ПРОАКТИВНЫЙ ПОИСК ВНУТРЕННИХ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В УСЛОВИЯХ ОГРАНИЧЕНИЙ

Исхаков А.Ю.[1], Гайдук К.А.[2]

Цель работы: исследование методов обнаружения внутренних угроз информационной безопасности и повышение их эффективности за счет модернизации подхода к построению системы выявления внутренних угроз.

Метод исследования: системный анализ открытых источников данных о методах обнаружения внутренних угроз информационной безопасности; построение модели на основе применения методов машинного обучения; методы формирования выборок и методы обучения, оценка аномальности на основе методов принятия решения.

Полученный результат: В данной работе была сформулирована задача выявления внутренних угроз, которая в рамках моделей машинного обучения формулируется как задача выявления аномалий. В статье проведено исследование существующих методов и моделей обнаружения внутренних угроз, основанных на неконтролируемом машинном обучении, в том числе на многослойных искусственных нейронных сетях (ИНС), приведен и формализован подход к построению системы выявления внутренних угроз, который затем применен в существующем наборе данных для оценки эффективности моделей. Особенность предложенного подхода заключается в возможности обнаружения внутренних угроз как для каждой записи отдельно, так и для каждого пользователя на основании числа аномальных записей. Описанные методы извлечения, агрегации и представления данных, сформулированные с учетом ограничения реальных систем были применены на наборе данных о действиях пользователей за сутки. Для подходов принятия решения об аномальности записей и каждого пользователя были посчитаны оценки эффективности моделей. Полученные оценки могут быть интерпретированы как следствие обобщающих способностей моделей и особенностями подсчета значений аномальности.

Научная новизна: Статья представляет подход обнаружения внутренних угроз информационной безопасности, отличающийся возможностью обнаружения внутренних угроз как для каждой записи отдельно, так и для каждого пользователя на основании числа аномальных записей.

Ключевые слова: внутренние угрозы, внутренний нарушитель, выявление угроз, информационная безопасность, аномалии, методы машинного обучения, искусственные нейронные сети.

DOI:10.21681/2311-3456-2023-4-105-119

Введение

Внутренние угрозы можно определить как любые потенциально опасные для организации действия субъекта (инсайдера), к ресурсам которой у него есть авторизованный доступ. К их числу относится: несанкционированная передача данных; нарушение целостности ресурсов и другие злонамеренные и непреднамеренные действия субъектов доступа. Задача выявления таких угроз усложняется тем, что лица способны тщательно скрывать следы, пытаясь моделировать нормальное поведение. Задача выявления таких угроз предполагает анализ большой объем данных системных журналов, зачастую в режиме реального времени. Внутренние угрозы можно разделить на две категории [1]: умышленные (злонамеренные), которые исходят от лиц, которые осознанно используют свой доступ к ресурсам компании для модификации, уничтожения и сбора конфиденциальной информации в личных целях и непреднамеренные (пассивные), когда сотрудники неосознанно предоставляют доступ злоумышленнику, игнорируя политику безопасности или пользуются ее недостатками.

1 Исхаков Андрей Юнусович, кандидат технических наук, старший научный сотрудник Института цифровых технологий Финансового университета, Москва, Россия. E-mail: iskhakovandrey@gmail.com, ORCID: 0000-0002-6603-265X

2 Гайдук Кирилл Алексеевич, магистрант НИЯУ МИФИ, Москва, Россия. E-mail: guydukk@gmail.com, ORCID: 0009-0002-3276-0891

В подходах к обнаружению внутренних угроз можно выделить три направления: подходы на основе правил, на основе графов и на основе методов машинного обучения [2]. Подходы, основанные на применении методов машинного обучения, получили широкое применение благодаря своему мощному математическому аппарату и высокой эффективности.

В статье проведено исследование существующих методов и моделей обнаружения внутренних угроз, основанных на неконтролируемом машинном обучении, в том числе на многослойных искусственных нейронных сетях (ИНС), приведен и формализован подход к построению системы выявления внутренних угроз, который затем применен существующем наборе данных для оценки эффективности моделей. Особенность предложенного подхода заключается в возможности обнаружения внутренних угроз как для каждой записи отдельно, так и для каждого пользователя на основании числа аномальных записей.

1. Выявление внутренних угроз

В настоящее время вместе со стремительным развитием методов анализа данных и современных возможностей вычислительной техники, наблюдается активное совершенствования средств защиты информации в рамках выявления внутренних угроз. По сравнению с традиционными методами анализа и последующего изменения политики безопасности, выявление угроз на основе данных дает преимущество непрерывной и своевременной оценки. Такой подход позволяет избежать различных проблем, присущих традиционным подходам, таких как их значительные административные накладные расходы и их зависимость от постоянной подстройки.

В контексте решения задачи выявления внутренних угроз машинное обучение используется для создания и применения моделей, идентифицирующих угрозы и статистически определяющих аномальное поведение. Классификация методов решения задачи выявления внутренних не может быть однозначной, так как количество этих методов постоянно растет и повторно анализируется в совокупности, однако с развитием таких методов наблюдаются определенные тенденции, подробно описанные в [3]. Подход на основе машинного обучения позволяет обнаруживать внутренние угрозы, анализируя данные журналов системы, отображающие действия пользователей, с достаточно высокой точностью без явного задания правил.

В контексте машинного обучения задача выявления внутренних угроз формулируется как задача детектирования аномалий, которые в зависимости от методов способны выявить различные типы атак [4].

Методы «глубокого» обучения как метод машинного обучения возникли на основе исследований искусственных нейронных сетей (ИНС) [5], которые, в свою очередь, появились в результате применения математического аппарата к исследованию функционирования нервной системы живых существ. Полученные при этом результаты успешно применяются при решении проблем классификации, выявления аномалий, прогнозирования, оптимизации и выявления внутренних угроз [6]. Таким образом модели глубокого обучение представляет собой любые многослойные модели, основанные на искусственных нейронных сетях.

Система выявления внутренних угроз предполагает обработку большого объема информации, которая представляет собой системные и сетевые журналы, отражающие действия пользователей информационной системы. Методы машинного обучения получают на вход преобразованные в вектор признаков данные и обучаются на них, чтобы затем вычислить значение аномальности каждого экземпляра данных и принять решение является ли такой вектор признаков аномальным. Принятие решения об аномальности экземпляра данных состоит в выборе порогового значения, выше или ниже которого все экземпляры будут считаться аномальными. Естественным образом можно выделить подход для принятия решения об аномальности пользователей на основе числа аномальных экземпляров данных у каждого пользователя. Таким образом, задача выявления аномалий ставит ряд вопросов о выборе парадигмы обучения, конкретного метода, вопросы обработки и представления данных, выбора порогового значения аномальности и подхода к принятию решения.

В машинном обучении существует две парадигмы: контролируемые (с учителем) и неконтролируемые методы (без учителя), которые отличаются способом представления обучающих данных. В контролируемых методах обучение происходит на заранее размеченных целевых данных. Разметка осуществляется экспертами в данной области и позволяет однозначно определить — несут ли конкретные действия внутреннюю угрозу или нет. Неконтролируемые методы применяются в условиях отсутствия возможности разметки данных. Обучение, как правило, требует более длительного времени и позволяет добиться меньшей точности за счет выявления любых аномальных действий, в том числе не представляющих угрозу информационной безопасности.

В реальных условиях система выявления аномалий работает с неразмеченными данными, для кото-

рых заранее неизвестно, являются ли данные аномальными или нет. По этой причине в данной работе рассматриваются только неконтролируемые алгоритмы выявления аномалий.

