ПОЛИТЕХ
Санкт-Петербургский
политехнический университет
Петра Великого

На правах рукописи

Раковский Дмитрий Игоревич

Многозначная классификация сетевых атак методами машинного обучения

2.3.6. Методы и системы защиты информации,
информационная безопасность

АВТОРЕФЕРАТ

Диссертации на соискание ученой степени

кандидата технических наук

Москва
2025

Работа выполнена на кафедре «Информационная безопасность» ордена Трудового Красного знамени Федерального государственного бюджетного образовательного учреждения высшего образования «Московский технический университет связи и информатики» (МТУСИ), г. Москва.

Научный руководитель:
Заслуженный деятель науки РФ, доктор технических наук, профессор Шелухин Олег Иванович.

Официальные оппоненты:
доктор технических наук, старший научный сотрудник Марков Алексей Сергеевич, акционерное общество «Научно-производственное объединение «Эшелон», президент, г. Москва.

кандидат технических наук, доцент Красов Андрей Владимирович, федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича», заведующий кафедрой «Защищённые системы связи», г. Санкт-Петербург.

Ведущая организация:
Федеральное государственное автономное образовательное учреждение высшего образования «Южный федеральный университет» (ЮФУ), г. Ростов-на-Дону.

Защита состоится «11» февраля 2025 года в 14:00 ч.
на заседании диссертационного совета У.2.3.6.12 федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого» (195251, г. Санкт-Петербург, ул. Политехническая, д. 29, Главный корпус, аудитория 175).

С диссертацией можно ознакомиться в библиотеке и на сайте www.spbstu.ru федерального государственного автономного образовательного учреждения высшего образования «Санкт- Петербургский политехнический университет Петра Великого»

Автореферат разослан «___» _____ 2025 г.

Ученый секретарь
диссертационного совета У.2.3.6.12,
канд. физ.-мат. наук, доцент Шенец Николай Николаевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Современные компьютерные сети (КС) порождают разнородные и многомерные наборы данных. Данные агрегируют с системных журналов, сетевых карт, системных датчиков, установленных на каждом из хостов. В следствие роста количества комбинированных компьютерных атак (КА), важно иметь возможность отслеживать их одновременную реализацию на один узел подконтрольной компьютерной сети. Под КА понимается целенаправленное несанкционированное воздействие на информацию или ресурс КС. В рамках диссертационного исследования рассматривается подвид КА, связанный с использованием протоколов межсетевого взаимодействия – сетевые атаки.

Для обеспечения информационной безопасности (ИБ) КС, при создании систем обнаружения вторжений (СОВ), необходимо учитывать возможность одновременного совершения КА. В случае успешной реализации КА в КС наблюдается нарушение функционирования за счет деструктивного воздействия на доступность и целостность поддерживающей инфраструктуры и циркулирующей в ней информации. Для обнаружения КА, из исходных данных извлекаются атрибуты, описывающие поведение сети за период агрегации.

Современные СОВ проектируются на базе методов машинного обучения (МО), основанных на предварительной корректировке своих параметров на основании обучающих данных (методы контролируемого обучения). Одним из семейств методов МО, релевантных задаче обнаружения КА, являются вычислительные графы, включающие искусственные нейронные сети (ИНС). Применение ИНС в ИБ перспективно, поскольку позволяет учитывать нелинейные связи между пространствами атрибутов и целевых атрибутов за счет своей структуры.

На основании мониторинга данных, поступающих на вход СОВ, происходит обнаружение и последующая классификация КА. Исследования показывают, что КА разных типов могут быть идентичны по значениям всех атрибутам, за исключением целевых атрибутов. Это приводит к ошибкам при работе СОВ, построенных на однозначных алгоритмах МО. Известны базы данных, широко используемые для разработки СОВ, в которых проявляется многозначность целевых атрибутов, однако данный факт, как правило, не учитывается разработчиками СОВ.

Степень разработанности темы исследования. Обнаружению и классификации КА для обеспечения информационной безопасности посвящены работы российских и зарубежных ученых: *Зегжда П.Д., Зегжда Д.П., Лаврова Д.С., Козачок А.В., Марков А.С., Крундышев В. М., Калинин М.О., Шелухин О.И., Mirsky Y., Doitshman T., Elovici Y., Shabtai A. и др.* Большая часть работ посвящена выявлению аномалий в одномерных временных рядах. Исследования *Котенко И.В., Хмырова С.С., Саенко И.Б., Лаута О.С., Скоробогатова С.Ю., Полониной Я.В., Петренко С.А.* посвящены обнаружению КА посредством бинарных и многоклассовых алгоритмов МО. В работах *Павленко Е.Ю.* и *Лавровой Д.С.* КА обнаруживаются при помощи спектрального анализа моделирующих графов.

