ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 004.9 ISSN 1995-5499
DOI: https://doi.org/10.17308/sait/1995-5499/2023/3/51-63
Поступила в редакцию 24.11.2022
Подписана в печать 30.09.2023

РАЗРАБОТКА ДВУХЭТАПНОГО МЕТОДА НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СИСТЕМЕ МОНИТОРИНГА КИБЕРБЕЗОПАСНОСТИ СУБЪЕКТОВ ЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

© 2023 В. А. Сизов, А. Д. Киров✉

Российский экономический университет имени Г. В. Плеханова
Стремянный пер., 36, 117997 Москва, Российская Федерация

Аннотация. Работа направлена на повышение эффективности управления кибербезопасностью субъектов экономической деятельности (СЭД) за счет организации эффективного мониторинга кибербезопасности (КБ), учитывающего такие особенности его процесса, как неоднородность источников исходных данных мониторинга КБ, их представление в разных форматах данных, их неточность, во многом неопределённость и зашумлённость, а также большое количество событий информационной безопасности (ИБ), обрабатываемых неоднородными компонентами системы мониторинга КБ СЭД.

В работе предлагается комплексный двухэтапный метод нечеткой кластеризации событий ИБ, учитывающий оценки критичности событий ИБ и функциональные возможности системы мониторинга КБ СЭД. На первом этапе используется модель кластеризации событий ИБ в системе мониторинга КБ СЭД на основе метода нечетких с-средних. Эта модель позволяет кластеризовать множество событий ИБ на 3 нечётких кластера: нечёткий кластер событий ИБ, являющихся инцидентами ИБ, нечёткий кластер событий ИБ, не являющихся инцидентами ИБ и нечёткий кластер событий ИБ, требующих дополнительного анализа. На втором этапе для уточнения результатов кластеризации событий ИБ, полученных на первом этапе, используется модель кластеризации событий ИБ в системе мониторинга КБ СЭД на основе метода выделения α-ядер нечетких кластеров. Эта модель позволяет выбирать вручную пороги степеней принадлежности событий ИБ нечётким кластерам с учетом дополнительной информации и особенностей обработки событий ИБ в системе мониторинга КБ конкретного СЭД. В работе приводится оценка работоспособности разработанного двухэтапного метода нечеткой кластеризации событий ИБ в системе мониторинга КБ СЭД на конкретном примере.

Предложенный подход позволяет повысить эффективность мониторинга КБ СЭД и сократить период времени, необходимый для принятия решения на управление ИБ СЭД за счет комплексного учёта особенностей обработки событий ИБ в системе мониторинга КБ конкретного СЭД.

Ключевые слова: кибербезопасность, информационная безопасность, субъект экономической деятельности, мониторинг, управление, нечёткая кластеризация, событие информационной безопасности, инцидент информационной безопасности.

✉ Киров Алексей Дмитриевич
e-mail: Kirov.AD@rea.ru

ВВЕДЕНИЕ

В настоящее время в IT-сфере, включая КБ активно используются технологии искусственного интеллекта (ИИ). Анализ развития методов ИИ показывает их практическую направленность в решении ряда задач управления КБ, многие из которых получили широкое одобрение профессионалов и высокую оценку их эффективности [1–2].

Создание на основе методов ИИ принципиально новых информационных технологий и их применение в различных высокотехнологичных областях, в том числе в обеспечении требуемого уровня защищенности информационных активов различных субъектов экономической деятельности, мониторинге КБ с целью предотвращения и/или минимизации ущерба от действий злоумышленников приводит к необходимости, с одной стороны, выявления наиболее целесообразных методов ИИ в названной предметной области, а с другой — определения тех задач, в решении которых эти методы могут быть применены наиболее эффективно.

В области управления КБ субъектов экономической деятельности в настоящее время требуется организация эффективного мониторинга КБ, позволяющего принимать рациональные решения на управление ИБ субъекта экономической деятельности (СЭД) в условиях ограниченных ресурсов [3]. Она должна комплексно учитывать такие особенности процессов мониторинга КБ, как неоднородность источников исходных данных мониторинга КБ, их представление в разных форматах данных, их неточность, во многом неопределённость и зашумлённость, а также большое количество событий ИБ, обрабатываемых системой мониторинга КБ СЭД [4]. Эти особенности процессов мониторинга КБ в условиях ограниченных ресурсов не позволяют осуществлять анализ всех данных мониторинга КБ в реальном масштабе времени [5].

Анализ научных методов исследования мониторинга ИБ показал широкое использование методов моделирования процессов мониторинга ИБ. Например, в работе [6] предлагается модель процесса мониторинга ИБ в информационно-телекоммуникационных системах на основе применения аппарата теории марковских случайных процессов. В ней используются методы анализа состояний марковских случайных процессов, протекающих в информационно-телекоммуникационных системах, которые позволяют использовать модель для получения вероятностных и временных зависимостей, описывающих состояния процессов мониторинга ИБ при варьируемых исходных данных входящих и выходящих потоков событий ИБ. Анализ предложенного в этой работе метода показывает, что он может быть использован для выявления аномалий, возникающих в системах мониторинга КБ СЭД. Однако, этот метод не учитывает большие объёмы данных о событиях ИБ, обрабатываемые в системах мониторинга КБ СЭД, которые существенно ограничивают возможности его применения на практике.