2. Неконтролируемые модели машинного обучения

В случае обучения без учителя алгоритм выявляет скрытые структуры данных, обнаруживает группы схожих объектов или взаимосвязанных свойств. Такой метод не требует заранее размеченных данных, что может значительно сэкономить трудозатраты. Так как эти методы не требуют знание меток, которое можно получить с помощью экспертов, они способны работать в режиме реального времени и осуществлять поиск как новых типов угроз, так и уже известных.

2.1 Одноклассовый метод опорных векторов

Метод опорных векторов (SVM) был разработан как алгоритм для классификации двух или более классов, который ищет максимальную границу разделения классов и является контролируемым алгоритмом [7]. Основная идея алгоритма состоит в использовании функции ядра, которая способна преобразовывать пространство признаков без их непосредственного преобразования, которая позволяет получать бесконечно размерные пространства признаков. В свою очередь, одноклассовый SVM может иметь лишь один класс и применяется для классификации выбросов, когда данные содержат только положительные точки [8]. Общая идея данного метода состоит в преобразовании пространства признаков и разделении гиперплоскостью так, чтобы наблюдения были наиболее удалены от начала координат. Полученная граница разделяет близко расположенные наблюдения из выборки и аномальные значения. Таким образом, данный алгоритм должен обучаться в большей мере на «нормальных» неразмеченных данных.

Одноклассовый SVM пытается построить наименьшую гиперсферу (в общем случае гиперплоскость), включающую все точки выборки x_i, поэтому проблема оптимизации определяется следующим образом (1):

$$r^2 \rightarrow \min \|F(x_i) - a\|2 \leq r2, \quad (1)$$

где i =1,...,N, N – размер выборки, r > 0, a — центр гиперсферы, F – функция ядра

Одноклассовый SVM дает возможность выбора различных функций ядра, которые позволяют использовать нелинейные границы принятия решений.

Мерой аномальности для одноклассового SVM является расстояние до разделяющей гиперсферы, причем расстояние сильно зависит от данных и количества признаков, поэтому для принятия решения об аномальности данных необходимо иметь предположение о верхней границе процента аномальности данных для вычисления порога аномальности. Таким образом алгоритм одноклассового SVM обладает возможностью обнаружения новизны и стремится определить, являются ли новые значения аномалией или нет.

2.2 Изоляционный лес (IF)

Алгоритмы «деревьев решений» позволяют прогнозировать поведение модели с высокой точностью и достаточно простой интерпретацией. Методы деревьев решений строят модель решений, принятых на основе значений атрибутов, а решения представляют собой древовидные структуры. Этот класс алгоритмов получил свою популярность ввиду достаточно большой скорости и точности при решении задач, решаемых с помощью машинного обучения [9].

Применительно к задаче выявления аномалий применяется алгоритм «Изоляционный лес» (Isolation forest, IF) [10]. Основываясь на предположении редкости аномалий и их отличия по значениям атрибутов от нормальных точек данных, IF разработан как ансамбль «деревьев изоляции», при этом предполагается, что аномалии, которые легче изолировать, находятся ближе к корням деревьев, чем нормальные экземпляры. Изоляционный лес отличается от других методов обнаружения аномалий, которые, в основном, строят модели на нормальных данных и определяют аномалии как любые экземпляры, не соответствующие модели. Каждое дерево в IF работает на подмножестве обучающих данных и наборе признаков, которые выбираются случайным образом. В каждом узле дерева генерируются двоичные разбиения по случайно выбранному признаку и значению разбиения. Процесс рекурсивно повторяется до тех пор, пока каждый экземпляр не окажется в листе дерева, то есть будет изолирован. После обучения всех деревьев изоляции балл аномалии экземпляра данных рассчитывается как средняя длина пути от корневых узлов до соответствующих листьев экземпляра в деревьях [11].

Значение аномальности для каждого образца подсчитывается по формуле следующим образом (2):

$$A(x,n) = 2^{\frac{E(h(x))}{c(n)}} \quad (2)$$

где n – число экземпляров, h(x) – глубина в конкретном дереве, E(h(x)) – средняя длина по деревьям,

$c(n) = 2H(n-1) - \frac{2(n-1)}{n}$, H(n) – n-ое гармоническое число.

В некоторых реализациях алгоритма формула для вычисления значения аномальности экземпляра данных берется с противоположным знаком. При правильном подборе количества деревьев и достаточно большой выборке данный алгоритм способен выявлять аномалии намного быстрее. Алгоритм изоляционного леса также имеет модификацию – расширенный изоляционный лес, который позволяет избежать некоторых недостатков, присущих модели [12].

3. Глубокие модели обучения без учителя

Многослойные ИНС, которые еще часто называют «глубокими» моделями, также разделяются на контролируемые и неконтролируемые модели. Основная идея неконтролируемых моделей заключается в обучении сети реконструкции входных данных на выходе.

3.1 Сеть автокодировщика

Автокодировщик представляет собой ИНС, которая призвана повторять на выходе сети сигнал, наиболее близкий к входному. Главная идея этой архитектуры, как и в множестве других сетей, состоит в обратном распространении ошибки от выходного слоя к входному [13].

Автокодировщик состоит из входного, скрытых и выходного слоев, причем количество нейронов входного и выходного слоя должны быть равны. На количество нейронов в скрытом слое также накладываются ограничения в зависимости от конкретного типа автокодировщика и поставленных задач. Автокодировщик состоит из двух частей: кодировщик (кодер), преобразующий входной сигнал в представление на скрытом слое (скрытое представление) и декодировщик (декодер), который реализует обратную операцию – восстановление сигнала из скрытого представления.

Работу кодировщика можно представить как обучение некоторой функции, аргументом которой является вектор размерности n, а значением этой функции является вектор размера m (m<n). Декодировщик же в свою очередь вычисляет обратную функцию получая значение исходного вектора.

В контексте детектирования аномалий автокодировщик применяется в два этапа: для начала сеть автокодировщика обучается на случайной выборке данных, а поиск аномалий происходит на обученной сети [14]. Желательно, чтобы обучающие данные были нормальными, хотя из предположения о редкости аномальных образцов это не обязательно, но с таким ограничением на обучающую выборку сеть точно будет обучена на повторение только нормальных образцов на выходе. Обучение сетей автокодировщика происходит с помощью обратного распространения ошибок, поэтому применяются градиентные методы, а также их модификации [15]. После обучения для каждого образца данных вычисляется отличие входного сигнала от выходного, в качестве которого выбирают среднюю квадратичную ошибку или перекрестную энтропию. Для аномальных образцов эта ошибка будет больше, чем для нормальных и, исходя из знаний об обучающей выборке, остается выбрать порог ошибки реконструкции, при котором образец будет считаться аномальным. Этот порог можно выбирать как статически, например, считая квантиль значений аномальности исходя из требований к системе, или экспериментально, имея часть тестовых размеченных данных и вычисляя оценки точности классификации для каждого порога.

3.2 Сеть долгой краткосрочной памяти (LSTM)

Сети долгой краткосрочной памяти (Long Short Term Memory, LSTM) – особый вид рекуррентных нейронных сетей, основным преимуществом которых является способность к обучению долгосрочным зависимостям [16].