Многозначной классификации в области кибернетики, анализа языка и текстов, медицины посвящены работы отечественных и зарубежных авторов: *Молодняков С.А., Tsoumakas G., Gibaja E., Ventura S., Молошников И.А., Олисеенко В.Д., Тулупьева Т.В., Карпович С.Н., Остапец А.А. и др.*

Однако в области ИБ работ, связанных с многозначной классификацией КА, крайне мало. Это определяет направление диссертационного исследования.

Областью исследования является разработка моделей и методов многозначной классификации КА.

Объект исследования: компьютерная сеть, подвергаемая одновременному воздействию нескольких компьютерных атак.

Предмет исследования: модели и методы многозначной классификации компьютерных атак и предупреждения нарушений функционирования КС.

Цель работы состоит в обеспечении ИБ путем повышения точности классификации сетевых атак в КС в условиях многозначности целевых атрибутов, маркирующих их тип.

Для достижения вышеуказанной цели представляется необходимым решить следующие задачи:

1. Провести анализ специфических особенностей многозначной классификации в задачах обнаружения и классификации КА.
2. Разработать модель табличного представления профилей функционирования КС, учитывающую многозначность целевых атрибутов, связанных с реализацией КА.
3. Разработать метод и алгоритм многозначной классификации КА, учитывающий нелинейную связь между множеством атрибутов КС и множеством целевых атрибутов; оценить точность многозначной классификации с использованием введенного множества метрик.
4. Разработать программно-аппаратный комплекс (ПАК) для сбора телеметрии и имитационного моделирования КА, обладающих свойством многозначности целевых атрибутов.

Методы исследований: системный подход, математический анализ, теория графов, теория множеств, математическая статистика, методы синтеза искусственных нейронных сетей, методы машинного обучения.

Научная новизна работы:

1. Разработана модель табличного представления профилей функционирования КС, учитывающая учитывающую многозначность целевых атрибутов, связанных с реализацией КА.
2. Предложены метод и алгоритм многозначной классификации КА, заключающиеся в многозначном отображении пространства атрибутов в пространство целевых атрибутов, отличающиеся от известных аналогов

дублированием и последующей декомпозицией пространства атрибутов по каждому целевому атрибуту.

Научные результаты, представленные в п. 1-2, соответствуют *пункту №6* паспорта научной специальности 2.3.6. «Методы и системы защиты информации, информационная безопасность» - «Методы, модели и средства мониторинга, предупреждения, обнаружения и противодействия нарушениям и компьютерным атакам в компьютерных сетях».

Теоретическая значимость работы состоит в разработке нового метода и алгоритма многозначной классификации компьютерных атак, расширяющих научно-методический инструментарий обеспечения защищенности КС различной конфигурации.

Разработана модель табличного представления профилей функционирования КС, учитывающая многозначность целевых атрибутов, связанных с реализацией КА.

Решена задача параллельной классификации нескольких КА в КС с помощью ИНС в условиях многозначности исходных данных и нелинейных связей между входными атрибутами и классовыми метками.

Практическая ценность определяется возможностью использования разработанной модели, метода и алгоритма для предотвращения воздействия КА на КС в условиях многозначности целевых атрибутов набора данных. Результаты работы позволяют:

– осуществить сбор телеметрии и результатов имитационного моделирования КА, обладающих свойством многозначности, в контролируемых условиях;

– автоматизировать маркировку КА, направленных на КС, позволяющую конфигурировать свойство многозначности данных в исследовательских целях.

Апробация результатов. Основные результаты диссертационной работы обсуждались на научных конференциях:

33-я Всероссийская конференция «Методы и технические средства обеспечения безопасности информации», Санкт-Петербург, 2024 г.; X Всероссийская научно-техническая конференция «Фундаментальные и прикладные аспекты компьютерных технологий и информационной безопасности», Таганрог, 2024 г.; VIII Международная отраслевая научно-техническая конференция «Технологии Информационного Общества», Москва, 2024 г.; XII Международная научно-техническая конференция, посвященная 25-летию кафедры ИУ8; III Всероссийская научно-практическая конференция «Теория и практика обеспечения информационной безопасности», 2023 г.; III Всероссийская научная школа-семинар «Современные тенденции развития методов и технологий защиты информации», Москва, 2023 г.; VII Международная отраслевая научно-техническая конференция «Технологии Информационного Общества», Москва, 2023 г.; International Scientific Conference «*2023 Systems Of Signals Generatingand Processing In The Fieldof On Board Communications*», Москва, 2023 г.; Научно-техническая конференция «Управление и безопасность в информационных и киберфизических системах», 2023 г.; II

Всероссийская научно-практическая конференция «Теория и практика обеспечения информационной безопасности», 2022 г.; XV Международная отраслевая научно-техническая конференция «Технологии информационного общества» Москва, 2021 г.

Основные положения, выносимые на защиту:

1. Модель табличного представления профилей функционирования КС, учитывающая многозначность целевых атрибутов, связанных с реализацией КА, повышающая точность классификации, в среднем, на 5% в сравнении однозначным представлением.