В работе [7] предлагается процессная модель мониторинга и реагирования на инциденты ИБ в системах мониторинга КБ СЭД. Предложенная модель основана на классификации процессов мониторинга и реагирования на инциденты ИБ и позволяет идентифицировать определённые ресурсы ИТ-инфраструктуры АИС и ОИ, которые необходимы для реализации системы мониторинга и реагирования на инциденты ИБ. Предложенная модель может быть использована для систематизации процессов мониторинга и управления ИБ в системах мониторинга КБ субъектов экономической деятельности. Однако, она не учитывает параметры времени обработки событий ИБ, напрямую влияющие на время протекания процессов в системах мониторинга КБ крупномасштабных гетерогенных информационно-телекоммуникационных сетей и систем СЭД.

В работе [8] представлены модели таргетного мониторинга ИБ и обработки инцидентов ИБ, построенные с применением методов нечёткой кластеризации.

1. ПОСТАНОВКА ЗАДАЧИ

Для организации эффективного мониторинга ИБ в системе мониторинга КБ СЭД и принятия решений на управление ИБ в ходе анализа событий ИБ и выявления инцидентов ИБ целесообразно учитывать оценки критичности событий ИБ и функциональные возможности системы мониторинга КБ СЭД.

Для решения этой задачи целесообразно использовать хорошо зарекомендовавшие на практике методы нечёткой логики и основанные на них методы нечеткой кластеризации событий ИБ в системе мониторинга КБ СЭД в реальном масштабе времени [9].

Суть задачи состоит в разработке моделей, позволяющих разбивать множества событий мониторинга КБ на кластеры по приоритету их обработки аналитическим отделом или инженером по кибербезопасности. При этом необходимо учитывать ограничения на имеющиеся временные и иные ресурсы системы КБ СЭД, а также, требования к оперативности обработки инцидентов ИБ и допустимому ущербу.

2. МЕТОДЫ И МАТЕРИАЛЫ

На основе анализа процессов в системе мониторинга КБ СЭД и принятия решений на управление ИБ предлагается комплексный метод, учитывающий оценки критичности событий ИБ и функциональные возможности системы мониторинга КБ СЭД. Данный метод включает 2 этапа последовательного применения моделей нечёткой кластеризации, которые позволяют на каждом этапе сужать множество событий ИБ, являющихся возможными инцидентами с учётом имеющихся данных мониторинга КБ СЭД и ограниченных ресурсов, выделяемых на их обработку. На первом этапе используется модель кластеризации нечётких с-средних, на втором этапе используется модель выделения α-ядер нечётких кластеров.

Критериями классификации данных результатов мониторинга КБ, представленных в виде событий ИБ, могут являться степень критичности события ИБ и время обработки события ИБ в системе мониторинга КБ СЭД. Степень критичности события ИБ может быть определена как риск события ИБ для конкретного информационного актива [10]. Такой риск может быть задан матрицей соответствия активов СЭД событиям ИБ, или их совокупностям и формулой расчёта риска события ИБ. Пусть M — матрица соответствия активов СЭД, об ущербе которым свидетельствует появление в базе данных системы мониторинга КБ СЭД записи об определённом событии ИБ. Тогда она может быть представлена в виде выражения 1:

$$M = \left\| \mathbf{m_{ij}} \right\|, \quad i = \overline{1,I}, j = \overline{1,J}, \qquad (1)$$

где $\mathbf{m_{ij}}$ — вероятность того, что появление в базе данных системы мониторинга КБ СЭД записи об i-м событии ИБ свидетельствует об ущербе j-му активу СЭД, $\mathbf{m_{ij}} \in [0..1]$, I — количество записей о событиях ИБ, содержащихся в базе данных системы мониторинга КБ СЭД, J — количество активов СЭД. Пример такой матрицы представлен в табл. 1.

Таким образом, риск i-го события ИБ для j-го информационного актива может быть представлен выражением 2:

$$R_{ij} = \sum_{i=1}^{I} \sum_{j=1}^{J} m_{ij} \times P_{ij} \times d_{ij}, \quad i = \overline{1,I}, j = \overline{1,J}, \qquad (2)$$

где R_{ij} — риск i-го события ИБ для j-го информационного актива, m_{ij} — вероятность того, что появление в базе данных системы мониторинга КБ СЭД записи об i-м событии ИБ свидетельствует о реальном или потенциальном ущербе j-му активу СЭД, $m_{ij} \in [0..1]$, P_{ij} — вероятность нанесения ущерба j-му активу СЭД в результате воздействия, о котором свидетельствует появление в базе данных системы мониторинга КБ СЭД записи об i-м событии ИБ, $P_{ij} \in [0..1]$, d_{ij} — максимальный ущерб, который может быть нанесён j-му активу СЭД в результате воздействия, о котором свидетельствует появление в базе системы мониторинга КБ СЭД записи об i-м событии ИБ, I — количество записей о событиях ИБ, содержащихся в базе данных системы мониторинга КБ СЭД, J — количество активов СЭД.

Предложенные критерии позволяют не только оценить степень влияния событий и

Таблица 1. Пример матрицы соответствия активов СЭД событиям ИБ (матрицы M)
[Table 1. Example of a matrix of correspondence between EDS assets and information security events (matrices M)]

Активы СЭД / События ИБ	Сервер	Маршрутизатор	Рабочая станция	База данных	Веб-ресурс
Со стороны сервера начата установка программного обеспечения на рабочей станции	1	0,6	1	0,2	0
Отказ во входе на рабочую станцию в связи с неправильным паролем	0,8	0,1	1	0,7	0
Изменение пароля учётной записи пользователя joomla_admin	1	0,4	0	0,9	0,8
Обнаружено заражение маршрутизатора вредоносным ПО	0,3	1	0,6	0	0
Обнаружен несанкционированный доступ к базе данных	0,6	0,3	0,8	1	0,4

инцидентов ИБ на информационные активы СЭД, но определить конкретные значения рисков ИБ по отношению к информационным активам СЭД.