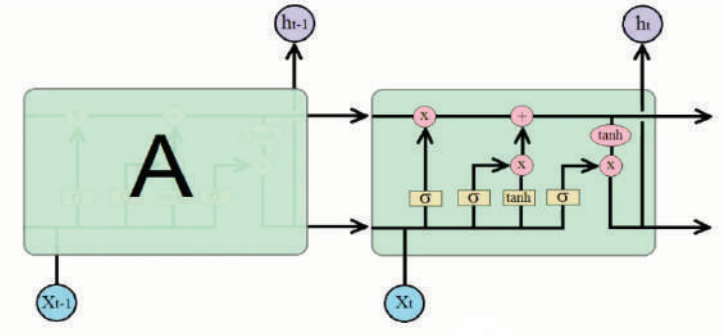


Рис. 1. Архитектура LSTM

Основная идея, благодаря которой рекуррентные ИНС получили популярность, заключается в возможности использования полученной в прошлом информации. Классические РНС умеют сохранять контекст лишь в небольшом временном промежутке, что подтверждается небольшими оценками при практическом применении [17]. Классические РНС представляют собой повторяющиеся модули с простой структурой, например, всего один слой с сигмоидной функцией активации.

Архитектура LSTM, в свою очередь, состоит из последовательно соединенных элементов, называемыми ячейками LSTM, каждая из которых состоит из четырех слоев, в которых они взаимодействуют особым образом. Архитектура сети LTSM с двумя ячейками представлена на рисунке 1.

Алгоритм обучения LTSM основан на методе обратного распространения ошибки во времени и его модификациях. Для выявления аномалий сеть LSTM обучают на подвыборке данных, структурированной по времени. После обучения в качестве входных данных подаются новые образцы и вычисляется разность между предсказанным значением сети и действительным, по аналогии выявления аномалий с помощью автокодировщиков. Большая разность реального значения и предсказанного является признаком аномальности и исходя из порога и этой величины делается предположение об аномальности.

Слои сети LSTM возможно объединить в структуру, напоминающую автокодировщик, то есть образующую кодер и декодер с помощью больших и меньших размерностей слоев. Полученную при этом архитектуру называют LSTM автокодировщиком (LSTM-AE), которая способна объединять преимущества обоих архитектур и может быть применима в контексте выявления аномалий [18].

4. Подход к построению системы выявления внутренних угроз

Рассмотренные модели машинного обучения не являются самостоятельными в контексте решения конкретных задач. В рамках задачи выявления внутренних угроз очень трудно представить единственное решение или создать универсальную модель машинного обучения, которая решала бы задачу выявления внутренних угроз однозначно [19].

Задача выявления внутренних угроз предполагает обработку журналов системы, которые несут в себе временную информацию с конкретными действиями пользователей. Система выявления внутренних угроз на основе моделей машинного обучения может быть предназначена для работы в автоматическом режиме в течение продолжительного времени, когда модель периодически обучается на основе новых данных, так и в случае однократного обучения на большом наборе данных с целью апостериорного анализа или расследования случившихся инцидентов безопасности.

К числу важнейших задач при использовании методов машинного обучения относятся процессы сбора, обработки и анализа данных, от которых в большей степени зависит качество модели машинного обучения, а впоследствии и эффективность выявления внутренних угроз. На первом этапе построения системы необходимо собрать набор данных, который представляет собой журналы активности пользователей, собранные из различных источников.

В исследованиях по данной тематике одной из главных проблем является получение реального набора данных, который без доступа к реальной корпоративной сети получить затруднительно из-за вопросов конфиденциальности и безопасности персональных данных. Кроме того, актуальным вопросом остается выбор источников данных, таких как данные сетевого трафика, данные запросов к системе (внутренние ресурсы), приложениям (выполняемые команды), внешним устройствам и т.д.

В настоящее время не существует общедоступного набора реальных данных для обнаружения внутренних угроз, поэтому исследователи в своих работах используют синтетически сгенерированные данные. Большинство наборов представляют собой данные в формате CSV, строки которых представляют различные типы действий, обычно это дата, время, информация о пользователе или устройстве и непосредственно действие в системе. Так как проблема выявления внутренних угроз остается актуальной, существует множество современных синтетических наборов данных, которые отличаются по типам источников и широко используются в современных исследованиях. Одним из таких и наиболее популярным набором данных является CERT [20], созданный Институтом программной инженерии Университета Карнеги-Меллона. Этот «свободный от ограничений конфиденциальности» набор данных разработан, чтобы позволить исследователям, изучающим тему внутренних угроз, экспериментировать и оценивать предлагаемые ими решения. Этот набор имеет несколько различных версий, данные в каждой версии генерируются с использованием различных сценариев. Каждый набор данных содержит журналы данных входа в систему, HTTP или историю просмотров

веб-сайтов, данные об обмене электронной почты, журналы доступа к файлам, использования устройств, данные аутентификации, а также дополнительную поведенческую информацию о пользователях.

В качестве источника данных для разработки и апробирования подхода к построению системы выявления внутренних угроз служит набор данных CERT версии R4.2, который содержит около 32 миллионов записей действий 1000 пользователей за период 16 месяцев. Данный набор данных содержит метки аномальности по каждому из сценариев для каждой записи. Набор данных поставляется в виде csv-файлов по каждому типу действий (email, file, http, device, logon).

Извлечение признаков должно проводиться с учетом ограничений реальных систем, когда у системы не может быть доступа к текстуальным данным действий пользователей из-за вопросов конфиденциальности, таких как данные о тексте электронных писем, содержимого контента посещаемых сайтов и т.д. Также стоит подчеркнуть, что анализ и преобразование текстуальных данных требует достаточных вычислительных ресурсов. Таким образом в предложенном подходе из текстовых данных извлекается только количественная и статистическая информация.

В научных публикациях, например в [21], сравниваются типы агрегации пользовательских данных. Этот вопрос также является важным, т.к. зачастую данные необходимо разделять на равные промежутки для эффективного представления в модели машинного обучения и хранения. В связи с этим исходные данные были агрегированы по дням для каждого пользователя. Таким образом были извлечены основные порядковые, количественные и категориальные признаки представлены в таблице 1. К извлеченным признакам были добавлены категориальные признаки, относящиеся к пользователям, такие как метка администратора, департамента, юнита и т.д. Поскольку алгоритмы машинного обучения работают с числовыми данными необходимо кодирование признаков для создания числовых векторов. Ошибки при кодировании переменных функций может привести к тому, что модели машинного обучения неправильно интерпретируют корреляцию между ними.

После кодирования, приведения к единому виду и агрегации данных переходят к процессу нормализации данных, к задачам которой можно отнести сокращение или увеличение объема данных, удаление повторных, противоречивых или некорректных данных. К алгоритмам нормализации можно отнести метод синтетической переборки меньшинства (SMOTE [22]), который применяется при несбалансированном наборе классов. Обратным к генерации действием является удаление данных для сокращения будущей обучающей выборки, которая может производиться как намеренное приближение к заданному соотношению нормальных и аномальных записей, так и случайной

Таблица 1

Извлеченные из набора данных CERT R4.2 признаки

Временной признак	Признак рабочей станции	Тип действия	Количественные признаки	Категориальные признаки
В рабочий день В нерабочий день	ПК пользователя Общий ПК ПК менеджера Другой ПК	HTTP	Количество посещенных сайтов, длина URL, глубина URL, длина контента. количество слов в контенте	Сайт по поиску работы, сайт по работе с файловым облаком, сайты с программами, другие сайты
		EMAIL	Количество получателей, количество получателей скрытых копий, количество вложений, размер сообщения, длина текста в символах, количество слов в сообщении	Получатель с внутренним доменом, получатель с внешнем доменом
		FILE	Количество обращений к файлу	Архивы, фото, документы, текстовые данные, приложения
		DEVICE	Длительность подключения к ПК, количество подключений	-
		LOGON	Число актов аутентификации	-

подвыборкой исходных данных, исходя из предположения о редкости аномальных записей. Так как мы подходим к построению системы с учетом реальных ограничений справедливо будет извлечь случайную подвыборку данных, так как зачастую мы имеем дело с неразмеченными образцами.

