|_{n \times n},$$

$$s^e(\mathbf{e}_i, \mathbf{e}_j) = \frac{\sum\limits_{q=1,p=1}^{z} c_{qp} x_{iq} x_{jp}}{\sqrt{\sum\limits_{q=1,p=1}^{z} c_{qp} x_{iq} x_{ip} \cdot \sum\limits_{q=1,p=1}^{z} c_{qp} x_{jq} x_{jp}}}.$$

При заданном пороге сходства $s^{\min}$ схожие между собой события объединяются в кластеры $H^E = \{\mathbf{H}_i\}$, где $\mathbf{H}_i = \{\mathbf{e}_j \mid s^e(\mathbf{e}_j, \mathbf{e}_l) \geq s^{\min} \ \forall \ \mathbf{e}_l \in \mathbf{H}_i\}$. Для любой пары событий в матрице кластера их сходство выше указанного порогового значения. Для любого регистрируемого вектора события $\mathbf{e}_r$ полученные значения располагаются в порядке возрастания: $s^e(\mathbf{e}_{(1:r)}, \mathbf{e}_r) \leq \leq s^e(\mathbf{e}_{(2:r)}, \mathbf{e}_r) \leq \ldots \leq s^e(\mathbf{e}_{(n:r)}, \mathbf{e}_r)$, где через $\mathbf{e}_{(i:r)}$ обозначается событие, являющееся i-м соседним с $\mathbf{e}_r$. Алгоритм ω^k относит $\mathbf{e}_r$ к тому классу, представителей которого окажется больше всего среди k наиболее схожих с $\mathbf{e}_r$ событий:

$$\omega^k(\mathbf{\Phi}^y, \mathbf{e}_r) = \operatorname*{argmax}_{y \in Y} \sum_{i=1}^{k} \left[y(\mathbf{e}_{i:r}) = y \right] = y^{e_r}.$$

Корреляция событий на основе сходства не учитывает последовательность событий во времени в отличие от причинно-следственной корреляции. Модель причинно-следственной корреляции событий безопасности на основе графо-ориентированного подхода представляется в следующем виде:

$$\mathrm{MGB}_\gamma = (\mathbf{\Phi}(t),\ Y,\ g,\ \mathrm{GNN}(H^U,\ W)),$$

где $\mathbf{\Phi}(t)$ — матрица векторов событий безопасности с временными метками; g: $\mathbf{\Phi}(t) \rightarrow \mathbf{G}$ — функция построения графа событий безопасности, $\mathbf{G}$ — граф событий безопасности; GNN — графовая нейронная сеть, заданная множеством слоев H^U, U — количество слоев сети, $W = \{\mathbf{W}_u\}$ — множество матриц весов.

Граф события безопасности определяется как

$$\mathbf{G} = (\mathbf{\Phi}(t),\ \mathbf{R}^G,\ w_E,\ w_R),$$

$$\mathbf{R}^G = \left\| r(e_i, e_j) \right\|_{n \times n} = \left\| R_{ij} \right\|_{n \times n},$$

где $\mathbf{R}^G$ — множество переходов между событиями безопасности, отраженными в вершины графа (ребра графа), R_{ij} — переход между $\mathbf{e}_i$ и $\mathbf{e}_{ij}$; w_E: $\mathbf{\Phi}(t) \rightarrow \mathbb{R}$ — функция, отображающая вершины в их весовые коэффициенты; w_R: $\mathbf{R}^G \rightarrow \mathbb{R}$ — функция, отображающая ребра в их весовые коэффициенты.

Применение функции w_R позволяет получить матрицу смежности графа $\mathbf{A}^G$:

$$\mathbf{A}^G = w_R(\mathbf{R}^G) = \left\| a_{ij} \right\|_{n \times n},$$

где a_{ij} — вес ребра, соединяющего вершины $\mathbf{e}_i$ и $\mathbf{e}_j$.

Для анализа графа событий безопасности предлагается использовать графовую нейронную сеть GNN, каждый слой которой может быть представлен следующей нелинейной функцией:

$$H^{(u+1)} = \sigma(\mathbf{A}^G \times H^{(u)} \times \mathbf{W}^{(u)}),$$

где $\mathbf{W}^{(u)}$ — матрица весов для u-го слоя сети, σ — функция активации; на входном слое $H^{(0)} = \mathbf{G}$, а на выходном слое $H^{(U)} = y^G$, где y^G — класс графа событий.