Для эффективной работы систем мониторинга КБ СЭД целесообразна разработка моделей кластеризации событий ИБ, обрабатываемых в таких системах с учетом их возможностей. Эти модели должны позволять подразделять события ИБ на события, с достаточно высокой вероятностью определяющие инциденты ИБ, обработка которых требуется в первую очередь, события, с достаточно высокой вероятностью не приводящие к инцидентам ИБ и остальные события ИБ, решение относительно которых может быть принято только по итогам дополнительного анализа [11]. В связи с особенностями процессов мониторинга КБ, зачастую, имеющих неопределенный и нечеткий характер, а также необходимостью комплексного учёта ограничений на ресурсы систем мониторинга КБ СЭД, в разрабатываемых моделях целесообразно использовать методы многоэтапной нечёткой кластеризации [12].

3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В данной работе предлагается двухэтапный метод нечеткой кластеризации событий ИБ. На первом этапе используется модель кластеризации событий ИБ в системе мониторинга КБ СЭД на основе метода нечетких с-средних. Применение метода нечетких с-средних позволяет автоматически определять пороги степеней принадлежности событий ИБ нечётким кластерам при помощи анализа свойств событий ИБ и тем самым выделять наиболее важные нечеткие множества событий ИБ: нечеткое множество событий — инцидентов ИБ и нечеткое множество событий ИБ, не представляющих угрозу ИБ.

На втором этапе для уточнения результатов кластеризации событий ИБ, полученных на первом этапе, используется модель кластеризации событий ИБ в системе мониторинга КБ СЭД на основе метода выделения α-ядер нечетких кластеров. Эта модель позволяет выбирать вручную пороги степеней принадлежности событий ИБ нечётким кластерам с учетом особенностей обработки событий ИБ в системе мониторинга КБ конкретного СЭД (возможностей оперативного наращивания сил и средств обработки событий ИБ; использования пакетов экспертиз; интеллектуального потенциала и др.).

3.1. Разработка модели кластеризации событий информационной безопасности в системе мониторинга кибербезопасности субъектов экономической деятельности на основе метода нечетких с-средних

Пусть $S = \{s_1, s_2, ..s_i\}$, $i = \overline{1, I}$, где S — множество событий ИБ, содержащихся в базе данных системы мониторинга КБ СЭД, s_i — i-е событие ИБ, содержащееся в базе данных системы мониторинга КБ СЭД, I — количество событий ИБ, содержащихся в базе данных системы мониторинга КБ СЭД. Тогда задачу нечеткой кластеризации событий ИБ в системе мониторинга КБ СЭД можно свести к задаче разбиения множества S на непересекающихся 3 нечетких подмножества (кластера): S_0 — нечёткий кластер событий ИБ, являющихся инцидентами ИБ, S_1 — нечёткий кластер событий ИБ, не являющихся инцидентами ИБ, S_r — нечёткий кластер событий ИБ, требующих дополнительного анализа и являющихся инцидентами ИБ со степенью уверенности r, $r \in]0,1[$.

Таким образом,

$$S = S_0 \cup S_1 \cup S_r,$$
$$S_0 \cap S_1 = \varnothing,$$
$$S_0 \cap S_r = \varnothing,$$
$$S_1 \cap S_r = \varnothing.$$

Разбиение множества событий ИБ S производится при помощи применения алгоритма нечёткой кластеризации с-средних [13].

Метод нечёткой кластеризации с-средних предназначен для разбиения имеющегося множества событий ИБ на заданное число нечётких подмножеств. Метод нечеткой кластеризации с-средних можно рассматривать как усовершенствованный метод нечеткой кластеризации k-средних, при котором для каждого элемента из рассматриваемого множества рассчитывается степень его принадлежности каждому из кластеров.

Нечеткая кластеризация с-средних объединяет в себе сущность нечеткой теории. По сравнению с жесткой кластеризацией k-средних, она обеспечивает более гибкие результаты кластеризации. Поскольку в большинстве случаев события ИБ в наборе данных не могут быть разделены на четко разделенные кластеры, назначение события ИБ конкретному кластеру является немного грубым, а также могут возникать ошибки. Следовательно, каждому событию ИБ и каждому кластеру присваивается вес, чтобы указать, в какой степени объект принадлежит кластеру. Конечно, вероятностные методы также могут давать такие веса, но в большинстве случаев на практике трудно определить подходящую статистическую модель, поэтому лучше использовать нечеткие с-средние с естественными и не вероятностными характеристиками, что аналогично минимизации целевой функции (в некоторых данных она представляет собой сумму квадратов ошибки).

Нечеткая кластеризация с-средних — это процесс итеративного вычисления степени членства и центра кластера до тех пор, пока они не достигнут оптимума.

Смысл алгоритма нечёткой кластеризации с-средних состоит в том, чтобы присвоить каждой выборке событий ИБ функцию принадлежности, которая принадлежит каждому кластеру. Образцы классифицируются по степени значимости членства.

На основе [14] можно выделить следующий обобщённый список шагов алгоритма нечеткой кластеризации с-средних.

1. Инициализация

Обычно используется случайная инициализация. То есть веса выбираются случайным образом. Количество кластеров нужно выбирать вручную.