В соответствии с вышеописанным был получен кодированный набор извлеченных данных, который содержит 308 тысяч записей о пользовательских днях в виде 509 признаков.

В зависимости от конкретной модели машинного обучения данные необходимо масштабировать. Классические модели машинного обучения, такие как изоляционный лес и одноклассовый метод опорных векторов, не требуют масштабирования, в то время как модели, построенные на многослойных ИНС, требуют масштабирования признаков в диапазон от нуля до единицы.

В данном случае популярным решением является минимаксное преобразование, которое может быть выражено следующим образом (3):

$$x_{i,scaled} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad , \qquad (3)$$

где $x_{(i,scaled)}$ — масштабированное значение переменной, x_{min} — минимальное значение переменной, x_{max} — максимальное значение переменной.

Описанные выше этапы могут быть обобщены и формализованы в виде схемы на рисунке 2.

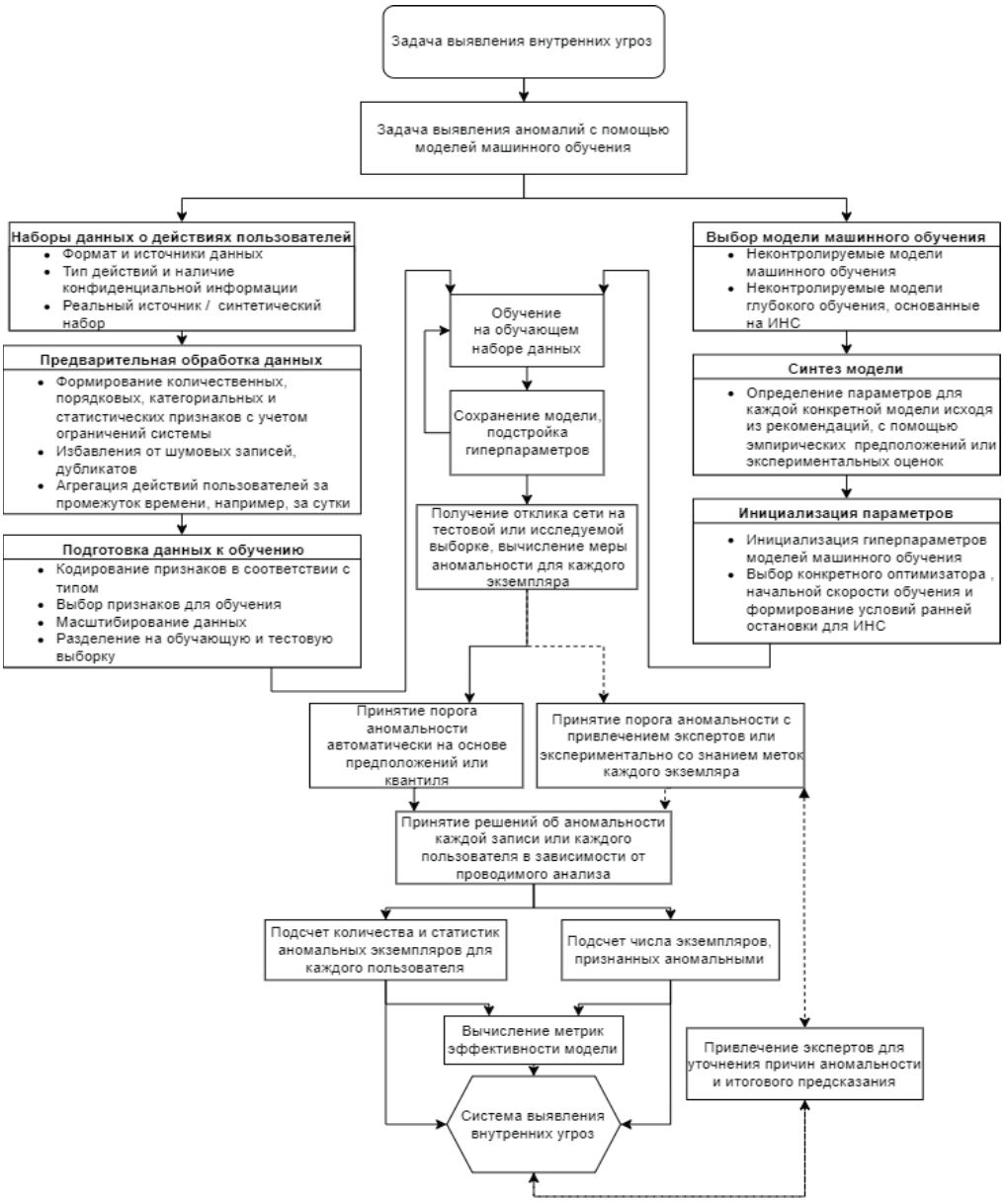


Рис. 2. Подход к построению системы выявления внутренних угроз

Для моделей машинного обучения важным аспектом являются извлеченные признаки и их количество. Для дальнейшего сравнения обобщающей способности из исходного набора было извлечено 100 признаков на основе их важности. «Важность» каждого признака можно считать, как корреляцию с прогнозируемой переменной или специальными методами, такими как метод среднего уменьшения по методу примесей [23] (Mean decrease impurity, MDI) или популярного в последнее время метода SHAP (SHapley Additive exPlanations), который основан на теории игр и ценности Шепли [24]. В данной работе был применен метод SHAP, в результате применения которого был получен ранжированный список по важности каждого признака. Стоит заметить, что из полученных результатов явно выделяются только 4 признака: общее количество действий типа HTTP, число действий типа HTTP за пользовательским ПК, число действий типа DEVICE и число всех действий в рабочее время, остальные признаки незначительно отличаются по вычисленной с помощью метода SHAP «важности».

5. Результаты обучения и предсказания моделей

Следующим действием является обучение алгоритмов и получение предсказаний от моделей. Для обучения моделей 75% исходных данных были взяты в качестве обучающих, а остальные 25% в качестве тестовых, для которых в дальнейшем строились оценки.

После получения предсказания стоит вопрос о принятии решения об аномальности. Классический подход на основе образцов заключается в признании аномальными всех записей, у которых значения аномальностей выше или ниже порогового. Описанный нами подход предполагает также принятие решение об аномальности каждого пользователя, для этого для каждого пользователей считается число записей, отнесенных к аномальным. Далее считается среднее число таких записей по всем пользователям, и, если у пользователя число аномальных записей больше, чем среднее – он признается аномальным. Сочетание этих подходов позволяет увеличить эффективность моделей без привлечения экспертной оценки. Поскольку нам важно обнаружить максимально число внутренних угроз среди всех важной оценкой для нас будет являться полнота, которая считается как отношение аномальных экземпляров (пользователей), выявленных верно, к общему числу аномальных экземпляров (пользователей). Также важной оценкой является общая точность модели, которая считается как число верных предсказаний к общему числу предсказаний.

В реализации, которая использовалась авторами, значение аномальности в формуле (2) берется со знаком минус, то есть более аномальные образцы имеют отрицательное значение аномальности. Пример полученного предсказания представлен в виде гистограммы на рисунке 3.а. В качестве автоматического порога для принятия решения об аномальности экземпляров был выбран ноль.