2. Метод и алгоритм многозначной классификации КА в КС, реализованные в виде архитектуры ИНС с множественным выходом, позволяющие повысить точность классификации до 16% в сравнении с известными алгоритмами за счет нелинейного многозначного отображения входных атрибутов в выходные целевые атрибуты.

3. Программно-аппаратный комплекс, позволяющий автоматизировать сбор телеметрии и имитационное моделирование КА, обладающих свойством многозначности целевых атрибутов.

Публикации. Основные положения диссертации опубликованы в ведущих рецензируемых научно-технических журналах, входящих в Перечень ВАК в количестве 9 публикаций; 1 публикация в международной базе данных *Scopus*; 4 публикации в *RSCI*.

Получено 5 свидетельств о государственной регистрации программы для ЭВМ. Всего по теме диссертационного исследования опубликовано 27 работ.

Внедрение результатов работы. Исследования выполнялись в рамках работы по гранту, направленного на обеспечение информационной безопасности для задач цифровой экономики, соглашение № 40469–21/23-К от 30.06.2023 г. аспирантам, соискателям и молодым ученым на исследования, направленные на обеспечение информационной безопасности для задач цифровой экономики и по государственной поддержке ведущих научных школ Российской Федерации в области информационной безопасности.

Результаты диссертации внедрены в ФГУП «Научно-исследовательский институт «Квант» при проектировании и создании автоматизированных систем в защищенном исполнении, моделировании вторжений в действующие корпоративные сети и выполнении НИР «Механизм-ФЛ», а также в учебный процесс МТУСИ при реализации рабочей программы дисциплины «Машинное обучение с использованием Python» в виде лекционных и практических занятий для студентов очной формы обучения, обучающихся по направлению 10.03.01.

Личный вклад. В опубликованные научные труды диссертантом внесен основной вклад, касающийся разработки алгоритмов и их практической реализации.

Вклад соавторов ограничивался постановкой задач на исследования и обсуждением полученных результатов.

Достоверность. Достоверность результатов диссертационной работы подтверждается совпадением результатов имитационного моделирования с результатами экспериментальных данных, корректным использованием современного математического аппарата, апробацией на научно-технических конференциях.

Объем и структура работы. Диссертация содержит 142 страницы текста без учета приложений и списка литературы, 30 рисунков, 23 таблицы и состоит из введения, четырех разделов, заключения, списка литературы и четырех приложений. Список литературы состоит из 248 наименований. Приложения содержат 13 страниц, включая акты о внедрении и использования результатов диссертационных исследований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы диссертационного исследования, сформулирована цель, определены основные задачи, научная новизна и практическая значимость полученных результатов, а также положения, выносимые на защиту.

В **первой главе** проведен аналитический обзор объекта и предмета исследования, выявлены основные предпосылки для разработки новых алгоритмов обнаружения КА и предупреждения нарушений функционирования КС на основе многозначных закономерностей.

Проанализированы подходы к работе с многозначными данными, и на его основании выбран подход *algorithm adaptation methods*, заключающийся в адаптации (разработке новых) методов и алгоритмов принятия решений в условиях многозначности целевых атрибутов.

В области ИБ источниками многозначных целевых атрибутов являются:

– возможность одновременной реализации КА на хосты КС, обладающие несколькими сетевыми интерфейсами;

– возможность одновременной реализации КА через несколько сетевых шлюзов КС.

– коллизии, возникающие при преобразовании данных, описывающих многопоточное функционирование сети (например, формат *.pcap*) в двухмерную табличную структуру;

– коллизии, возникающие при передискретизации данных табличной структуры в следствие их предварительной обработки;

– рост доли многозначных целевых атрибутов в следствие сокращения размерности исходных данных.

На основании анализа первоисточников сформулирована математическая постановка задачи повышения точности классификации сетевых атак в КС в условиях многозначности целевых атрибутов:

$$METHOD:\langle \boldsymbol{A},\boldsymbol{L},\Phi,S,\boldsymbol{\Theta}\rangle \to \max \Phi^{it}, \quad (1)$$

где $\boldsymbol{A}$ – непустое пространство атрибутов, $\boldsymbol{A}\in R^{\Lambda}$;

$\boldsymbol{L}$ – пространство целевых атрибутов, $\boldsymbol{L}\in R^{\Xi}$;

Φ – множество метрик оценки точности классификации, $\Phi=\{\varphi_\zeta\}\,|\,\forall\varphi_\zeta\in[0;1],\zeta=\overline{1,Z}$; определено содержимое множества $\Phi=\{accuracy_{ovo}, precision_{ovo}, recall_{ovo}, f_{meanovo}, AUC_{ovo}\}$.

S – граничные условия метода;

$\boldsymbol{\Theta}$ – варьируемые параметры, уточняемые под реализацию метода;

Z – количество показателей оценки точности;

Φ^{it} – интегральный показатель оценки точности классификации, $\Phi^{it}=\sum_{1}^{Z}\varphi_\zeta$, рассчитываемый как сумма всех показателей оценки точности классификации, каждый из которых должен быть нормированной величиной: $\forall\varphi_\zeta\in[0;1]$.