2. Вычисление центроида.

Центроид в методе нечётких с-средних отличается от традиционного центроида тем, что он использует степень принадлежности в качестве веса для получения средневзвешенного значения.

3. Обновление нечеткого псевдоделения.

Таким образом, алгоритм нечёткой кластеризации с-средних можно представить следующим образом.

Вход: множество S, коэффициент фаззификации m.

Выход: разбиение множества S на 3 кластера: S_0, S_1, S_r.

Шаг 1: начало;

Шаг 2: инициализация значений вектора центров нечётких кластеров V. пусть $l=0$, $\varepsilon>0$, $m=1$, $D^2(i,k)=\left|s_i-V_k\right|^2$, $U(k)$ — центры нечётких кластеров, $k=1,2,3$, $i=\overline{1,I}$;

Шаг 3: определить $U(k)$ согласно формуле 3:

$$U(i,k)=\left(\sum_{j=1}^{C}\left(\frac{D(i,k)}{D(i,j)}\right)^{\frac{2}{m-1}}\right)^{-1}, \qquad (3)$$

где $C=3$ — количество нечётких кластеров;

Шаг 4: обновить значение V согласно формуле 4:

$$V_k=\frac{\sum_{i=1}^{I}U(i,k)^m\times s_i}{\sum_{i=1}^{I}U(i,k)^m}. \qquad (4).$$

Шаг 5: если $\left|V^l-V^{l+1}\right|<\varepsilon$, то перейти к шагу 6, иначе увеличить l на 1 и перейти к шагу 3;

Шаг 6: конец.

3.2. Разработка модели кластеризации событий информационной безопасности в системе мониторинга кибербезопасности субъектов экономической деятельности на основе метода выделения α-ядер нечетких кластеров

Для разработки модели кластеризации событий ИБ в системе мониторинга КБ СЭД целесообразно использовать метод выделения α-ядер нечетких кластеров [15]. Суть этого метода состоит в том, что исходя из степеней квалификации злоумышленника и специалиста по КБ СЭД выбираются пороги критичности событий ИБ $\alpha\in(0,1]$ [16]. Затем для каждого из выбранных пороговых значений $\alpha\in(0,1]$ определяется расстояние от степени принадлежности (степени членства в нечётком кластере) каждого события ИБ до α. Если вычисленное расстояние от степени членства каждого события ИБ в нечётком кластере до α меньше, чем расстояние от степени членства каждого события ИБ в нечётком кластере до центра нечёткого кластера, вычисленного при помощи применения алгоритма нечёткой кластеризации с-средних, то соответствующее событие ИБ считается принадлежащим нечёткому кластеру с соответствующим значением α.

Поэтому его применение в модели кластеризации событий ИБ в системе мониторинга КБ СЭД позволяет учитывать такие особенности обработки событий ИБ в системе мониторинга КБ, как учёт связанной с событиями ИБ дополнительной информации (например, пакета экспертиз), а также, обеспечить возможность обработки в таких системах событий ИБ различных типов и структур.

Задача формулируется следующим образом.

Дано: S_r — множество событий ИБ, являющихся инцидентами ИБ со степенью уверенности r, $r\in]0,1[$.

Задача: необходимо разбить множество S_r на n непересекающихся нечётких подмножеств (кластеров):

$$S_r=S_0\cup S_1\cup..\cup S_n.$$

Разбиение множества событий ИБ S_r, полученного при помощи применения алгоритма нечёткой кластеризации с-средних, на n непересекающихся нечётких подмножеств (кластеров) производится методом выделения α-ядер нечетких кластеров.

Метод выделения α-ядер нечетких кластеров предполагает нахождение такого порога α, $\alpha\in(0,1]$, чтобы выполнялось условие 5:

$$\sum_{l=1}^{C}card\left(Supp\left(B^k\left(\alpha\right)\right)\right)\geq card\left(S_r\right), \qquad (5)$$

где $S_r=\{s_r^1,s_r^2,...,s_r^j\}$, $j=\overline{1,J}$ — множество событий ИБ, являющихся инцидентами ИБ со степенью уверенности r, $r\in]0,1[$, s_r^j — j-е событие ИБ, являющееся инцидентом ИБ со степенью уверенности r, $r\in]0..1[$, J — количество событий ИБ, являющихся инцидентами ИБ со степенью уверенности r, α-ядра $B^k(\alpha)$, $k\in\{1,2\}$ нечётких кластеров $B^k\in P$, $k\in\{1,2\}$ для некоторого $\alpha\in]0..1[$ представляют собой нечёткие множества уровня, определяемые как $B^k(\alpha)=\{(s_j,\mu_{kj}^{\alpha})\mid\mu_{kj}^{\alpha}\geq\alpha\}$.

Для выбранного порога $\alpha\in]0..1[$ определяется расстояние от степени принадлежности (степени членства в нечётком кластере)

каждого события ИБ до α. Если вычисленное расстояние от степени членства каждого события ИБ в нечётком кластере до α меньше, чем расстояние от степени членства каждого события ИБ в нечётком кластере до центра нечёткого кластера, вычисленного при помощи применения алгоритма нечёткой кластеризации с-средних, то соответствующее событие ИБ считается принадлежащим нечёткому кластеру с соответствующим значением α.

Применительно к задаче кластеризации событий ИБ, α целесообразно представить как величину, зависящую от характеристик сил и средств в области ИБ, анализирующих и обрабатывающих события ИБ [16].

