DOI 10.37882/2223-2966.2024.05.23

РАЗРАБОТКА УНИВЕРСАЛЬНОГО МЕТОДА ПРОГНОЗИРОВАНИЯ ПОВЕДЕНИЯ ОБЪЕКТОВ В ИНФОРМАЦИОННОЙ СИСТЕМЕ: АНАЛИЗ АКТУАЛЬНЫХ РАЗРАБОТОК В ОБЛАСТИ ПОВЕДЕНЧЕСКОЙ АНАЛИТИКИ И МАТЕМАТИЧЕСКИХ ТЕОРИЙ

DEVELOPMENT OF A UNIVERSAL METHOD FOR PREDICTING THE BEHAVIOR OF OBJECTS IN AN INFORMATION SYSTEM: ANALYSIS OF CURRENT DEVELOPMENTS IN THE FIELD OF BEHAVIORAL ANALYTICS AND MATHEMATICAL THEORIES

D. Mokhorev

Summary. The study considers the possibility of creating a universal method for predicting the behavior of objects in an information system. To this end, an analysis of current developments in the field of behavioral analytics is carried out and the main mathematical theories underlying them are highlighted. The criteria of universality of algorithms for analyzing the behavior of objects are presented, which make it possible to assess the possibilities and limitations of their use in various conditions and software environments. As a result of the analysis, the problems preventing the creation of a universal method for predicting the behavior of objects in an information system are formulated.

Keywords: information security, machine learning, behavioral analysis, UEBA, Bayes theorem, time series analysis, fuzzy clustering method.

Мохорев Дмитрий Евгеньевич
Аспирант,
ФГБОУ ВО «РЭУ им. Г.В. Плеханова» РЭУ им. Г.В. Плеханова
mokhorev @rea.ru

Аннотация. В исследовании рассмотрена возможность создания универсального метода прогнозирования поведения объектов в информационной системе. С этой целью проведен анализ актуальных разработок в области поведенческой аналитики и выделены основные математические теории, лежащие в их основе. Представлены критерии универсальности алгоритмов анализа поведения объектов, позволяющие оценить возможности и ограничения их использования в различных условиях и программных средах. В результате анализа сформулированы проблемы, препятствующие созданию универсального метода прогнозирования поведения объектов в информационной системе.

Ключевые слова: информационная безопасность, машинное обучение, поведенческий анализ, UEBA, теорема Байеса, анализ временных рядов, метод нечеткой кластеризации.

Введение

Экспоненциальный рост объема обрабатываемых данных уже давно является неотъемлемой частью развития любой современной информационной системы. На первый взгляд, увеличение количества информации только усложняет ее анализ, однако это не так: к большим массивам данных эффективнее применяются функции статистической математики, что позволяет точнее выявлять закономерности и тренды в их изменении. Вопросы, связанные со сбором, хранением и обработкой больших объемов данных, а также получение на их основе ценной для принятия решений информации, рассматриваются в рамках анализа больших данных (Big data analysis). Анализ больших данных, как область научного знания, объединяет математику, статистику и информатику.

Решение задач анализа больших данных представляет собой поиск и выявление закономерностей в массиве данных. Для этого, на основании исходной выборки значений, создается математическая функция, описывающая их изменения. Это процесс называется машинным обучением (ML, Machine Learning), а его результат — моделью искусственного интеллекта (AI, Artificial Intelligence) [2, с. 2]. В связи с устойчивым ростом объема обрабатываемой информации анализ больших данных является одним из наиболее актуальных направлений исследований в сфере информационных технологий, а машинное обучение — основным и активно развивающимся методом анализа.

Глобальная цифровизация привела к тому, что все современные бизнес-процессы организованы через построение информационных систем, в основе которых лежат цифровые данные. Это способствует повсеместному применению машинного обучения для решения различных задач. Существенный результат от использования данной технологии можно получить в сфере защиты информации. Это связано с тем, что традиционные ме-

тоды выявления нарушений безопасности информации основаны на сигнатурном анализе, который заключается в отслеживании ограниченного и заранее известного набора признаков реализации угрозы, индикаторов компрометации (IOC, Indicator of Compromise) [3].