Оценка выигрыша по точности производится на основании вычисления $\Delta\Phi^{it}=\Phi^{itП}-\Phi^{itД}$ - разницы между интегральным показателем оценки точности после использования метода классификации КА $\Phi^{itП}$ и до его использования - $\Phi^{itД}$;

Основными граничными условиями (1) являются $S=\{s_1,s_2,s_3,s_4\}$:

$s_1:\Xi>1$ - размерность пространства целевых атрибутов больше 1;

$s_2:\left|\Phi^{it}-\widehat{\Phi}^{it}\right|\le eps$ - точность классификации на произвольном наборе данных, релевантном исходному, не должна отличаться более, чем на *eps*;

$s_3:|\boldsymbol{L}|\le|\boldsymbol{A}|$ - размерность пространства целевых атрибутов $\boldsymbol{L}$ не должна превышать $\boldsymbol{A}$;

s_4: Данные полностью описывают функционирование компьютерной сети.

Дополнительным условием, накладываемым на выбор показателей оценки точности, является их интерпретируемость. Чем ближе значение φ_ζ к единице, тем точнее метод классификации решает поставленную задачу.

Во второй главе, в соответствии с поставленной задачей (1), введена в рассмотрение модель табличного представления поведения КС, подвергаемой КА, основанная на методе представления многозначных данных *binary relevance и* учитывающая свойство многозначности целевых атрибутов.

Показано, что данные о поведении КС могут быть представлены в виде объединения двух таблиц: таблицы атрибутов размером Λ (столбцов) N (строк) $\boldsymbol{A}$, и таблицы целевых атрибутов размером Ξ (столбцов) N (строк) $\boldsymbol{L}$:

$$\boldsymbol{D}_{NM}=\left\{\left(\boldsymbol{A}(n,),\boldsymbol{L}(n,)\right);\boldsymbol{A}=(a_{n\lambda}),\boldsymbol{L}=\left(l_{n\xi}\right),\lambda=\overline{1,\Lambda},\xi=\overline{1,\Xi},n=\overline{1,N},M=\Lambda+\Xi\right\}, \quad (1)$$

где: $\boldsymbol{A}(n,)=(a_{n1},a_{n2},\dots,a_{n\Lambda})$ - n-ный вектор-строка матрицы атрибутов экспериментальных данных $\boldsymbol{A}$, состоящий из Λ столбцов. Элемент этой строки, $a_{ni}\in\boldsymbol{A}(n,)$, представляет собой метрическое значение <u>i-го атрибута на n-ой строке</u> экспериментальных данных (например, связанного с загрузкой оперативной памяти хоста);

$\boldsymbol{L}(n,\ ) = (l_{n1}, l_{n2}, \ldots, l_{n\Xi})$ - n-ый вектор-строка матрицы целевых атрибутов экспериментальных данных $\boldsymbol{L}$ состоящая из Ξ столбцов;

Элемент $l_{ni} \in \boldsymbol{L}(n,\ ); l_{ni} \in \{0,1\}; n = \overline{1,N}$, означает значение i-го целевого атрибута (классовой метки) на n-той строке экспериментальных данных (например, бинарное значение о совершении компьютерной атаки типа «отказ в обслуживании»); N – количество записей экспериментальных данных; M – совокупное количество столбцов в (1), $M = \Lambda + \Xi$.

Если $\Xi=1$, то задача анализа таблицы (1) является однозначной, поскольку существует только один целевой атрибут. При $\Xi>1$ задача анализа (1) является многозначной.

Предложенная модель (1) требует выполнения двух шагов, связанных с предобработкой данных. На первом шаге алгоритмом поиска полных дубликатов по атрибутному пространству выявляются многозначные закономерности - записи, имеющие одинаковые значения по всем атрибутам за исключением целевых. На втором шаге дубли группируются и каждой записи присваиваются все маркеры КА, обнаруженные внутри группы.

В рамках решения задачи классификации КА, предложенная модель требует тождественности Ξ количеству уникальных КА (уникальных классовых меток). Каждый из Ξ столбцов включается в модель (1) в виде булева вектора $(l_{n\xi})$, где «1» означает присутствие ξ-ой атаки в момент внесения вектора-строки матрицы атрибутов $\boldsymbol{A}(n,\ )$ в таблицу (1), а «0» означает отсутствие указанной атаки.

Итогами анализа модели (1) является:

– обнаружение одновременно реализующихся КА в КС, а также построение ПФ КС;

– формирование статистики о распределении КА по однозначным и многозначным записям;

– принятие на основании полученной информации решений о мерах противодействия воздействию КА.

При решении задач методами МО важную роль играют данные, соответствующие решаемой задаче. Выявлено, что многозначность «исторических данных» ранее не учитывалась в задачах информационной безопасности. При создании СОВ, на основе данных в табличном представлении, исходные данные, как правило, считаются однозначными, что приводит к дополнительным ошибкам.