Модель корреляции на основе интеллектуального анализа данных может быть представлена в следующем виде:

$$\mathrm{MDM}_\gamma = <\mathbf{\Phi}(t),\ Y,\ \mu^e,\ \mathrm{BiLSTM}(\mathbf{x}_t,\ \mathbf{h}_t,\ \mathbf{y}_t,\ \mathbf{c}_t,\ \Omega,\ W,\ B,\ \Sigma)>,$$

где μ^e: $\mathbf{\Phi}(t) \rightarrow \boldsymbol{x}_t$ — функция преобразования матрицы векторов событий во входной вектор сети; BiLSTM — двунаправленная нейронная сеть LSTM, имеющая в качестве параметров входной вектор $\mathbf{x}_t$, скрытый вектор $\mathbf{h}_t$, выходной вектор $\mathbf{y}_t$, вектор состояний $\mathbf{c}_t$,, векторы

вентилей $\Omega = (\mathbf{f}_t, \mathbf{i}_t, \mathbf{o}_t)$, множество матриц весов $W = \{\mathbf{W}_{ij}\}$, множество векторов смещения $B = \{\mathbf{b}_j\}$ и множество функций активации $\Sigma = \{\sigma_j\}$.

Долгая краткосрочная память (Long Short-Term Memory — LSTM) является разновидностью рекуррентных нейронных сетей и позволяет эффективно анализировать данные в случае, когда важные события разделены временными лагами неопределенной продолжительности. LSTM-блоки содержат вентили $\mathbf{f}_t$, $\mathbf{i}_t$, $\mathbf{o}_t$, реализованные в виде логистических функций, для контроля потоков информации на входах и выходах данных блоков, а также вектор состояний $\mathbf{c}_t$ в качестве внутренней памяти, которая обновляется с использованием текущего и предыдущего состояний:

$$\mathrm{LSTM} = \begin{cases} \mathbf{f}_t = \sigma_g(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{W}_{cf}\mathbf{c}_{t-1} + \mathbf{b}_f); \\ \mathbf{i}_t = \sigma_g(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{W}_{ci}\mathbf{c}_{t-1} + \mathbf{b}_i); \\ \mathbf{o}_t = \sigma_g(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{W}_{co}\mathbf{c}_{t-1} + \mathbf{b}_o); \\ \mathbf{c}_t = \mathbf{f}_t \circ \mathbf{c}_{t-1} + \mathbf{i}_t \circ \sigma_c(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c); \\ \mathbf{h}_t = \mathbf{o}_t \circ \sigma_c(\mathbf{c}_t); \\ \mathbf{y}_t = \mathbf{h}_t, \end{cases}$$

где $\mathbf{f}_t$ — вектор вентиля забывания (вес запоминания старой информации); $\mathbf{i}_t$ — вектор входного вентиля (вес получения новой информации); $\mathbf{o}_t$ — вектор выходного вентиля (кандидат на выход); $\mathbf{W}_{jk}$ — матрица весов для преобразования вектора j в компоненту вектора k; $\mathbf{b}_j$ — вектор смещения вектора j; σ_g — функция активации на основе сигмоиды; σ_c — функция активации на основе гиперболического тангенса; знак „$\circ$“ — произведение Адамара (покомпонентное произведение двух матриц).

Для анализа причин и последствий событий безопасности целесообразно анализировать и последующую информацию. В этом случае предлагается использовать двунаправленную модель долгой краткосрочной памяти (bidirectional LSTM, BiLSTM), позволяющую определять последовательность событий безопасности в двух направлениях:

$$\mathrm{BiLSTM} = \begin{cases} \mathbf{h}_t^{\rightarrow} = \mathrm{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}^{\rightarrow}); \\ \mathbf{h}_t^{\leftarrow} = \mathrm{LSTM}(\mathbf{x}_t, \mathbf{h}_{t+1}^{\leftarrow}), \end{cases}$$

где $\mathbf{h}_t$, $\mathbf{h}_{t-1}$ и $\mathbf{h}_{t+1}$ — векторы текущего, предыдущего и последующего скрытых состояний соответственно; индекс „$\rightarrow$“ обозначает прямое распространение информации, индекс „$\leftarrow$“ — обратное.

Взаимосвязь входных и выходных данных в комбинированной модели корреляции представлена в таблице.