Концепция α-ядер нечетких кластеров событий ИБ позволяет, с одной стороны, отнести каждое событие ИБ к наименьшему числу нечетких кластеров, являющегося результатом классификации, а с другой — сохранить значения принадлежности каждого события ИБ кластеру инцидентов ИБ или кластеру событий, «похожих» на инциденты ИБ или кластеру других событий ИБ.

Именно на этом этапе события ИБ категорируются, путём их привязки к угрозам ИБ и техникам, применяемым злоумышленниками для реализации этих угроз.

3.3. Оценка работоспособности двухэтапного метода нечеткой кластеризации событий информационной безопасности в системе мониторинга кибербезопасности субъектов экономической деятельности

Для оценки работоспособности двухэтапного метода нечеткой кластеризации событий информационной безопасности в системе мониторинга кибербезопасности субъектов экономической деятельности проведен компьютерный эксперимент.

Пусть $S = \{s_1, s_2, s_3, s_4, s_5\}$, где S — множество, состоящее из 5 событий ИБ. Тогда для решения задачи разбиения множества S на непересекающихся 3 подмножества (кластера): S_0 — подмножество событий ИБ, являющихся инцидентами ИБ со степенью уверенности 0, S_1 — подмножество событий ИБ, являющихся инцидентами ИБ со степенью уверенности 1, S_r — подмножество событий ИБ, являющихся инцидентами ИБ со степенью уверенности r, $r \in]0,1[$, используется программный макет, представляющий собой реализацию метода кластеризации нечётких с-средних, выполненную на языке программирования высокого уровня Python [17]. В качестве исходных данных для программного макета выступает S — множество, состоящее из 5 случайно выбранных событий ИБ.

На рис. 1 показаны результаты применения метода нечёткой кластеризации с-средних, реализованного в виде программного макета, выполненного на языке программирования высокого уровня Python ко множеству событий ИБ S, где подмножество событий ИБ S_0 обозначено как кластер 1, подмножество событий ИБ S_1 обозначено как кластер 3, подмножество событий ИБ S_r обозначено как кластер 2, что связано с особенностями используемой реализации метода кластеризации нечётких с-средних.

Анализ данных, представленных на рис. 1 показывает, что все события из множества S разделены на 3 кластера. В первый кластер входят события, с достаточно высокой степенью уверенности, не входящие в состав инцидентов ИБ. Во второй кластер входят события ИБ, решение относительно которых может быть принято только по итогам дополнительного анализа, который должен быть проведён на 2 этапе алгоритма многоэтапной нечёткой кластеризации. В третий кластер входят события, с достаточно высокой степенью уверенности входящие в состав инцидентов ИБ, обработка которых требуется в первую очередь.

На рис. 2 показаны результаты применения алгоритма нечёткой кластеризации с-средних, а затем, применения к полученным результатам метода выделения α-ядер нечетких кластеров ко множеству событий ИБ, являющихся инцидентами ИБ со степенью уверенности r S_r.

Анализ данных, представленных на рис. 2 показал, что после применения к событиям ИБ, входящим в множество S, алгоритма нечёткой кластеризации с-средних, а затем, применения метода выделения α-ядер нечет-

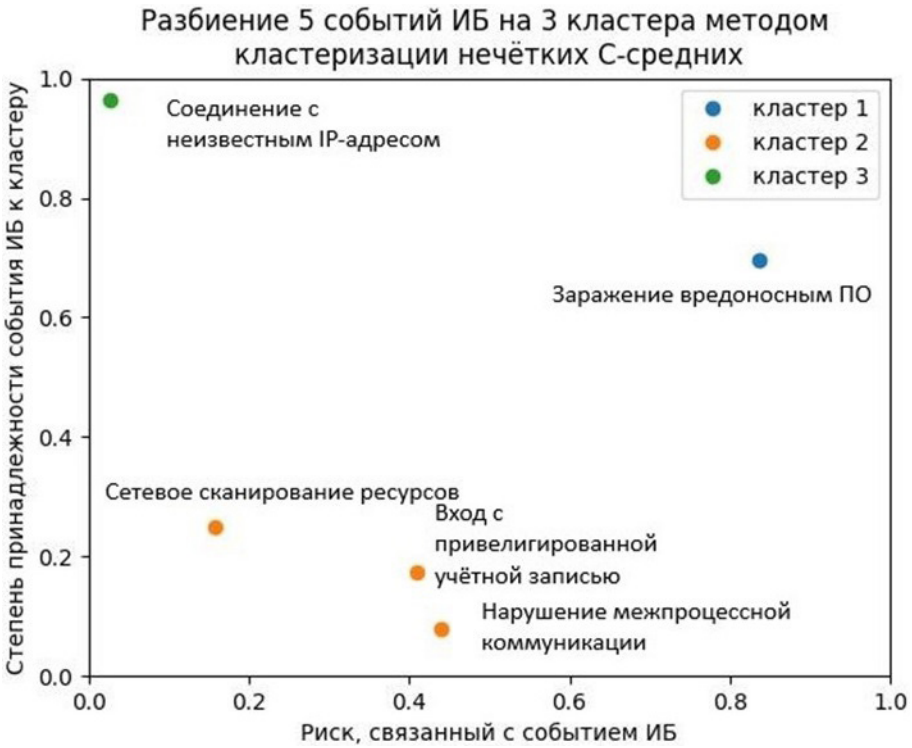

Рис. 1. Кластеризованное методом нечёткой кластеризации с-средних множество событий ИБ S

[Fig. 1. Clustered set of information security events S using the c-means fuzzy clustering method]