Предложенная модель требует предобработки целевых атрибутов: выполнения операции поиска полных дубликатов по атрибутному пространству и последующей переразметки данных, что снижает ошибки первого и второго рода в силу корректного учета многозначных закономерностей.

Проанализированы известные наборы данных, находящиеся в открытом доступе, (*SR-BH 2020; UNSW-NB15; NF-UQ-NIDS* и др.), на многозначность целевых атрибутов с использованием алгоритма поиска полных дубликатов. Выявлена многозначность целевых атрибутов во всех анализируемых наборах данных.

Проведенный анализ показал, что доля многозначных целевых атрибутов составила от 1,3% до 86%. Тем самым обоснована возможность повышения уровня

информационной безопасности КС от воздействия КА в условиях многозначности данных с помощью разработки алгоритмов, учитывающих многозначность целевых атрибутов в исходных данных.

В третьей главе на основе разработанной модели (1) предложен метод многозначной классификации на основании дублирования и последующей декомпозиции атрибутного пространства (далее – метод), учитывающий многозначность исходных данных и нелинейные связи между входными атрибутами и классовыми метками. На основании предложенного метода разработан алгоритм классификации КА и архитектура ИНС.

Суть предлагаемого метода заключается в представлении данных согласно модели (1) и последующем отображении пространства атрибутов в пространство целевых атрибутов. Отображение осуществляется посредством дублирования и последующей декомпозиции пространства атрибутов по каждому целевому атрибуту. Декомпозиция дубликатов выполняется за счет последовательного нелинейного отображения входных данных в значение целевого атрибута. На этапе декомпозиции учитывается нелинейная связь между множеством атрибутов КС и множеством целевых атрибутов. Процедура отображения осуществляется в следующей последовательности.

Пусть исходные данные представляются в виде модели (1), а немаркированные «новые» данные одинаковой размерности, представлены в виде:

$$\hat{\boldsymbol{D}}_{\hat{N}M} = \left\{ \hat{\boldsymbol{A}}(\hat{n},\ );\hat{\boldsymbol{A}} = (\hat{a}_{\hat{n}\lambda}), \lambda = \overline{1,\Lambda}, \hat{n} = \overline{1,\hat{N}} \right\}, \quad (2)$$

где $\hat{N} \in \mathbb{N}$ - количество неразмеченных данных, $\hat{\boldsymbol{A}}(\hat{n},\ )$ - $\hat{n}$ -я строка в наборе неразмеченных экспериментальных данных; количество строк идентично (1). Тогда классификация немаркированной записи является процессом отображения немаркированной $\hat{n}$ -ой записи (вектора-строки) $\hat{\boldsymbol{A}}(\hat{n},\ )$ в соответствующий набор целевых атрибутов $set_{\hat{n}}$.

Процесс маркировки $\hat{\boldsymbol{A}}(\hat{n},\ )$ может быть выполнен посредством аппроксимации $\hat{\boldsymbol{A}}(\hat{n},\ )$ по некоторой метрике φ, ассоциированной с истинным набором многозначных целевых атрибутов для $\hat{\boldsymbol{A}}(\hat{n},\ )$ - $set_{\hat{n}}^{true}$.

«Алфавит», образованный уникальными строками $\boldsymbol{L}_{,}$ сводится к отдельному множеству $S = \bigcup_{n=1}^{N} \boldsymbol{L}(n,\ ); n = \overline{1,N}$, в котором объединены все строки $\boldsymbol{L}_{.}$ Результатом классификации является точечно-множественное отображение $\hat{\boldsymbol{A}}(\hat{n},\ ) \to \hat{\boldsymbol{L}}(\hat{n},\ )$ - такое, что при пересечении предсказанных строк целевых атрибутов $\hat{\boldsymbol{L}}(\hat{n},\ )$ и строк с истинным метками $\hat{\boldsymbol{L}}(\hat{n},\ )^{true}$, результирующее множество равно истинному: $\hat{\boldsymbol{L}}(\hat{n},\ ) \cap \hat{\boldsymbol{L}}(\hat{n},\ )^{true} = \hat{\boldsymbol{L}}(\hat{n},\ )^{true}$.

Для учета ситуаций, когда $\hat{\boldsymbol{L}}(\hat{n},\ ) \cap \hat{\boldsymbol{L}}(\hat{n},\ )^{true} \neq \hat{\boldsymbol{L}}(\hat{n},\ )^{true}$, проведен анализ функций потерь и метрик оценки качества классификации.

Анализ выявил соответствие поставленной задаче метода «один против одного» (англ. *One-vs-One*), когда метрика, вычисляемая для каждой уникальной пары элементов *S*, затем усредняется. В качестве основной меры оценки качества классификации выбрана площадь AUC_{ovo} под *ROC*-кривой.