Входные данные	Модель	Выходные данные
Уровень представления событий безопасности		
E — множество событий безопасности, Y — множество классов состояний безопасности системы	MR_γ	**Φ** — матрица векторов событий безопасности; **Φ**(t) — матрица векторов событий безопасности с временными метками,
	MX_γ	$\mathbf{\Phi}^y$ — обучающая матрица векторов событий безопасности с метками классов состояний
Уровень корреляции событий безопасности		
Φ, $\mathbf{\Phi}^y$	MSB_γ	H^E — множество кластеров схожих событий безопасности, y^e — класс события безопасности
Φ(t), Y	MGB_γ	$eS = \mathbf{G}$ — граф событий безопасности, y^G — класс графа событий безопасности
Φ(t), Y	MDM_γ	eS = BiLSTM — модель нейронной сети событий безопасности; $\mathbf{y}_t$ — текущий вектор классов состояний безопасности, $\mathbf{y}_{t+1}$ — прогнозируемый вектор событий безопасности

Пример практической реализации. На основе предложенной комбинированной модели был определен подход к прогнозированию событий безопасности и реализован программный прототип на языке Python. Входными данными являются параметры состояний безопасности, собранные за некоторое время, предшествующее прогнозированию, выходными данными — прогнозируемые параметры состояний. Для предсказания параметров событий в следующий момент времени необходимо проанализировать предшествующие события во временном окне размерностью N. Построение обучающей и тестовой выборок осуществляется таким образом, что обучающая выборка содержит временной ряд данных, предшествующий временному ряду в тестовой выборке. Метриками качества прогнозирования являются среднеквадратическая ошибка (MSE) и средняя абсолютная ошибка (MAE):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \; ; \quad \mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|,$$

где Y и $\hat{Y}$ — ряд наблюдаемых и ряд прогнозируемых значений переменной соответственно, n — длина ряда.

В качестве экспериментального набора данных использованы журналы событий, регистрирующие параметры процесса вождения „умного крана“ для 8 ездовых циклов [14]. Каждый ездовой цикл состоит из повторения процессов подъема груза, движения из точки A в точку B по маршруту, опускания груза, подъема груза, возвращения в точку A и опускания груза. Признаковое пространство содержит 12 характеристик ездового процесса, а состояние безопасности обозначается бинарной меткой (норма/аномалия).

Ошибки прогнозирования рассчитываются между реальными и прогнозируемыми параметрами как для характеристик каждого процесса по отдельности, так и в целом для состояния безопасности системы. Результаты по среднеквадратической ошибке прогнозирования каждой характеристики представлены на рис. 1. Для MSE по индивидуальным признакам установлено пороговое значение в 0,05, а для MAE — в 0,22 ($\approx\sqrt{0,05}$). Можно отметить, что ошибки в прогнозировании 8 признаков ниже пороговых значений. Наименьшие ошибки прогнозирования достигаются для признака 4 — MSE=0,017 и MAE=0,094, признака 6 — MSE=0,013 и MAE=0,055 и признака 12 — MSE=0,014 и MAE=0,081.

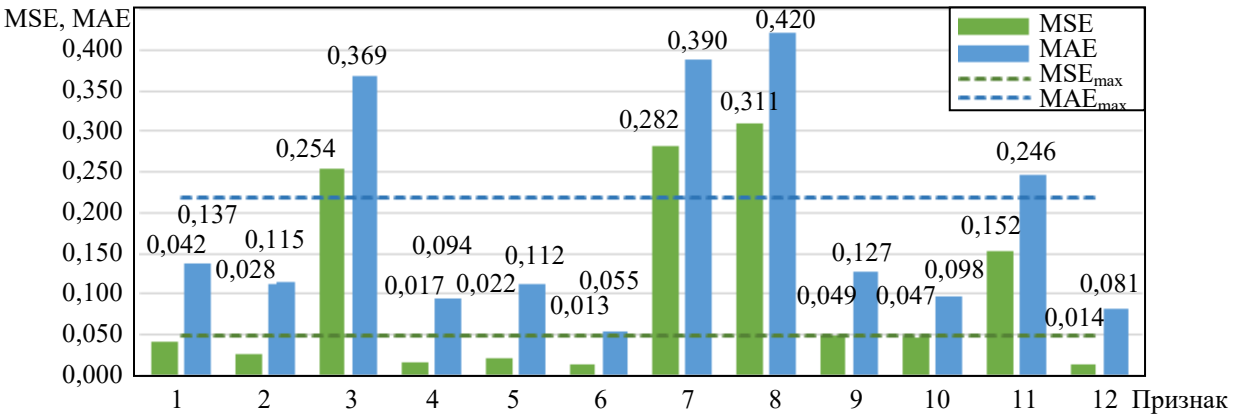

Рис. 1

На рис. 2 приведены ошибки прогнозирования состояния безопасности для каждого из 8 ездовых циклов (N — номер цикла). Для MSE установлен порог в 0,075 (0,05·12/8), а для MAE — в 0,27 ($\approx\sqrt{0,075}$). Для четырех циклов значения MSE ниже пороговых, а для семи циклов значения MAE ниже пороговых. Таким образом, прогнозирование состояний безопасности достигается при низком уровне ошибок.