Обучение модели одноклассового SVM было проведено с радиально базисной функцией ядра, коэффициент ядра вычисляется автоматически как величина, обратная числу признаков. Пример полученного предсказания представлен в виде гистограммы на рисунке 4.б. В качестве автоматического порога для принятия решения об аномальности экземпляров был выбран десятипроцентный квантиль значений аномальностей. Как видно из рисунка 4.б автоматически вычисленный порог и порог с лучшими оценками практически совпадают.

Модели, основанные на ИНС, обучались с помощью оптимизатора Nadam, начальная скорость обучения 0.01, ранняя остановка с минимальной дельтой ошибок 0.001. Для модели автокодировщика ошибка считалась по формуле среднеквадратичной ошибки, а в случае с моделью LSTM ошибка считалась как средняя абсолютная ошибка. Результаты предсказания моделей представлены на рисунках 4а и 4б.

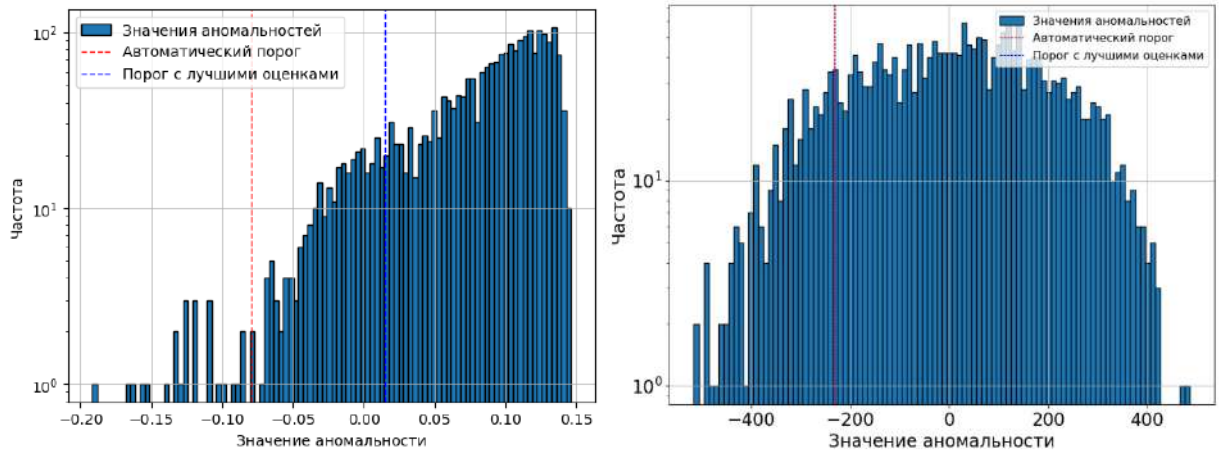
После обучения моделей, основанных на ИНС необходимо оценить значение аномальности для каждой записи, которая в данном случае является ошибкой реконструкции сети. Для каждого образца тестовых данных считается эта ошибка по формулам (4) и (5) для моделей автокодировщика и LSTM соответственно:

$$RecError_{AE} = \left| \sum X_{test} - \sum Y_{test} \right| , \quad (4)$$

$$RecError_{LSTM} = \frac{\left| \sum X_{test} - \sum Y_{test} \right|}{t} , \quad (5)$$

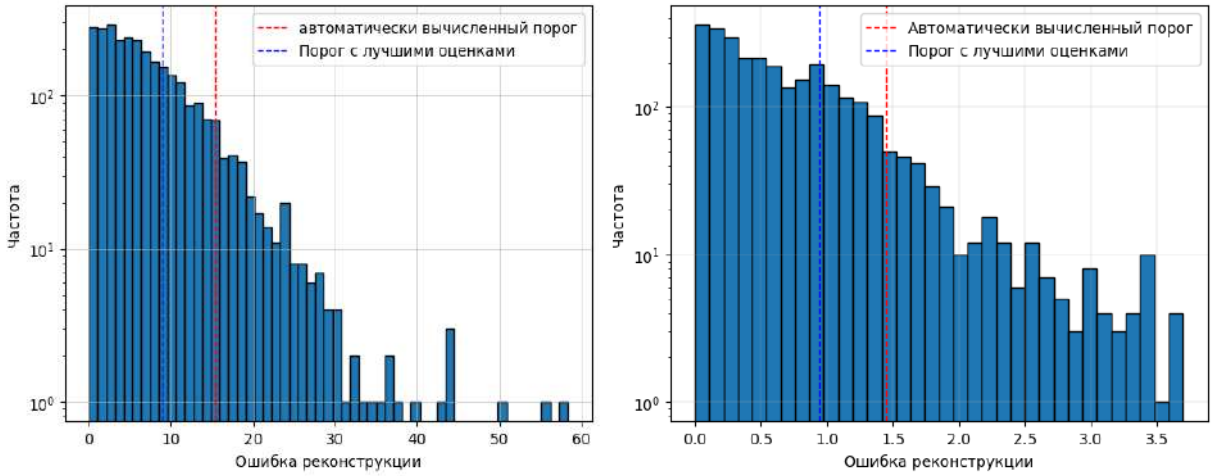
где X_{test} – вход сети, Y_{test} – выход сети, а t – временное окно, т.е. число подряд идущих записей, которые представляются в сеть LSTM как один обучающий образец.

По итогам обучения и полученных предсказаний лучшие оценки эффективности моделей для разного числа признаков и при разном подходе принятия решения представлены в таблице 2. Исходя из полученных результатов, можем сказать, что при большем числе признаков, т.е. при большей информации о действиях пользователей модели машинного обучения изоляционного леса и одноклассового SVM могут терять обобщающую способность, что обусловлено



а) *б)*

Рис. 3. Результат предсказания моделей машинного обучения на наборе данных со всеми признаками: а) модель изоляционного леса, б) модель одноклассового SVM



а) *б)*

Рис. 4. Результат предсказания моделей глубокого обучения на наборе данных со всеми признаками: а) модель Автокодировщика б) модель LSTM

особенностями вычисления значения аномальности. При этом для модели одноклассового SVM автоматически полученный порог близок к порогу с лучшими оценками, что может быть преимуществом для автоматической системы выявления внутренних угроз при минимальном участии экспертов

Для моделей, основанных на ИНС, напротив, большее число признаков дает больше информации для сетей, что увеличивает их обобщающую способность. При этом оценки при принятии решении об аномальности каждой записи заметно меньше, что также является следствием их обобщающей способности и предрасположенности к усреднению. В основе принятия решений моделей машинного обучения лежит принятие решения о каждом экземпляра без контекста ближайших экземпляров, в то время как модели, основанные на ИНС за счет большого числа подстраиваемых параметров, могут усреднять значения на выходе, в том числе из-за эффекта переобучения.

6. Описание применения предложенного подхода

В действительности система выявления внутренних угроз не может быть полностью автономна, как минимум в процессе принятия решений и применение соответствующих мер к пользователю, помеченному как аномальным, то есть несущим внутреннюю угрозу. Автоматическое и постоянное функционирование системы при данном подходе представляет специалистам информацию о записях пользовательских действий, агрегированных за сутки, значение аномальности для каждой такой записи, а также статистику записей, признанных системой аномальными на основе автоматически вычисленного порога.