Применение искусственного интеллекта позволяет осуществлять качественно иной подход к мониторингу — поведенческий анализ. Анализ и прогнозирование поведения объектов в информационных системах как метод контроля защищенности информационной инфраструктуры заключается в сборе сведений о состоянии элементов автоматизированных систем (АС), формировании с помощью машинного обучения базовых моделей поведения данных элементов и регистрации аномалий при выявлении отклонений от них. Среди отслеживаемых объектов могут быть, как отдельные пользователи АС, автоматизированные рабочие места, сервера и сетевые устройства, так и их совокупность. Важная особенность поведенческого анализа — это возможность выявления ранее неизвестных угроз, для которых отсутствуют IOC. На фоне роста обрабатываемых объемов данных и увеличения количества и сложности кибератак эта отличительная черта является перспективным преимуществом поведенческой аналитики.

Интерес к поведенческой аналитике обусловлен начавшемся в 2021 и 2022 годах ростом уровня внутренней угрозы для организаций, на который повлияли такие факторы, как уход сотрудников на удаленную работу, активное привлечение к работе сотрудников сторонних компаний и нестабильность рынка труда в целом. В 2021 году организации потратили в среднем 15,38 миллионов долларов на борьбу с инсайдерскими угрозами, что на 34 % больше, чем 11,45 миллионов долларов, потраченных в 2020 году [14]. В 2022 году некоторые вендоры зафиксировали наивысший уровень внутренней угрозы: до 35 % инцидентов вызваны несанкционированным доступом внутренних сотрудников [15]. На основании данного тренда, к 2026 году аналитики прогнозируют рост рынка UEBA с 1,2 миллиардов долларов в 2022 году до 4,2 миллиарда долларов США [6, с. 480]. Однако, согласно отчету «2023 Cyberthreat Defense Repor», использование поведенческой аналитики среди опрошенных организаций уступает сигнатурным методам [16]. На рисунке 1 представлена статистика использования технологий сетевой безопасности в опрошенных организациях.

Программные средства защиты информации, в которых поведенческий анализ используется в качестве самостоятельной технологии относятся к системам UEBA (User and Entity Behavioral Analytics, ранее, когда объектом наблюдения были только пользователи, UBA, User Behavioral Analytics). Наиболее популярным сценарием применения UEBA является выявление инсайдерской деятельности и утечек информации. В связи с этим системы UEBA наиболее часто относят к решениям класса IRM (Insider Risk Management) [14].

Рис. 1. Статистика использования в организациях технологии сетевой безопасности

В настоящий момент в России нет систем защиты информации, реализовавших в полной мере концепцию поведенческого анализа. Средства класса UBA, UEBA, которые использовали российские организации, были разработаны зарубежными компаниями и ушли с отечественного рынка в результате санкций. [4]. Поэтому разработка универсального метода прогнозирования поведения объектов является важным и актуальным вопросом.

Создание инструментов, основанных на поведенческом анализе и машинном обучении, представляет собой сложный процесс и требует затрат большого объема ресурсов. В данном исследовании проведен анализ научных исследований и разработок с целью определения возможности создания универсального для современной типовой информационной инфраструктуры метода прогнозирования поведения объектов.

Методика исследования

Целью данного исследования является оценка возможности создания универсального метода прогнозирования поведения объектов в современной информационной системе. Для достижения поставленной цели проведен анализ актуальных научных публикаций, посвящённых проблемам поведенческой аналитики. Рассмотрены возможности и ограничения моделей машинного обучения, лежащих в основе уже разработанных средств защиты информации. В связи с тем, что универсальное средство должно быть совместимо с различным программно-аппаратным обеспечением, особое внимание в исследовании уделялось условиям использования данных разработок. Проанализирован потенциал применения рассмотренных подходов к созданию поведен-

ческих моделей в информационных средах, содержащих компьютеры под управлением разных операционных систем семейств Windows и Unix.