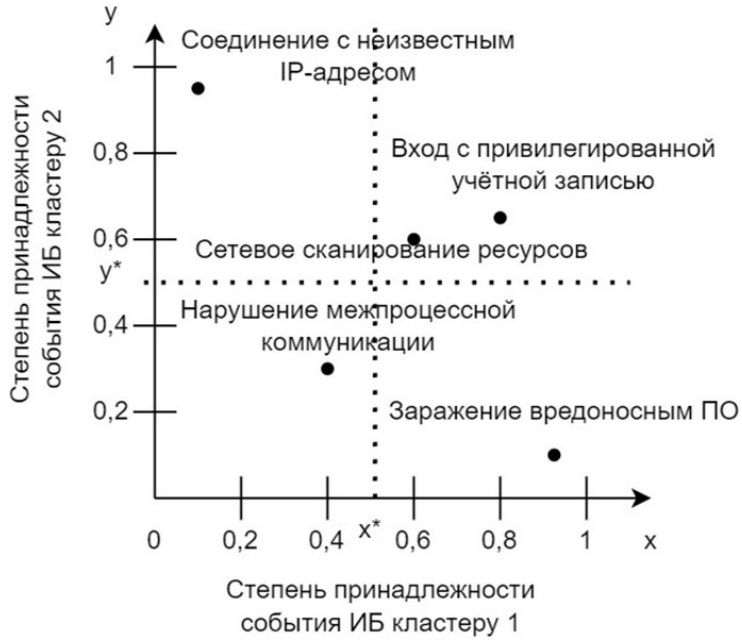

Рис. 2. Множество событий ИБ, являющихся инцидентами ИБ со степенью уверенности r, кластеризованное методом выделения α-ядер нечетких кластеров

[Fig. 2. A set of information security events that are information security incidents with a degree of confidence, clustered by the method of extracting -kernels of fuzzy clusters]

ких кластеров ко множеству событий ИБ, являющихся инцидентами ИБ со степенью уверенности r S_r, это множество событий ИБ разделено на 2 подмножества: S_0 — нечеткое подмножество событий ИБ, являющихся возможно инцидентами ИБ, S_1 — нечеткое под-

множество событий ИБ, скорее всего не являющихся инцидентами ИБ. Соответственно, обработка событий ИБ, входящих в подмножество S_0 требуется в первую очередь.

ЗАКЛЮЧЕНИЕ

Таким образом, предложенный подход к организации эффективного мониторинга ИБ в системе мониторинга КБ СЭД на основе многоэтапной нечёткой кластеризации позволяет последовательно решать задачи выделения критически важных данных результатов мониторинга КБ СЭД и комплексно учитывать не только параметры процессов мониторинга ИБ, но и параметры времени обработки событий ИБ и рисков ИБ, связанных с рассматриваемыми событиями ИБ в условиях ограниченных ресурсов системы мониторинга КБ.

В целом такой подход позволяет повысить эффективность мониторинга КБ СЭД и сократить периоды времени, необходимые для принятия решений на управление ИБ СЭД за счет комплексного учёта особенностей обработки событий ИБ в системе мониторинга КБ конкретного СЭД.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Chiheb Chebbi*. Mastering Machine Learning for Penetration Testing / Chebbi Chiheb; Birmingham B3 2PB. – UK : Packt Publishing. – 2018. – 264 p.

2. *Onyango Oscar*. Artificial Intelligence and its Application to Information Security Management / Oscar Onyango. 10.13140/RG.2.2.12066.09921.

3. *Сизов В. А.* Проблемы внедрения SIEM-систем в практику управления информационной безопасностью субъектов экономической деятельности / В. А. Сизов, А. Д. Киров // Открытое образование. – 2020. – Т. 2, № 1. – С. 69–79. https://doi.org/10.21686/1818-4243-2020-1-69-79

4. *Вульфин А. М.* Обнаружение сетевых атак в гетерогенной промышленной сети на основе технологий машинного обучения / А. М. Вульфин // Программная инженерия. – 2022. – Т. 13, № 2. – С. 68–80. – DOI 10.17587/prin.13.68-80. – EDN LYURNN

5. *Kotenko Igor*. Model of security information and event management system / Igor Kotenko, Igor Parashchuk, // Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics. 2020. – P. 84-94. – 10.24143/2072-9502-2020-2-84-94.

6. *Ерышов В. Г.* Модель процесса мониторинга информационной безопасности в информационно-телекоммуникационных системах на основе применения аппарата теории марковских случайных процессов / В. Г. Ерышов, Д. В. Ильина // Волновая электроника и инфокоммуникационные системы : Сборник статей XXIII международной научной конференции, Санкт-Петербург, 01–05 июня 2020 года. – Санкт-Петербург: Санкт-Петербургский государственный университет аэрокосмического приборостроения, 2020. – С. 236-242. – EDN BDOQWB.

7. *Королев В. И.* Процессная модель мониторинга и реагирования на инциденты информационной безопасности / В. И. Королев // Информационная безопасность: вчера, сегодня, завтра : Сборник статей по материалам III Международной научно-практической конференции, Москва, 23 апреля 2020 года. – Москва: Российский государственный гуманитарный университет, 2020. – С. 18–25. – EDN APZTCW.