Таблица 2

Оценки эффективности для моделей

Модель	Принятие решения об аномальности	Число признаков	Точность (Accuracy)	Полнота (Recall)
Изоляционный лес	По каждой записи	509	84.8%	78%
		100	92.3%	80%
	По каждому пользователю	509	79.6%	52%
		100	92.8%	80%
Одноклассовый SVM	По каждой записи	509	85%	72.2%
		100	91%	73%
	По каждому пользователю	509	89%	63%
		100	79%	79%
Автокодировщик	По каждой записи	509	68%	47%
		100	62%	44%
	По каждому пользователю	509	96.6%	91%
		100	96.4%	91%
LSTM	По каждой записи	509	71.4%	50%
		100	68%	44%
	По каждому пользователю	509	97%	96%
		100	96%	94%

Данная информация может служить предупреждением для специалистов, которые в зависимости от уточнения обстоятельств и причин признания таких действий аномальными должны реагировать и предпринимать определенные действия в отношении пользователя. Такой режим соответствует мониторингу инцидентов информационной безопасности, связанных с внутренними угрозами и аномальным поведением пользователей. Исходя из большого числа ложноположительных или ложноотрицательных результатов специалист также может изменять автоматический порог для более эффективной работы системы.

Система с подходом к принятию решения о внутренней угрозе каждого пользователя требует достаточного набора исторических данных и может быть использована в качестве инструмента проактивного поиска внутренних угроз. Проактивный поиск угроз исходит из предположения о том, что каждый пользователь несет внутреннюю угрозу для информационной системы и задача специалиста состоит в поиске следов и конкретных действий для проверки данной гипотезы. В данном случае система представляет ранжированный по количеству признанных системой аномальных записей список пользователей, на основе которого специалист итеративно проверяет каждого подозрительного пользователя.

В то время как исследователи по тематике внутренних угроз используют различные психометрические данные пользователей, преобразованные текстовые данные и другие специфические признаки, в предложенном подходе учитывается ограничения системе на доступ к таким данным. В условиях существования ограничений для системы выявления внутренних угроз, таких как наличие доступа только к количественным данным и невозможность получения конфиденциальных данных предложенный подход способен решать поставленную задачу с высокой эффективностью и менее требователен к вычислительным ресурсам в части обработки данных.

Большинство современных исследований не задается вопросом места моделей в системе выявления внутренних угроз, а задается вопросом эффективности конкретных моделей и подходов к обработке данных. В качестве основных результатов авторы приводят сравнения с другими публикациями, использующих аналогичные модели и подходы. Такое сравнение делается на основе лучших собственных и сравниваемых результатов. В таблице 3 приведено сравнение результатов настоящей работы и других работ для принятия решения по каждой записи (записи) и по каждому пользователю (пользователи).

Таблица 3

Сравнение результатов публикаций

Публикация	Оценки для моделей			
[25]	*Полнота (записи)*		*Полнота (пользователи)*	
	ИНС прямого распространения			
	91.38%		42.27%	
	Случайный лес			
	79.69%		37.5%	
[21]	*Полнота (пользователи)*		*Полнота (записи)*	
	Изоляционный лес			
	-		51%	
	ИНС прямого распространения			
	86%		40%	
[10]	*Полнота (записи)*			
	Изоляционный лес			
	70.18%			
	Автокодировщик			
	55.36%			
[26]	*Полнота (записи)*		*Точность (записи)*	
	Изоляционный лес			
	22.5%		92.8%	
	Одноклассовый SVM			
	99.2%		23.8%	
Настоящая работа	*Точность (записи)*	*Полнота (записи)*	*Точность (пользователи)*	*Полнота (пользователи)*
	Изоляционный лес			
	92.3%	80%	92.8%	80%
	Одноклассовый SVM			
	91%	73%	79%	79%
	Автокодировщик			
	68%	47%	96.6%	91%

В [25] исследуется аналогичный тип агрегации действий пользователей по дням и принятие решения на основе записей и пользователей, при этом полнота нейросетевой модели при принятии решения об аномальности пользователей больше, чем при принятии решения об аномальности записей, что соответствует полученным в настоящей работе результатам. При этом полнота всех моделей, представленных в данной работе выше при обоих подходах к принятию решения, что говорит о их преимуществе.

Работа [21] в свою очередь проводится более подробное исследование разных временных агрегаций данных, при этом лучшие оценки, показанные в таблице 3, получаются при агрегации данных именно за сутки, что подтверждает правильность выбора временной агрегации настоящей работы. Продолжение исследований тех же авторов в [10] исследует большее число представлений данных при тех же подходах к принятию решения, при этом полученные оценки эффективности близки к оценкам настоящей работы.

Большинство других исследований, например [26], сравнивает подходы к построению моделей и обработке, и извлечению данных с рассмотрением проблем выбора обучающих данных, зачастую включаю дополнительную синтетическую генерацию данных и принимая решения лишь по записям, что не соответ-

ствует реалистичному подходу к построению системы выявления внутренних угроз.

Из сравнений результатов видно, что при сохранении вариативности в подходах к принятию решения и аналогичном типе агрегации результаты применения настоящего подхода сопоставимы с аналогичными публикациями и в большинстве превосходят их. Данное сравнение, как и сравнение в других работах не может быть полностью справедливым из-за большой вариативности применяемых методах, которые разнятся от работы к работе. Полученное сравнение и оценки эффективности моделей могут служить хорошей мерой достоверности и применимости предложенного подхода.

Таким образом предложенный подход к построению системы выявления внутренних угроз не только обладает высокими оценками эффективности, но и способен помочь специалистам в извлечении скрытой в условиях ограничения доступа системы к источникам конфиденциальных данных информации. Поход предполагает возможность работы как в режиме постоянного мониторинга пользовательский действий, представленных как агрегированные за день количественные и категориальные признаки, так и в режиме проактивного поиска угроз с принятием решения о внутренней угрозе пользователей, который позволяет специалистам эффективно и наглядно проверять свои гипотезы относительно каждого пользователя.

Заключение

Использование мощного математического аппарата, заложенного в структуру моделей машинного обучения, в том числе основанных на ИНС, позволяет построить автономную систему выявления внутренних с минимальным привлечением экспертных оценок, предоставляя для них данные о записях журналов действий в системе, признанными аномальным, а также данные статистики о числе таких записей для каждого пользователя, на основе среднего которых система принимает решение об аномальности пользователя.

В данной работе была сформулирована задача выявления внутренних угроз, которая в рамках моделей машинного обучения формулируется как задача выявления аномалий, проведен анализ популярных неконтролируемых моделей машинного обучения, в том числе основанных на ИНС. Описанные методы извлечения, агрегации и представления данных, сформулированные с учетом ограничения реальных систем, были применены на существующем наборе данных, содержащем данные о действиях пользователей за сутки.

Для подходов принятия решения об аномальности записей и каждого пользователя были посчитаны оценки эффективности моделей. Полученные оценки могут быть интерпретированы как следствие обобщающих способностей моделей и особенностями подсчета значений аномальности. В зависимости от поставленной перед системой задачи следует выбирать различные модели и подходы к принятию решений, например, в случае систематического выявление внутренних угроз по каждой записи следует обратить внимание на классические модели машинного обучения с тщательно подобранными признаками, а в случае апостериорного анализа инцидентов безопасности с выявлением внутренней угрозы среди пользователей, целесообразным будет применение моделей машинного обучения, основанных на многослойных ИНС.

Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета.