Разработанный метод основан на последовательном нелинейном многозначном отображении входных атрибутов в выходные целевые атрибуты (классовые метки). Вектор входных атрибутов $\hat{A}(\hat{n},)$ кодируется в вектор установленной размерности $\overline{NET}_0 = (net_1, net_2, ..., net_{64})$ нелинейной функцией отображения. $\overline{NET}_0$ дублируется *Tree* раз (*Tree* соответствует количеству целевых атрибутов Ξ), и каждый дубликат подвергается последовательным нелинейным отображениям, сокращающим размерность $\overline{NET}_0$ до 1. В результате преобразований *Tree* размерность векторов сокращаются до Ξ скаляров. Ξ скаляров являются результатом работы предложенного метода, и объединяются в множество целевых атрибутов.

На основе предложенного метода разработан алгоритм многозначной классификации (АМК) целевых атрибутов КА работающий в двух режимах: рабочем (АК-1) и режиме корректировки весовых коэффициентов (АК-2).

Режим АК-1 отображает входной вектор значений атрибутов КС в вектор выходных целевых атрибутов.

Режим АК-1 включает четыре шага и принимает в качестве входных данных:

- строку данных $\hat{A}(\hat{n},)$, которую необходимо классифицировать;
- размерность пространства целевых атрибутов *Tree*;
- векторы весовых коэффициентов АМК:

$$\Theta^{(1)} = \begin{pmatrix} (\vec{W}_{1,0}, w_{смещ.\ 1,0}) \\ (\vec{W}_{2,0}, w_{смещ.\ 2,0}) \\ ... \\ (\vec{W}_{64,0}, w_{смещ.\ 64,0}) \end{pmatrix};\ \Theta^{(tree)}(,1) = \begin{pmatrix} (\vec{W}_{1,1}, w_{смещ.\ 1,1}) \\ (\vec{W}_{2,1}, w_{смещ.\ 2,1}) \\ ... \\ (\vec{W}_{32,1}, w_{смещ.\ 32,1}) \end{pmatrix};\ \Theta^{(tree)}(,2) = \begin{pmatrix} (\vec{W}_{1,2}, w_{смещ.\ 1,2}) \\ (\vec{W}_{2,2}, w_{смещ.\ 2,2}) \\ ... \\ (\vec{W}_{16,2}, w_{смещ.\ 16,2}) \end{pmatrix};$$

$$\Theta^{(tree)}(,3) = \left((\vec{W}_{1,3}, w_{смещ.\ 1,3})\right);\ tree = \overline{1, Tree}.$$

Векторы функций отображения АМК:

$$NN^{(1)} = \begin{pmatrix} AF_1^{\ ReLU}(net_1) \\ AF_2^{\ ReLU}(net_2) \\ ... \\ AF_{64}^{\ ReLU}(net_{64}) \end{pmatrix};\ NN^{(tree)}(,1) = \begin{pmatrix} AF_{1,1}^{\ ReLU}(net_{1,1}) \\ AF_{2,1}^{\ ReLU}(net_{2,1}) \\ ... \\ AF_{32,1}^{\ ReLU}(net_{32,1}) \end{pmatrix};\ NN^{(tree)}(,2) = \begin{pmatrix} AF_{1,2}^{\ ReLU}(net_{1,2}) \\ AF_{2,2}^{\ ReLU}(net_{2,2}) \\ ... \\ AF_{16,2}^{\ ReLU}(net_{16,2}) \end{pmatrix};$$

$$NN^{(tree)}(,3) = \left(AF_{1,3}^{\ sigm}(net_{1,3})\right);\ tree = \overline{1, Tree}.$$

Размерность каждого вектора весовых коэффициентов и векторов функций отображения АМК определяется экспериментально.

<u>Алгоритм АК-1 - рабочий режим АМК.</u>

1. *Вычислить скалярное произведение* $\hat{A}(\hat{n},)$ *и каждого вектора весовых коэффициентов в* $\Theta^{(1)}$*, получив* $\overline{NET}_0 = (net_1, net_2, ..., net_{64})$.

2. Отобразить $\overline{NET}_0$ *по каждой функции в* $NN^{(1)}$ - $\overline{Out}_0$.

3. Для tree-го целевого атрибута, $tree = \overline{1, Tree}$*:*

3.1. Вычислить скалярное произведение $\overline{Out}_0$ *и каждого вектора весовых коэффициентов, входящих в* $\Theta^{(tree)}(,1)$ - $\overline{NET}_1^{(tree)}$.

3.2. Отобразить $\overline{NET}_1^{(tree)}$ *по каждой функции* $NN^{(tree)}(,1)$ - $\overline{Out}_1^{(tree)}$.

3.3. Нормализовать $\overline{Out_1^{(tree)}} \to \overline{nOut_1^{(tree)}}$ *и заменить нулями 25% выбранных случайно элементов полученного вектора.*

3.4. Вычислить скалярное произведение $\overline{nOut_1^{(tree)}}$ *и каждого вектора весовых коэффициентов, входящим в* $\boldsymbol{\Theta}^{(tree)}(\,,2)$ *-* $\overline{NET_2^{(tree)}}$.

3.5. Отобразить $\overline{NET_2^{(tree)}}$ *по каждой функции* $\boldsymbol{NN}^{(tree)}(\,,2)$ *-* $\overline{Out_2^{(tree)}}$.