8. *Kirov A.* Development of a method for targeted monitoring and processing of information security incidents of economic entities / A. Kirov, V. Sizov // J Comput Virol Hack Tech. – 2022. – P. 1–6. https://doi.org/10.1007/s11416-022-00449-8

9. *Kaffah F. M.* Implementation of the fuzzy logic for measuring instrument evaluation results in Information Security Index / F. M. Kaffah, M. Irfan, C. Slamet, C. Berhat, A. B. A. Rahman // IOP Conference Series Materials Science and

Engineering. – 2021. – 1098.062003 – P. 1–8 – 10.1088/1757-899X/1098/6/062003.

10. *Lee Ming-Chang*. Information Security Risk Analysis Methods and Research Trends: AHP and Fuzzy Comprehensive Method / Ming-Chang Lee // International Journal of Computer Science and Information Technology. – 2014. – № 6. – P. 29–45.

11. *Сидорова Д. Н.* Алгоритмы и методы кластеризации данных в анализе журналов событий информационной безопасности / Д. Н. Сидорова, Е. Н. Пивкин // Безопасность цифровых технологий. – 2022. – № 1(104). – С. 41–60. – DOI 10.17212/2782-2230-2022-1-41-60. – EDN RMDHEC.

12. *Вятченин Д. А.* Методология анализа данных на основе многоэтапной нечеткой кластеризации / Д. А. Вятченин // Объединенный институт проблем информатики Национальной академии наук Беларуси. Искусственный интеллект. – 2009.

13. *Аль-Раммахи Али Абидалкарим Хабиб Х*. Модификация метода нечеткой кластеризация с-средних с использованием метода роя частиц для обработки больших данных / Х. Аль-Раммахи Али Абидалкарим Хабиб, А. Сари Фарах Аббас, Ю. В. Минин // Современная наука: теория, методология, практика : Материалы 1-й Всероссийской (национальной) научно-практической конференции, Тамбов, 26–27 ноября 2019 года. – Тамбов: Издательство Першина Р.В., 2019. – С. 231–233. – EDN RTAOOJ.

14. *Tran Khang*. Fuzzy C-Means Clustering Algorithm with Multiple Fuzzification Coefficients / Tran Khang, Vuong Nguyen, Tran Manh-Kien, Fowler, Michael // Algorithms. – 2020. – V. 13 – P. 158. – 10.3390/a13070158

15. *Вятченин Д. А.* Нечеткие методы автоматической классификации / Д. А. Вятченин. – Минск : УП Технопринт, 2004. – 219 с.

16. *Сизов В. А.* Разработка моделей аналитической системы обработки данных для мониторинга ИБ объекта информатизации, использующего облачную инфраструктуру / В. А. Сизов, А. Д. Киров // Российский технологический журнал. – 2021. – С. 16–25.

17. *James C.* Bezdek FCM: The fuzzy c-means clustering algorithm / James C. Bezdek, Robert Ehrlich, William Full. // Computers & Geosciences. – № 10. – P. 191–203.

Сизов Валерий Александрович — д-р техн. наук, проф., профессор кафедры прикладной информатики и информационной безопасности, Российский экономический университет им. Г. В. Плеханова.
E-mail: Sizov.VA@rea.ru
ORCID iD: https://orcid.org/0000-0002-4844-4714

Киров Алексей Дмитриевич — аспирант кафедры прикладной информатики и информационной безопасности, Российский экономический университет им. Г. В. Плеханова.
E-mail: Kirov.AD@rea.ru
ORCID iD: https://orcid.org/0000-0002-8424-3071

DOI: https://doi.org/10.17308/sait/1995-5499/2023/3/51-63 ISSN 1995-5499
Received 24.11.2022
Accepted 30.09.2023

DEVELOPMENT OF A TWO-STAGE METHOD OF FUZZY CLUSTERING OF INFORMATION SECURITY EVENTS IN THE CYBER SECURITY MONITORING SYSTEM OF ECONOMIC ACTIVITIES SUBJECTS

© 2023 V. A. Sizov, A. D. Kirov✉

Plekhanov Russian Univercity of Economics
36, Stremyanny Lane, 117997 Moscow, Russian Federation

Annotation. The work is aimed at improving the efficiency of managing the cyber security of economic entities (EDS) by organizing effective cyber security (CS) monitoring, taking into account such features of its process as the heterogeneity of sources of initial data for monitoring CB, their presentation in different data formats, their inaccuracy, in many respects uncertainty and noise, as well as a large number of information security (IS) events processed by heterogeneous components of the EDMS KB monitoring system.

The paper proposes a comprehensive two-stage method for fuzzy clustering of IS events, considering the assessment of the criticality of IS events and the functionality of the monitoring system of KB EDMS. At the first stage, the IS event clustering model is used in the EDMS KB monitoring system based on the fuzzy c-means method. This model allows clustering a set of IS events into 3 fuzzy clusters: a fuzzy cluster of IS events that are IS incidents, a fuzzy cluster of IS events that are not IS incidents, and a fuzzy cluster of IS events that require additional analysis. At the second stage, to refine the results of IS event clustering obtained at the first stage, the IS event clustering model is used in the EDMS KB monitoring system based on the method of extracting α-kernels of fuzzy clusters. This model allows you to manually select the thresholds for the degree of belonging of IS events to fuzzy clusters, considering additional information and features of processing IS events in the IS monitoring system of a particular EDMS. The paper provides an assessment of the performance of the developed two-stage method of fuzzy clustering of IS events in the monitoring system of KB EDMS using a specific example.

The proposed approach makes it possible to increase the efficiency of ERMS IS monitoring and reduce the period of time required to decide on the management of IS EDMS due to the complex consideration of the features of processing IS events in the IS monitoring system of a particular EDMS.