В соответствии с целью сформулированы следующие результаты, которое должны быть достигнуты в рамках данного исследования:

— список основных математических теорий, которые используются при создании средств поведенческой аналитики;
— оценка возможности использования каждой теории при построении универсального метода прогнозирования поведения объектов в информационной системе;
— перечень проблем, препятствующих созданию универсального метода прогнозирования поведения объектов в информационной системе.

Список математических теорий был получен в результате анализа доступных в электронной библиотеке «eLIBRARY.RU» научных статей, опубликованных в период с 2021 по 2024 годы. Отбор публикаций выполнялся по таким ключевым словосочетаниям, как: «информационная безопасность», «анализ поведения пользователей», «система мониторинга за действиями пользователя», «ueba», «машинное обучение» и т.д. [4].

Каждая математическая теория, на примере разработанного алгоритма, была оценена с точки зрения возможности использования в различных информационных системах. Для этого были определены следующие критерии:

1. Возможность использования в различных операционных системах. Соответствие данному критерию подразумевает, что применяемый в алгоритме метод выявления угроз безопасности информации не привязан к функциональным особенностям одной операционной системы и может одинаково эффективно работать в различных информационных системах.
2. Возможность использования с различными протоколами и форматами данных. Рассматриваются ограничения для входных данных алгоритма. Соответствие данному критерию означает, что алгоритм поддерживает работу со всеми типами IOC.
3. Поддержка различных типов кибератак. Данный критерий предполагает, что использование метода анализа позволит выявить большинство современных типов кибератак.

Алгоритм, соответствующий данным критериям, можно считать универсальным, а математическую модель, лежащую в его основе, можно использовать при создании универсального метода прогнозирования поведения объектов в информационной системе. Важно отметить, что в данной статье рассматриваются только функциональные особенности методов анализа и не учитывается соответствие требованием законодательства и регулирующих стандартов. Эти вопросы требуют отдельного исследования.

Результаты исследования

В рамках анализа актуальных научных публикаций, посвященных применению поведенческого анализа в области защиты информации, определены теории математической статистики, которые наиболее часто используются для построения модулей машинного обучения. Результаты анализа представлены на рисунке 2.

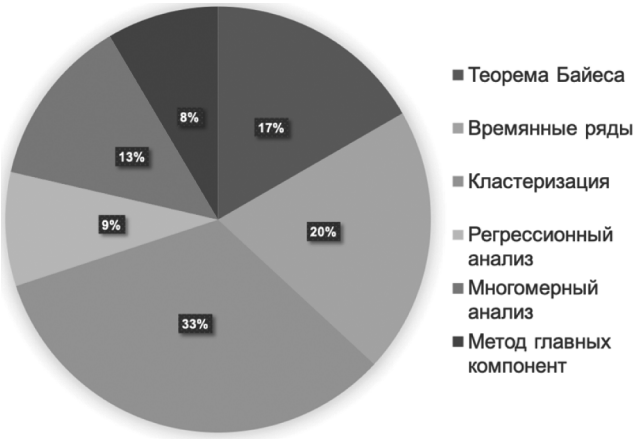

Рис. 2. Частота использования математических теорий в алгоритмах поведенческого анализа

Можно выделить следующие теории математической статистики, лежащие в основе большинства моделей поведенческого анализа в информационной безопасности:

— теорема Байеса;
— метод нечёткой кластеризации;
— анализ временных рядов.

Данные теории были рассмотрены в контексте информационной безопасности и выявления аномалий. Изучены модели поведенческого анализа, построенные на их основе. С помощью обозначенных критериев была произведена оценка возможности использования каждой теории в построении универсального метода прогнозирования поведения объектов в информационной системе.