3.6. Нормализовать $\overline{Out_2^{(tree)}} \to \overline{nOut_2^{(tree)}}$ *и заменить нулями 15% выбранных случайно элементов полученного вектора.*

3.7. Вычислить скалярное произведение $\overline{nOut_2^{(tree)}}$ *и вектора весовых коэффициентов, входящим в* $\boldsymbol{\Theta}^{(tree)}(\,,3)$ *-* $\overline{NET_3^{(tree)}}$.

3.8. Отобразить $\overline{NET_3^{(tree)}}$ *по единственной функции* $\boldsymbol{NN}^{(tree)}(\,,3)$ *-* $\overline{Out_3^{(tree)}}$, *округлить-* $round(\overline{Out_3^{(tree)}})$.

4. Выдать $\hat{\boldsymbol{L}}(\hat{n},) = \{round(\overline{Out_3^{(tree)}})\}$ *в качестве результата многозначной классификации.*

Алгоритм АК-2 включает 3 шага и принимает в качестве входных данных:

- *набор исходных данных* $\boldsymbol{D}_{NM}$;
- *размерность пространства целевых атрибутов Tree;*
- *количество эпох обучения Epoch;*
- *функция потерь lossFunc;*
- *векторы весовых коэффициентов АМК* $\boldsymbol{\Theta}^{(1)}, \boldsymbol{\Theta}^{(tree)}(\,,1), \boldsymbol{\Theta}^{(tree)}(\,,2), \boldsymbol{\Theta}^{(tree)}(\,,3)$; $\boldsymbol{NN}^{(1)}, \boldsymbol{NN}^{(tree)}(\,,1), \boldsymbol{NN}^{(tree)}(\,,2), \boldsymbol{NN}^{(tree)}(\,,3)$.

Режим АК-2 предназначен для корректировки весовых коэффициентов методом обратного распространения ошибки.

Алгоритм АК-2 – режим корректировки весовых коэффициентов АМК.

1. Инициализировать:

- размерность векторов, входящих в $\boldsymbol{\Theta}^{(1)}$ ($\vec{W}_{1,0}$, …, $\vec{W}_{64,0}$,), *размерностью атрибутного пространства входных данных* $\boldsymbol{D}_{NM}$;

- начальные значения каждого элемента векторов весовых коэффициентов $\boldsymbol{\Theta}^{(1)}, \boldsymbol{\Theta}^{(tree)}(\,,1), \boldsymbol{\Theta}^{(tree)}(\,,2), \boldsymbol{\Theta}^{(tree)}(\,,3)$ *случайными числами в диапазоне от -1 до 1.*

- правила преобразования атрибутов в выходные значения $\boldsymbol{NN}^{(1)}$, $\boldsymbol{NN}^{(tree)}(\,,1)$, $\boldsymbol{NN}^{(tree)}(\,,2)$ *функцией ReLU;* $\boldsymbol{NN}^{(tree)}(\,,3)$ *- сигмоидальной функцией sigm.*

2. Для каждой $epoch = \overline{1, Epoch}$, *для каждого* $n = \overline{1, N}$:

2.1 Преобразовать $\boldsymbol{A}(n,)$ *в* $\hat{\boldsymbol{L}}(\hat{n},)$.

2.2 Вычислить для tree-го целевого атрибута:

- Функцию потерь для соответствующего целевого атрибута $lossFunc^{(tree)}$.

- Градиент весовых коэффициентов между $lossFunc^{(tree)}$ *и* $\boldsymbol{NN}^{(tree)}(\,,3)$ *-* $Dif_1^{(tree)} = lossFunc^{(tree)\prime}(\,,3) \odot \left[\nabla_{lossFunc^{(tree)}} \boldsymbol{NN}^{(tree)}(\,,3)\right]$ *и скорректировать веса* $\boldsymbol{\Theta}^{(tree)}(\,,3)$.

- *Градиент весовых коэффициентов между* $Dif_1^{(tree)}$ *и* $\boldsymbol{NN}^{(tree)}(\,,2)$ $Dif_2^{(tree)} = Dif_1^{(tree)\prime} \odot \left[\nabla_{Dif_1^{(tree)}} \boldsymbol{NN}^{(tree)}(\,,2) \right]$ *и скорректировать веса* $\boldsymbol{\Theta}^{(tree)}(\,,2)$.

- *Градиент весовых коэффициентов между* $Dif_2^{(tree)}$ *и* $\boldsymbol{NN}^{(tree)}(\,,1)$ $Dif_3^{(tree)} = Dif_2^{(tree)\prime} \odot \left[\nabla_{Dif_2^{(tree)}} \boldsymbol{NN}^{(tree)}(\,,1) \right]$ *и скорректировать веса* $\boldsymbol{\Theta}^{(tree)}(\,,1)$.