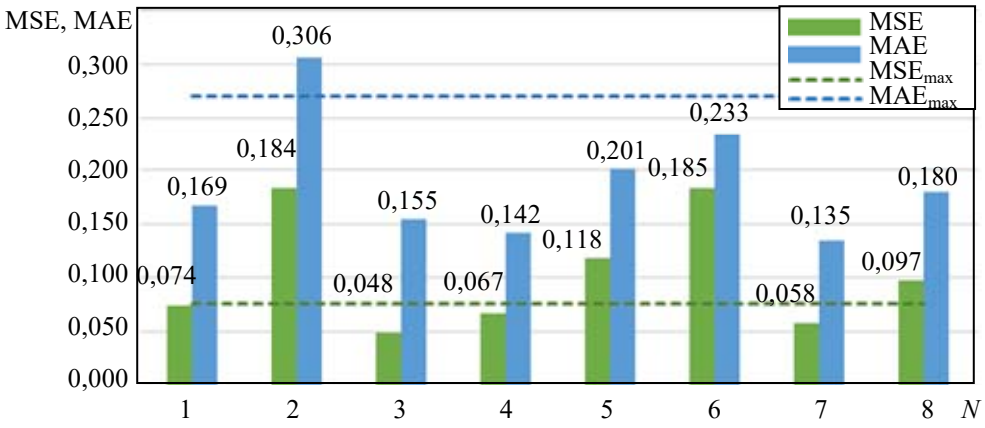

Рис. 2

Заключение. Предложено решение задачи корреляции событий безопасности с применением комбинированной модели корреляции, включающей в себя два уровня моделей с позиций а) представления знаний о событиях безопасности, б) методов корреляции событий безопасности. Представление знаний о событиях достигается путем объединения их структурного и семантического анализа. Уровень корреляции событий безопасности основан на применении методов корреляции, базирующихся на сходстве элементов векторов событий, нейросетевом графо-ориентированном подходе и анализе данных с помощью рекуррентных нейронных сетей. Эффективность предложенной комбинированной модели проиллюстрирована результатами эксперимента по прогнозированию событий безопасности системы на основе двунаправленной BiLSTM, входящей в состав модели корреляции.

СПИСОК ЛИТЕРАТУРЫ

1. *Котенко И. В., Саенко И. Б., Коцыняк М. А., Лаута О. С.* Оценка киберустойчивости компьютерных сетей на основе моделирования кибератак методом преобразования стохастических сетей // Информатика и автоматизация. 2017. Т. 6, № 55. С. 160—184.

2. *Albasheer H., Siraj Md M., Mubarakali A., Elsier Tayfour O., Salih S., Hamdan M., Kamarudeen S.* Cyber-Attack Prediction Based on Network Intrusion Detection Systems for Alert Correlation Techniques: A Survey // Sensors. 2022. Vol. 22, N 4. P. 1494 (1—15).

3. *Москвичев А. Д., Долгачев М. В.* Алгоритмы корреляции событий информационной безопасности // Автоматизация процессов управления. 2020. № 3. С. 50—59.

4. *Гайфулина Д. А., Котенко И. В.* Анализ моделей глубокого обучения для задач обнаружения сетевых аномалий интернета вещей // Информационно-управляющие системы. 2021. №. 1 (110). С. 28—37.

5. *Kovačević I., Groš S., Slovenec K.* Systematic review and quantitative comparison of cyberattack scenario detection and projection / /Electronics. 2020. Vol. 9, N 10. P. 1722 (1—32).

6. *Kotenko I., Gaifulina D., Zelichenok I.* Systematic Literature Review of Security Event Correlation Methods // IEEE Access. 2022. Vol. 10. P. 43387—43420.

7. *Охтилев М. Ю.* Системы искусственного интеллекта и их применение в автоматизированных системах мониторинга состояния сложных организационно-технических объектов. СПб: СПбГУАП, 2018. 261 с.

8. *Tanwar P., Prasad T. V., Aswal M. S.* Comparative study of three declarative knowledge representation techniques // Intern. Journal on Computer Science and Engineering. 2010. Vol. 2, N 07. P. 2274—2281.