Теорема Байеса

Теорема Байеса предназначена для подсчета вероятности наступления определенного события при происшествии другого события из множества взаимоисключающих событий [6, с. 480]. Данная теорема основана на формуле полной вероятности и является одной из основных в теории вероятностей. Если определить, что $P(A)$ — это вероятность события A, $P(B)$ — вероятность произошедшего события B, статистически связанного

с A, то, при i=1,2,...,n и j=1,2,...,n, вероятность наступления события A при наступлении события B можно определить по следующей формуле:

$$P(A_j / B) = \frac{P(A_j)P(B / A_j)}{\sum_{i=1}^{n} P(A_i)P(B / A_i)}, \quad (4)$$

где $P(B / A_j)$ — это вероятность наступления события B при наступлении события A.

Данная формула называется формулой Байеса. Ее важная особенность заключается в связи вероятности одного события с другим, уже наступившим. Именно эта связь позволяет использовать формулу Байеса при решении задач поведенческого анализа, в которых требуется предсказать сценарий развития ситуации на основании анализа уже произошедших событий.

На основании теоремы Байеса создана Байесовская сеть. Байесовская сеть — это графическая модель, которая представляет собой совокупность вершин и направленных ребер, описывающих зависимости между случайными величинами [6, с. 480]. Определение вероятностей событий в данной модели выполняется с учетом зависимости между различными случайными величинами с помощью формулы Байеса. Байесовская сеть позволяет представить зависимости в виде графа, где каждая вершина представляет собой случайную величину, а направленные ребра указывают на степень зависимости между ними.

Пример использования теоремы Байеса для решения задач кибербезопасности представлен в работе «Исследование действий пользователей в информационной среде» [7, с. 52]. Автор составил математическую модель обнаружения внутреннего нарушителя информационной безопасности, основанную на байесовской сети. Модель предназначена для анализа поведения пользователей в информационной системе под управлением операционной системы Microsoft Windows. В качестве вершин графа выбраны определенные события информационной безопасности (события ИБ), детектируемые программно-техническими средствами сбора и анализа данных мониторинга. Данные события указывают на возможное нарушение безопасности информации, в том числе реализации угрозы безопасности информации, сбое программно-технических средств, обеспечивающих функционирование информационной инфраструктуры, средств защиты информации [1]. В модели отслеживаются следующие события ИБ:

— события журналов Microsoft Windows с идентификаторами 5136–5145, указывающие на попытки и успехи получения доступа пользователя к объектам Active Directory [8];
— события подключения к компьютеру съемных носителей, детектируемые антивирусным программным обеспечением;

На основании предположения, что все отслеживаемые события ИБ являются условно независимыми, можно заключить, что реализация угрозы возможна при наступлении одного или нескольких событий ИБ одновременно. Таким образом, при присвоении каждой вершине графа значения вероятности, возможно составить байесовскую сеть, позволяющую относить события ИБ к стандартной и аномальной активности пользователя.

Результатом применения данной модели является рекомендация сотруднику информационной безопасности по принятию мер, соответствующих одному из трех возможных вердиктов на зафиксированное событие: реализованная угроза, подозрение на реализацию угрозы и ложноположительное срабатывание. Предусмотрены следующие мероприятия:

1. Блокировка учетной записи пользователя (УЗ) — при реализованной угрозе;
2. Проведение дополнительного анализа — при подозрении на реализацию угрозы;
3. Мероприятия по реагированию не требуются — при ложноположительном срабатывании.

Данная модель позволяет анализировать поведение пользователей в информационной системе под управлением Microsoft Windows и вычислять вероятность причастности зафиксированной аномальной активности к реализации угрозы ИБ.

Для определения возможности использования теоремы Байеса в разработке универсального метода прогнозирования поведения объектов в информационной системе проведем оценку данного подхода по обозначенным критериям.

Возможность использования в различных операционных системах. Несмотря на то, что сам алгоритм ориентирован на работу только в среде Windows, рассмотренный принцип выявления аномалий может использоваться и в других операционных системах. Функционал регистрации событий безопасности имеется во всех операционных системах.

Возможность использования с различными протоколами и форматами данных. Рассмотренный алгоритм работает только с событиями безопасности операционной системы и не подразумевает анализ таких типов данных, как объем передаваемого сетевого трафика, содержание сетевых пакетов и сведения о пользователе. Поэтому данный алгоритм и теорема Байеса не могут использоваться с различными протоколами и форматами данных.