Литература

1. Al-Mhiqani M.N., Ahmad R., Abidin Z.Z., Abdulkareem K.H., Mohammed M.A., Gupta D., Shankar K. A new intelligent multilayer framework for insider threat detection // Computers & Electrical Engineering. – 2022. – Vol. 97. – P. 107597. – DOI: 10.1016/j.compeleceng.2021.107597.
2. Kim A., Oh J., Ryu J., Lee K. A Review of Insider Threat Detection Approaches with IoT Perspective // IEEE Access. – 2020. – Vol. 8. – P. 78847-78867. – DOI: 10.1109/ACCESS.2020.2990195.
3. Kim J., Park M., Kim H., Cho S., Kang P. Insider Threat Detection Based on user Behavior Modeling and Anomaly Detection Algorithms // Applied Sciences. – 2019. – Vol. 9. – P. 4018. – DOI: 10.3390/app9194018.
4. Al-Mhiqani M.N., Ahmad R., Abidin Z.Z., Yassin W., Hassan A., Abdulkareem K.H., Ali N.S., Yunos Z. A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations // Applied Sciences. – 2020. – Vol. 10, no. 15. – P. 5208. – DOI: 10.3390/app10155208.
5. Goodfellow I., Bengio Y., Courville A. Deep learning. – MIT press, 2016.

6. Yuan S., Wu X. Deep learning for insider threat detection: Review, challenges and opportunities // Computers & Security. – 2021. – Vol. 104. – P. 102221. – DOI: 10.1016/j.cose.2021.102221.
7. Chauhan V. K., Dahiya K., Sharma A. Problem formulations and solvers in linear SVM: a review // Artificial Intelligence Review. – 2019. – Vol. 52, no. 2. – P. 803-855. – DOI: 10.1007/s10462-018-9614-6.
8. Khan S.S., Madden M.G. One-class classification: taxonomy of study and review of techniques // The Knowledge Engineering Review. – 2014. – Vol. 29, no. 3. – P. 345-374. – DOI: 10.1017/S026988891300043X.
9. Hurst W., Tekinerdogan B., Alskaif T., Boddy A. Securing electronic health records against insider-threats: A supervised machine learning approach // Smart Health. – 2022. – Vol. 26, no. 9. – P. 100354. – DOI: 10.1016/j.smhl.2022.100354.
10. Le D. C., Zincir-Heywood N. Anomaly detection for insider threats using unsupervised ensembles // IEEE Transactions on Network and Service Management. – 2021. – Vol. 18, no. 2. – P. 1152-1164. – DOI: 10.1109/TNSM.2021.3071928.
11. Sadaf K., Sultana J. Intrusion detection based on autoencoder and isolation forest in fog computing // IEEE Access. – 2020. – Vol. 8. – P. 167059-167068. – DOI: 10.1109/ACCESS.2020.3022855.
12. Hariri S., Kind M. C., Brunner R. J. Extended isolation forest // IEEE Transactions on Knowledge and Data Engineering. – 2019. – Vol. 33, no. 4. – P. 1479-1489. – DOI: 10.1109/TKDE.2019.2947676.
13. Pouyanfar S., Sadiq S., Yan Y., Tian H., Tao Y, Reyes M.P., Shyu M.-L., Chen S.-C., Iyengar S.S. A survey on deep learning: Algorithms, techniques, and applications // ACM Computing Surveys (CSUR). – 2018. – Vol. 51, no. 5. – P. 1-36. – DOI: 10.1145/3234150.
14. Merrill N., Eskandarian A. Modified autoencoder training and scoring for robust unsupervised anomaly detection in deep learning // IEEE Access. – 2020. – Vol. 8. – P. 101824-101833. – DOI: 10.1109/ACCESS.2020.2997327.
15. Pantelidis E., Bendiab S., Kolokotronis N. Insider threat detection using deep autoencoder and variational autoencoder neural networks // 2021 IEEE International Conference on Cyber Security and Resilience (CSR). – IEEE, 2021. – P. 129-134. – DOI: 10.1109/CSR51186.2021.9527925.
16. Tang T.A., Mhamdi L., McLernon D., Zaidi S.A.R., Ghogho M. Deep Recurrent Neural Network for Intrusion Detection in SDN-based Networks // 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft). – IEEE, 2018. – P. 202-206. – DOI: 10.1109/NETSOFT.2018.8460090.
17. Buber E., Diri B. Web page classification using RNN // Procedia Computer Science. – 2019. – Vol. 154. – P. 62-72. – DOI: 10.1016/j.procs.2019.06.011.
18. Sharma B., Pokharel P., Joshi B. User behavior analytics for anomaly detection using LSTM autoencoder-insider threat detection // Proceedings of the 11th International Conference on Advances in Information Technology. – Bangkok, Thailand, 2020. – P. 1–9. – DOI: 10.1145/3406601.3406610.
19. Гайдук К.А., Исхаков А.Ю. К вопросу о реализации алгоритмов выявления внутренних угроз с применением машинного обучения // Вестник СибГУТИ. – 2022. – № 16(4). – С. 80-95. – DOI: 10.55648/1998-6920-2022-16-4-80-95.
20. Lindauer B. Insider Threat Test Dataset. Carnegie Mellon University. Dataset. – URL: https://doi.org/10.1184/R1/12841247.v1 (accessed: 19.07.2023).
21. Le D.C., Zincir-Heywood N., Heywood M.I. Analyzing data granularity levels for insider threat detection using machine learning // IEEE Transactions on Network and Service Management. – 2020. – Vol. 17, no. 1. – P. 30-44. – DOI:10.1109/TNSM.2020.2967721.
22. Al-Shehari T., Alsowail R.A. An Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques // Entropy. – 2021. – Vol. 23, no. 10. – P. 1258. – DOI: 10.3390/e23101258.
23. Li X., Wang Y., Basu S., Kumbier K., Yu B. A debiased MDI feature importance measure for random forests // Advances in Neural Information Processing Systems. – 2019. – Vol. 32. – P. 1-19.
24. Lundberg S.M., Lee S.I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 1-10.
25. Le D.C., Zincir-Heywood A.N. Machine learning based insider threat modelling and detection // 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM). – IEEE, 2019. – P. 1-6.
26. Bartoszewski F.W., Just, M., Lones, M.A., Mandrychenko, O. Anomaly Detection for Insider Threats: An Objective Comparison of Machine Learning Models and Ensembles // ICT Systems Security and Privacy Protection: 36th IFIP TC 11 International Conference, SEC 2021. – Cham : Springer International Publishing, 2021. – P. 367-381. – DOI: 10.1007/978-3-030-78120-0_24.