9. *Sikos L. F.* AI in digital forensics: Ontology engineering for cybercrime investigations // Wiley Interdisciplinary Reviews: Forensic Science. 2021. Vol. 3, N 3. P. e1394 (1—11).

10. *Zeng J., Wu S., Chen Y., Zeng R., Wu C.* Survey of attack graph analysis methods from the perspective of data and knowledge processing // Security and Communication Networks. 2019. Vol. 2019. P. 1—16.

11. *Lallie H. S., Debattista K., Bal J.* A review of attack graph and attack tree visual syntax in cyber security // Computer Science Review. 2020. Vol. 35. P. 100219 (1—41).

12. *Маликов А. В., Авраменко В. С., Саенко И. Б.* Модель и метод диагностирования компьютерных инцидентов в информационно-коммуникационных системах, основанные на глубоком машинном обучении // Информационно-управляющие системы. 2019. № 6 (103). С. 32—42.

13. *Бутусов И. В., Романов А. А.* Предупреждение инцидентов информационной безопасности в автоматизированных информационных системах // Вопр. кибербезопасности. 2020. № 5 (39). С. 45—51.

14. *Ala-Laurinaho R., Keski-Heikkilä T.* Driving smart crane with various loads // IEEE Dataport [Электронный ресурс]: <https://ieee-dataport.org/documents/driving-smart-crane-various-loads>.

Сведения об авторе

Диана Альбертовна Левшун — аспирант; СПбФИЦ РАН, СПИИРАН, лаборатория проблем компьютерной безопасности; мл. научный сотрудник;
E-mail: gaifulina@comsec.spb.ru

Поступила в редакцию 18.07.2022; одобрена после рецензирования 28.07.2022; принята к публикации 30.09.2022.

REFERENCES

1. Kotenko I.V., Saenko I.B., Kotsynyak M.A., Lauta O.S. *Informatics and Automation*, 2017, no. 6(55), pp. 160–184. (in Russ.)
2. Albasheer H., Siraj Md M., Mubarakali A., Elsier Tayfour O., Salih S., Hamdan M., Kamarudeen S. *Sensors*, 2022, no. 4(22), pp. 1494(1–15).
3. Moskvichev A.D., Dolgachev M. V. Automation of Control Processes, 2020, no. 3, pp. 50–59. (in Russ.)
4. Gaifulina D.A., Kotenko I.V. *Information and Control Systems*, 2021, no. 1(110), pp. 28–37. (in Russ.)
5. Kovačević I., Groš S., Slovenec K. *Electronics*, 2020, no. 10(9), pp. 1722(1-32).
6. Kotenko I., Gaifulina D., Zelichenok I. *IEEE Access*., 2022, vol. 10, pp. 43387–43420.
7. Okhtilev M.Yu. *Sistemy iskusstvennogo intellekta i ikh primeneniye v avtomatizirovannykh sistemakh monitoringa sostoyaniya slozhnykh organizatsionno-tekhnicheskikh ob"yektov* (Artificial Intelligence Systems and Their Application in Automated Systems for Monitoring the State of Complex Organizational and Technical Objects), St. Petersburg, 2018, 261 p. (in Russ.)
8. Tanwar P., Prasad T.V., Aswal M.S. *International Journal on Computer Science and Engineering*, 2010, no. 07(2), pp. 2274–2281.
9. Sikos L.F. *Wiley Interdisciplinary Reviews: Forensic Science*, 2021, no. 3(3), pp. e1394(1-11).
10. Zeng J., Wu S., Chen Y., Zeng R., Wu C. *Security and Communication Networks*, 2019, vol. 2019, pp. 1–16.
11. Lallie H.S., Debattista K., Bal J. *Computer Science Review*, 2020, vol. 35, pp. 100219(1-41).
12. Malikov A.V., Avramenko V.S., Saenko I.B. *Information and Control Systems*, 2019, no. 6(103), pp. 32–42. (in Russ.)
13. Butusov I., Romanov A. *Voprosy kiberbezopasnosti*, 2020, no. 5(39), pp. 45–51. (in Russ.)
14. Ala-Laurinaho R., Keski-Heikkilä T. *Driving smart crane with various loads*, https://ieee-dataport.org/documents/driving-smart-crane-various-loads.

Data on author

Diana A. Levshun — Post-Graduate Student; St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Laboratory of Computer Security Problems; Junior Researcher; E-mail: gaifulina@comsec.spb.ru

Received 18.07.2022; approved after reviewing 28.07.2022; accepted for publication 30.09.2022.