Поддержка различных типов кибератак. Ввиду того, что в рассмотренном алгоритме отсутствует возможность проводить анализ данных сетевого трафи-

ка, выявление атак на соответствующем уровне также невозможно. Можно заключить, что этот алгоритм не поддерживает различные типы атак.

Таким образом, с точки зернения выделенных критериев, данный алгоритм нельзя считать универсальным. Однако, теорема Байеса является мощным инструментом для подсчета вероятностей реализации различных событий. В сочетании с другими математическими методами анализа теорема Байеса может использоваться для создания универсальной модели прогнозирования поведения объектов в информационной системе.

Метод нечёткой кластеризации

Задача выявления угроз на основании аномального поведения объектов информационной системы может быть решена через проведение кластеризации. Кластеризация — это процесс подбора признаков схожести и разбиения объектов на кластеры [9, с. 4; 10, с. 372]. В рамках решения данной задачи можно выделить, например, следующие кластеры:

— стандартная активность;
— аномальная активность.

К первому кластеру должны быть отнесены все события, соответствующие безопасному функционированию информационной системы: легитимные действия пользователя, штатная работа сервера и стандартный объем передаваемых по сети данных. Все события, которые указывают на нестандартное поведение объектов информационной системы должны быть отнесены к аномальной активности. Таким образом, отклонения от нормальной работы будут сигнализировать о возможной реализации угрозы информационной безопасности. Точные параметры стандартной и аномальной активностей неизвестны. Таким образом для каждого объекта необходимо посчитать вероятность, с которой он относится к каждому кластеру, количество которых известно и ограничено. Для этого может использоваться метод нечеткой кластеризации [9, с. 9]. Данный метод заключается в подборе оптимального разбиения объектов на группы с целью минимизации ошибки разбиения. [10, с. 373]

Для описанной выше задачи алгоритм применения метода нечеткой кластеризации можно описать следующим образом:

1. выбор двух событий ИБ по количеству кластеров и назначение их центром соответствующего кластера;
2. присвоение к определенному кластеру событий ИБ, наиболее близких по признакам к его центру;
3. пересчет центров на основании обновленного состава кластеров
4. подсчет и проверка ошибки на соответствие критерию.

Таким образом события ИБ, имеющие наиболее похожие признаки будут отнесены к стандартной активности, а остальные — к аномальной.

Метод нечеткой кластеризации широко применяется в рамках решения задач поведенческого анализа [11, с. 76; 12, с. 123]. В работе «Обнаружение аномальных запросов к базам данных на основе методов машинного обучения» с помощью кластеризации выполняется обнаружение аномальных запросов к базам данных [11, с. 77]. Автор представляет алгоритм анализа регистрационного журнала системы управления базами данных (СУБД), позволяющий с помощью методов машинного обучения выявлять признаки реализации угроз безопасности информации. Кластеризация в данном примере используется для отнесения запросов к определенному типу компьютерных атак.

Регистрационный журнал СУБД содержит сведения о всех действиях, выполняемых в СУБД. Каждая запись имеет описание, дату и время запуска каждой операции в СУБД.

Для работы алгоритма создается обучающая выборка, содержащая характерные для реализации угрозы безопасности информации записи регистрационного журнала СУБД. Она формируется на основании анализа компьютерных атак, направленных на базу данных. Данная выборка необходима для обучения классификатора — математической модели, составляющей правила отнесения записей регистрационного журнала СУБД к определенной компьютерной атаке. В основе данной модели используется метод нечёткой кластеризации. С его помощью для каждого анализируемого объекта определяется кластер, состоящий из наибольшего числа похожих экземпляров обучающей выборки (запросов к СУБД). Таким образом выполняется идентификация активности в СУБД по кластеру, который был определен с помощью метода нечёткой кластеризации.

Проведем оценку возможности использования нечёткой кластеризации в разработке универсального метода прогнозирования поведения объектов в информационной системе с помощью обозначенных критериев.