Keywords: cybersecurity, information security, economic entity, monitoring, management, fuzzy clustering, information security event, information security incident.

CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. *Chebbi C.* (2018) Mastering machine learning for penetration testing: Develop an extensive skill set to break self-learning systems using Python. Birmingham: Packt Publishing Ltd.

2. *Onyango Oscar*. Artificial Intelligence and its Application to Information Security Management. 10.13140/RG.2.2.12066.09921.

3. *Sizov V. A. and Kirov A. D.* (2020) Problems of implementing SIEM systems in the practice of managing information security of Economic Entities. Open Education. 24(1). P. 69–79. (In Russian) https://doi.org/10.21686/1818-4243-2020-1-69-79.

4. *Vulfin A. M.* (2022) Detection of network attacks in a heterogeneous industrial network based on machine learning. Programmnaya Ingeneria. 13(2). P. 68–80. (In Russian) https://doi.org/10.17587/prin.13.68-80.

✉ Kirov Alexey D.
e-mail: Kirov.AD@rea.ru

5. *Kotenko Igor and Igor Parashchuk* (2020). Model of security information and event management system. Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics. P. 84–94. 10.24143/2072-9502-2020-2-84-94.

6. *Eryshov V. G. and Ilina D. V.* (2020) Method of the information security monitoring process in information and telecommunication systems based on the application of methods of Markov Random Processes. 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) [Preprint]. (In Russian) https://doi.org/10.1109/weconf48837.2020.9131492.

7. *Korolev V. I.* (2020) Protsessnaia model' monitoringa i reagirovaniia na intsidenty informatsionnoi bezopasnosti. Informatsionnaia bezopasnost': vchera, segodnia, zavtra : Sbornik statei po materialam III Mezhdunarodnoi nauchno-prakticheskoi konferentsii, Moskva, 23 aprelia 2020 goda [Information Security: Yesterday, Today, Tomorrow: Collection of articles based on the materials of the III International Scientific and Practical Conference, Moscow, April 23, 2020]. Moskva: Rossiiskii gosudarstvennyi gumanitarnyi universitet. P. 18–25. (In Russian) – EDN APZTCW.

8. *Kirov A. and Sizov V.* (2022) Development of a method for targeted monitoring and processing of information security incidents of economic entities. J Comput Virol Hack Tech. P. 1–6. https://doi.org/10.1007/s11416-022-00449-8

9. *Kaffah F. M., Irfan M., Slamet C., Berhat C., Rahman A. B. A.* (2021) Implementation of the fuzzy logic for measuring instrument evaluation results in Information Security Index. IOP Conference Series Materials Science and Engineering. 1098.062003. P. 1–8. 10.1088/1757-899X/1098/6/062003.

10. *Lee Ming-Chang* (2014) Information Security Risk Analysis Methods and Research Trends: AHP and Fuzzy Comprehensive Method. International Journal of Computer Science and Information Technology. № 6. P. 29–45.

11. *Sidorova D. N., Pivkin E. N.* (2022) Algoritmy i metody klasterizatsii dannykh v analize zhurnalov sobytii informatsionnoi bezopasnosti [Algorithms and methods of data clustering in the analysis of information security event logs]. Bezopasnost' tsifrovykh tekhnologii = Digital Technology Security. No 1 (104). P. 41–60. (In Russian) DOI: 10.17212/2782-2230-2022-1-41-60.

12. *Viattchenin D. A.* (2009) Methodology of Data Analysis Based on Multistage Fuzzy Clustering. United Institute of Informatics Problems of the National Academy of Sciences of Belarus. Artificial intelligence. (In Russian)

13. *Al'-Rammakhi Ali Abidalkarim Khabib Kh., Sari Farakh Abbas A. and Minin Iu. V.* (2019) Modifikatsiia metoda nechetkoi klasterizatsiia s-srednikh s ispol'zovaniem metoda roia chastits dlia obrabotki bol'shikh dannykh. Sovremennaia nauka: teoriia, metodologiia, praktika : Materialy 1-i Vserossiiskoi (natsional'noi) nauchno-prakticheskoi konferentsii, Tambov, 26–27 noiabria 2019 goda [Modern science: theory, methodology, practice: Proceedings of the 1st All-Russian (national) scientific and practical conference, Tambov, November 26–27, 2019.]. Tambov : Izdatel'stvo Pershina R.V., P. 231–233. (In Russian) – EDN RTAOOJ.

14. *Tran Khang, Vuong Nguyen, Tran Manh-Kien and Fowler Michael* (2020) Fuzzy C-Means Clustering Algorithm with Multiple Fuzzification Coefficients. Algorithms. V. 13. P. 158. 10.3390/a13070158

15. *Viattchenin D. A.* (2004) Fuzzy automatic classification methods. Minsk : Techno print. 219 p.

16. *Sizov V. A. and Kirov A. D.* (2021) The development of models of an analytical data processing system for monitoring information security of an informatization object using cloud infrastructure. Russian Technological Journal. 9(6). P. 16–25. (In Russian) https://doi.org/10.32362/2500-316X-2021-9-6-16-25

17. *James C. Bezdek, Ehrlich Robert and Full William*. FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences. No 10. P. 191–203.

Sizov Valerii A. — doctor of sciences in technology, professor, professor of the Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics.
E-mail: Sizov.VA@rea.ru
ORCID iD: https://orcid.org/0000-0002-4844-4714

Kirov Alexey D. — postgraduate student of the Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics.
E-mail: Kirov.AD@rea.ru
ORCID iD: https://orcid.org/0000-0002-8424-3071