PROACTIVE SEARCH FOR INTERNAL THREATS TO INFORMATION SECURITY IN CONDITIONS OF CONSTRAINTS

Iskhakov A.Yu. [3], Gaiduk K.A. [4]

Purpose of work: *research of methods for detecting internal threats to information security and improving their effectiveness through the modernization of the approach to building an internal threat detection system.*

3 Andrey Yu. Iskhakov, Ph.D. in Engineering, Financial University, Moscow, Russia, E-mail: iskhakovandrey@gmail.com, ORCID: 0000-0002-6603-265X
4 Kirill A. Gayduk, National Research Nuclear University MEPHI, Moscow, Russia, E-mail: guydukk@gmail.com, ORCID: 0009-0002-3276-0891

Research method: *system analysis of open data sources on methods for detecting internal threats to information security; construction of a model based on the application of machine learning methods; methods for sample formation and training, assessment of anomaly based on decision-making methods.*

The result obtained: *The article presents a study of existing methods and models for detecting internal threats based on unsupervised machine learning, including multi-layer artificial neural networks. An approach to building an internal threat detection system is described and formalized, which is then applied to an existing dataset to evaluate the effectiveness of the models. The uniqueness of the proposed approach lies in the ability to detect internal threats for each record individually, as well as for each user based on the number of anomalous records. The described methods for data extraction, aggregation, and representation, formulated with the constraints of real systems in mind, were applied to a dataset of user actions over a day. Evaluations of model effectiveness were calculated for anomaly record decision-making approaches and each user. The obtained evaluations can be interpreted as a result of the models' generalization abilities and the peculiarities of calculating anomaly values.*

Scientific novelty: *the article presents an approach to detecting internal threats to information security that differs in the ability to detect internal threats for each record individually, as well as for each user based on the number of anomalous records.*

Keywords: *internal threats, insider, threat detection, information security, anomalies, machine learning, artificial neural networks.*

References

1. Al-Mhiqani M.N., Ahmad R., Abidin Z.Z., Abdulkareem K.H., Mohammed M.A., Gupta D., Shankar K. A new intelligent multilayer framework for insider threat detection // Computers & Electrical Engineering. – 2022. – Vol. 97. – P. 107597. – DOI: 10.1016/j.compeleceng.2021.107597.
2. Kim A., Oh J., Ryu J., Lee K. A Review of Insider Threat Detection Approaches with IoT Perspective // IEEE Access. – 2020. – Vol. 8. – P. 78847-78867. – DOI: 10.1109/ACCESS.2020.2990195.
3. Kim J., Park M., Kim H., Cho S., Kang P. Insider Threat Detection Based on user Behavior Modeling and Anomaly Detection Algorithms // Applied Sciences. – 2019. – Vol. 9. – P. 4018. – DOI: 10.3390/app9194018.
4. Al-Mhiqani M.N., Ahmad R., Abidin Z.Z., Yassin W., Hassan A., Abdulkareem K.H., Ali N.S., Yunos Z. A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations // Applied Sciences. – 2020. – Vol. 10, no. 15. – P. 5208. – DOI: 10.3390/app10155208.
5. Goodfellow I., Bengio Y., Courville A. Deep learning. – MIT press, 2016.
6. Yuan S., Wu X. Deep learning for insider threat detection: Review, challenges and opportunities // Computers & Security. – 2021. – Vol. 104. – P. 102221. – DOI: 10.1016/j.cose.2021.102221.
7. Chauhan V. K., Dahiya K., Sharma A. Problem formulations and solvers in linear SVM: a review // Artificial Intelligence Review. – 2019. – Vol. 52, no. 2. – P. 803-855. – DOI: 10.1007/s10462-018-9614-6.
8. Khan S.S., Madden M.G. One-class classification: taxonomy of study and review of techniques // The Knowledge Engineering Review. – 2014. – Vol. 29, no. 3. – P. 345-374. – DOI: 10.1017/S026988891300043X.
9. Hurst W., Tekinerdogan B., Alskaif T., Boddy A. Securing electronic health records against insider-threats: A supervised machine learning approach // Smart Health. – 2022. – Vol. 26, no. 9. – P. 100354. – DOI: 10.1016/j.smhl.2022.100354.
10. Le D. C., Zincir-Heywood N. Anomaly detection for insider threats using unsupervised ensembles // IEEE Transactions on Network and Service Management. – 2021. – Vol. 18, no. 2. – P. 1152-1164. – DOI: 10.1109/TNSM.2021.3071928.
11. Sadaf K., Sultana J. Intrusion detection based on autoencoder and isolation forest in fog computing // IEEE Access. – 2020. – Vol. 8. – P. 167059-167068. – DOI: 10.1109/ACCESS.2020.3022855.
12. Hariri S., Kind M. C., Brunner R. J. Extended isolation forest // IEEE Transactions on Knowledge and Data Engineering. – 2019. – Vol. 33, no. 4. – P. 1479-1489. – DOI: 10.1109/TKDE.2019.2947676.
13. Pouyanfar S., Sadiq S., Yan Y., Tian H., Tao Y, Reyes M.P., Shyu M.-L., Chen S.-C., Iyengar S.S. A survey on deep learning: Algorithms, techniques, and applications // ACM Computing Surveys (CSUR). – 2018. – Vol. 51, no. 5. – P. 1-36. – DOI: 10.1145/3234150.
14. Merrill N., Eskandarian A. Modified autoencoder training and scoring for robust unsupervised anomaly detection in deep learning // IEEE Access. – 2020. – Vol. 8. – P. 101824-101833. – DOI: 10.1109/ACCESS.2020.2997327.
15. Pantelidis E., Bendiab S., Kolokotronis N. Insider threat detection using deep autoencoder and variational autoencoder neural networks // 2021 IEEE International Conference on Cyber Security and Resilience (CSR). – IEEE, 2021. – P. 129-134. – DOI: 10.1109/CSR51186.2021.9527925.
16. Tang T.A., Mhamdi L., McLernon D., Zaidi S.A.R., Ghogho M. Deep Recurrent Neural Network for Intrusion Detection in SDN-based Networks // 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft). – IEEE, 2018. – P. 202-206. – DOI: 10.1109/NETSOFT.2018.8460090.
17. Buber E., Diri B. Web page classification using RNN // Procedia Computer Science. – 2019. – Vol. 154. – P. 62-72. – DOI: 10.1016/j.procs.2019.06.011.
18. Sharma B., Pokharel P., Joshi B. User behavior analytics for anomaly detection using LSTM autoencoder-insider threat detection // Proceedings of the 11th International Conference on Advances in Information Technology. – Bangkok, Thailand, 2020. – P. 1–9. – DOI: 10.1145/3406601.3406610.

19. Gajduk K.A., Ishakov A.Ju. K voprosu o realizacii algoritmov vyjavlenija vnutrennih ugroz s primeneniem mashinnogo obuchenija // Vestnik SibGUTI. – 2022. – № 16(4). – S. 80-95. – DOI: 10.55648/1998-6920-2022-16-4-80-95.
20. Lindauer B. Insider Threat Test Dataset. Carnegie Mellon University. Dataset. – URL: https://doi.org/10.1184/R1/12841247.v1 (accessed: 19.07.2023).
21. Le D.C., Zincir-Heywood N., Heywood M.I. Analyzing data granularity levels for insider threat detection using machine learning // IEEE Transactions on Network and Service Management. – 2020. – Vol. 17, no. 1. – P. 30-44. – DOI:10.1109/TNSM.2020.2967721.
22. Al-Shehari T., Alsowail R.A. An Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques // Entropy. – 2021. – Vol. 23, no. 10. – P. 1258. – DOI: 10.3390/e23101258.
23. Li X., Wang Y., Basu S., Kumbier K., Yu B. A debiased MDI feature importance measure for random forests // Advances in Neural Information Processing Systems. – 2019. – Vol. 32. – P. 1-19.
24. Lundberg S.M., Lee S.I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 1-10.
25. Le D.C., Zincir-Heywood A.N. Machine learning based insider threat modelling and detection // 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM). – IEEE, 2019. – P. 1-6.
26. Bartoszewski F.W., Just, M., Lones, M.A., Mandrychenko, O. Anomaly Detection for Insider Threats: An Objective Comparison of Machine Learning Models and Ensembles // ICT Systems Security and Privacy Protection: 36th IFIP TC 11 International Conference, SEC 2021. – Cham : Springer International Publishing, 2021. – P. 367-381. – DOI: 10.1007/978-3-030-78120-0_24.