Возможность использования в различных операционных системах. Алгоритм, основанный на методе нечёткой кластеризации, с точки зрения применимости в различных информационных системах является универсальным, так как в разных операционных системах регистрационные журналы СУБД имеют идентичный набор полей. Использующиеся в анализе данные из записи журнала имеются во всех типах системах: описание, дата и время запуска операции в СУБД.

Возможность использования с различными протоколами и форматами данных. С точки зрения возмож-

ности использования с различными типами данных, рассмотренный алгоритм ориентирован только на анализ регистрационных журналов СУБД. Поддержка других протоколов и типов данных не предусмотрена в работе данной модели.

Поддержка различных типов кибератак. Часть компьютерных атак, например перехват сетевого трафика, DDoS, эксплуатация уязвимостью приложения, невозможно выявить на основании записей регистрационного журнала СУБД, так как они не влияют на работу базы данных. Поэтому данный алгоритм нельзя назвать универсальным по критерию поддержки различных типов атак.

Однако несмотря на то, что описанный выше алгоритм решает узконаправленную задачу и применим только в анализе регистрационного журнала СУБД, математическая модель, лежащая в его основе, может использоваться для более широкого набора задач. Метод нечёткой кластеризации является универсальным способом кластеризации данных и находит широкое применение в системах поведенческого анализа.

Анализ временных рядов

Широко распространенным в области машинного обучения математическим методом является анализ временных рядов. Если представить состояние одного параметра наблюдаемого объекта в определенный момент времени t, как y_t, то накопленные результаты наблюдений за период времени $t \in [1:n]$ можно представить следующим образом:

$$y_1, y_2, \ldots, y_n.$$

Данная последовательность называется временным рядом. Временной ряд — это расположенные в хронологическом порядке значения того или иного показателя, изменение которого отражает ход развития изучаемого явления [12, с. 123]. С помощью временных рядов возможно осуществлять анализ данных с учетом времени, в которое они были записаны. Благодаря этому данный метод также находит широкое применение в поведенческом анализе.

В работе «Модель и метод оценивания защищённости киберфизических систем от информационных угроз на основе анализа временных рядов» представлен алгоритм классификации состояний объектов замкнутой информационной системы (ЗИС), основанный на анализе временных рядов [13, с. 133]. Он используется для выявления нарушений ИБ при идентификации соответствующих состояний ЗИС.

Каждый процесс, протекающий в ЗИС, характеризуется набором параметров объектов ЗИС, изменяющихся с течением времени. Данные параметры собираются и накапливаются в базу данных средствами мониторинга объектов ЗИС. Для конечного набора процессов, протекающих в ЗИС, определено множество временных рядов X, отражающих состояние ИБ объектов ЗИС:

$$\{o_1, \ldots, o_m\} = \\ = x_1(t_1), x_2(t_1), \ldots, x_n(t_1)\}, \ldots, \{x_1(t_m), x_2(t_m), \ldots, x_n(t_m)\}\} \subset X,$$

где m — количество объектов ЗИС.

На основании обучающей выборки из l идентифицируемых состояний ИБ КФС выделены два класса состояний ИБ ЗИС: $\{c_1, c_2, \ldots, c_k\} \subset C_0$ и $\{c_{k+1}, c_{k+2}, \ldots, c_l\} \subset C_1$ — множество безопасных состояний и множество аномальных и потенциально опасных состояний объектов ЗИС соответственно.

Таким образом, для определенного времени t работу разработанного алгоритма μ можно представить следующим образом:

$$c(t) = \mu(x_{1,t}, x_{2,t}, \ldots, x_{s,t}), c \in C, x_{i,t} \in D_f, s \ll n,$$

где: D_f — множество допустимых значений признака, s — количество анализируемых признаков [13, с. 133].

Результатом работы алгоритма является идентификация небезопасного состояния объекта ЗИС $c(t) \subset C_1$, указывающего на реализацию угрозы безопасности информации ЗИС.

Рассмотрим теорию временных рядов с точки зрения использования в разработке универсального метода прогнозирования поведения объектов в информационной системе, исходя из обозначенных критериев.

Возможность использования в различных операционных системах. Описанный выше подход позволяет выявлять аномальное поведение объектов замкнутой информационной системы, в которой протекает известный и ограниченный набор процессов. Поэтому данный алгоритм не может использоваться в динамических системах, которыми являются большинство современных информационных инфраструктур. Однако, временные метки событий регистрационных журналов имеются во всех операционных системах. Благодаря этому теория временных рядов может использоваться в любых операционных системах.

Возможность использования с различными протоколами и форматами данных. В рассмотренном алгоритме нет зависимости от определенного типа и источника данных, так как временные ряды возможно применять к неограниченному набору параметров объектов. По-

этому алгоритмы, основанные на анализе временных рядов, можно считать универсальными по критерию поддержки различных протоколов и форматов данных.

Поддержка различных типов кибератак. Несмотря на то, что рассмотренный подход к использованию теории временных рядов не привязан к определенному типу данных, выявление некоторых кибератак требует более глубокой аналитики. Например, для обнаружения нелегитимной выгрузки данных внутренним нарушителем необходимо вычислить объем переданного трафика. Распределение на временной шкале различных событий безопасности является важной составляющей мониторинга, но должно использоваться в совокупности с другими методами анализа. Поэтому теорию временных рядов нельзя считать достаточной для выявления всех типов кибератак.

Выводы

На основании результатов проведенных исследований текущее состояние разработок в области поведенческого анализа можно охарактеризовать следующими образом:

1. Системы поведенческого анализа в настоящий момент используются для решения узконаправленных задач;
2. Системы поведенческого анализа используются только в информационных системах со строго ограниченным составом программного и аппаратного обеспечения;
3. В настоящий момент отсутствует единый подход к построению аналитических модулей. В основе моделей машинного обучения используются различные теории и методы математической статистики.

Таким образом, можно выделить следующие проблемы, препятствующие созданию универсального метода прогнозирования поведения объектов в информационной системе:

1. Недостаток данных. Универсальный метод прогнозирования поведения объектов должен быть основан на данных, которые имеют одинаковое содержание в различных сценариях работы и не зависят от состава программно-аппаратного обеспечения.
2. Разнообразие информационных систем. В современных информационных инфраструктурах используется множество различных информационных систем, каждая из которых имеет свои особенности и специфические данные. Это может затруднить создание универсального метода прогнозирования поведения объектов. Необходимо провести комплексный анализ каждой из этих систем для выбора наиболее подходящего для этой задачи состава данных.
3. Сложность математических теорий. Создание универсального метода прогнозирования поведения объектов в информационной системе должно быть основано на совокупности математических теорий. Для эффективного использования различных теорий совместно требуется углубленный анализ в области математики и статистики.

Обсуждение и заключение

В результате исследований определены основные математические методы и подходы к разработке, используемые при создании систем поведенческой аналитики. Составлен следующий список математических теорий, лежащих в основе большинства алгоритмов поведенческого анализа в информационной безопасности:

— теорема Байеса;
— метод нечёткой кластеризации;
— анализ временных рядов.

На основании обозначенных критериев выполнена оценка возможности использования каждой математической теории при построении универсального метода прогнозирования поведения объектов в информационной системе. Результаты оценки представлены в таблице 1.

Таблица 1.

Соответствие математических моделей критериям универсальности

	Теорема Байеса	Метод нечёткой кластеризации	Анализ временных рядов
Возможность использования в различных операционных системах	+	+	+
Возможность использования с различными протоколами и форматами данных.	–	–	+
Поддержка различных типов кибератак	–	–	–

Сформулированы актуальные проблемы, препятствующие созданию универсального метода прогнозирования поведения объектов. Полученные результаты являются основной для дальнейших исследований в данной области. Последующие работы должны быть сосредоточены вокруг решения следующих проблем:

— определение требований к универсальной модели прогнозирования поведения объектов в информационной системе;
— выбор достаточного для поведенческого анализа состава данных, который не зависит от перечня используемого в информационной системе программного обеспечения;

— разработка комплексного алгоритма прогнозирования поведения объектов, основанного на совокупности математических теорий.

Таким образом, на основании анализа научных публикаций, посвященных проблемам поведенческой аналитики, а также текущих разработок в данной области, произведена оценка возможности создания универсального метода прогнозирования поведения объектов в информационной инфраструктуре.

ЛИТЕРАТУРА

1. ГОСТ Р 59547-2021 Национальный стандарт Российской Федерации. Защита информации. Мониторинг информационной безопасности. Общие положения — Введ. 2022-04-01. М.: Стандартинформ, 2022
2. C. Chio, D. Freeman Machine Learning & Security PROTECTING SYSTEMS WITH DATA AND ALGORITHMS. — First Release изд. — Sebastopol: O'Reilly Media, Inc., 2018. — (Chapter 2, Papper 25–29, 32–34, 35–37).
3. UNDERSTANDING INDICATORS OF COMPROMISE (IOC) PART I // RSA URL: https://web.archive.org/web/20170914034202/https://blogs.rsa.com/understanding-indicators-of-compromise-ioc-part-i/ (дата обращения: 17.11.2023).
4. UEBA-системы: что это, принципы работы, обзор рынка // Инсайдер.рф URL: https://инсайдер.рф/news/ueba_sistemy/ (дата обращения: 31.01.2024).
5. Научная электронная библиотека eLIBRARY.RU // eLIBRARY.RU URL: https://elibrary.ru (дата обращения: 12.11.2023).
6. Гмурман В.Е. Теория вероятностей и математическая статистика. — 12 изд. — М.: Юрайт, 2016. — 480 с.
7. Емелина, А.А. Исследование действий пользователей в информационной среде / А.А. Емелина, Н.Е. Карпова, А.А. Саранский // Интерэкспо Гео-Сибирь. — 2023. — Т. 6. — С. 50–57.
8. Doug White. Three Key Areas in Active Directory Security/ Security Weekly. URL: https://securityweekly.com/2018/09/06/three-key-areas-in-active-directory-security/ (дата обращения 02.11.2023).
9. Кудинов Ю.И., Кудинов И.Ю. Нечеткое моделирование и кластеризации // Пробл. управл., — 2008, выпуск 6, 2–10.
10. Мохорев, Д.Е. Применение метода нечеткой кластеризации для ранжирования индикаторов компрометации по уровню потенциальной угрозы / Д.Е. Мохорев // Инжиниринг предприятий и управление знаниями (ИП&УЗ-2021) : Сборник научных трудов XXIV Международной научной конференции, Москва, 02–03 декабря 2021 года / Под научной редакцией Ю.Ф. Тельнова. — Москва: Российский экономический университет имени Г.В. Плеханова, 2022. — С. 371–375.
11. Саенко, И.Б. Обнаружение аномальных запросов к базам данных на основе методов машинного обучения / И.Б. Саенко, М.Х. Аль-Барри // Труды ЦНИИС. Санкт-Петербургский филиал. — 2021. — Т. 2, № 12. — С. 75–79.
12. Куприенко Н.В., Пономарева О.А., Тихонов Д.В. Статистика. Временные ряды. Анализ тенденций и прогнозирование. — 1 изд. — СПБ: Издательство Политехнического университета, 2015. — 123 с.
13. Семенов, В.В. Модель и метод оценивания защищённости киберфизических систем от информационных угроз на основе анализа временных рядов: специальность 23.60.00: диссертация на соискание ученой степени кандидата технических наук / Семенов Виктор Викторович, 2022. — 133 с.
14. 2023 Market Guide for Insider Risk Management Published 13 November 2023 // Gartner URL: https://www.gartner.com/en/documents/4931631 (дата обращения: 14.11.2023).
15. Q3 2022 Threat Landscape: Insider Threat, The Trojan Horse of 2022
16. 2023 Cyberthreat Defense Report // CyberEDGE URL: https://cyber-edge.com/cdr/ (дата обращения: 14.10.2023).

Журнал «Современная наука: актуальные проблемы теории и практики»