- *Градиент весовых коэффициентов между всеми* $Dif_3^{(tree)}$ *и* $\boldsymbol{NN}^{(1)}$ $Dif_4^{(1)} = \prod_{tree} Dif_3^{(tree)\prime} \odot \left[\nabla_{Dif_3^{(tree)}} \boldsymbol{NN}^{(1)} \right]$ *и скорректировать веса* $\boldsymbol{\Theta}^{(1)}$.

3. По достижению epoch корректировок, завершить корректировку весов. Выдать $\boldsymbol{\Theta}^{(1)}$, $\boldsymbol{\Theta}^{(tree)}(\,,1)$, $\boldsymbol{\Theta}^{(tree)}(\,,2)$, $\boldsymbol{\Theta}^{(tree)}(\,,3)$ *в качестве результата.*

Учет нелинейных связей как между целевыми атрибутами на выходе и значениями атрибутов на входе АК происходит на шаге 2.2 алгоритма АК-2. Градиенты весовых коэффициентов, поступивших с каждой *tree*-той «ветви» АК, объединяются и воздействуют на весовые коэффициенты $\boldsymbol{\Theta}^{(1)}$, преобразующие $\boldsymbol{A}(n,\,)$ в $\overline{NET}_0$.

На базе АМК и режимов его работы (АК-1, АК-2) разработана и программно реализована архитектура ИНС на языке python с использованием библиотек *keras* и *tensor flow*. Архитектура ИНС представленная на рис. 1. Разработанный АМК решает задачу многозначной классификации КА за счет присвоения $\hat{\boldsymbol{A}}(\hat{n},\,)$ вектора-строки классовых меток $\hat{\boldsymbol{L}}(\hat{n},\,)$.

Достоинством разработанного алгоритма и соответствующей архитектуры ИНС является многозначная логика их работы, а также высокая обобщающая способность (англ. *generalization ability, generalization performance*).

Оценка точности разработанного алгоритма на базе ИНС с множественным выходом осуществлялась с использованием множества метрик оценки точности классификации $\Phi = \{accuracy_{ovo}, precision_{ovo}, recall_{ovo}, f_{meanovo}, AUC_{ovo}\}$.

Для количественной оценки точности предложенных модели и алгоритма рассмотрена задача многозначной классификации КА в КС на примере наборов данных *UNSW-NB15*, *SR-BH* 2020.

В рамках апробации АМК на наборе данных *UNSW-NB15* решалась задача многозначной классификации 10 видов КА. Исходный набор содержал 257.674 записей по 10 видам КА. Выполнялась предварительная обработка исходных данных при помощи алгоритма поиска дубликатов для выявления многозначных записей.

Проводилось сравнение АК-1 в виде разработанной архитектуры ИНС в условиях изменяющейся атрибутной размерности со стандартными алгоритмами классификации. Рассматривалось 3 случая: исходные данные не сокращались; проводилось сокращение до 5 и 10 наиболее значимых атрибутов.

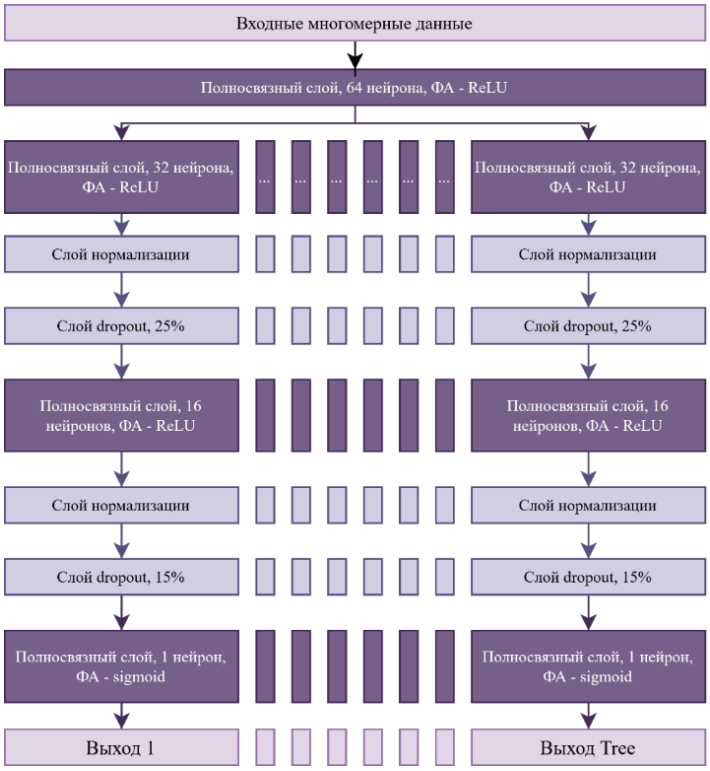

Рисунок 1 – Общий вид разработанной архитектуры ИНС с несколькими выходами

Разделение на обучающую и тестовую выборки проводилось в пропорции 2:1. Записи равномерно перемешивались и нормировались по всему массиву данных.

В эксперименте участвовали следующие алгоритмы МО: *Decision Tree; Extra Tree; ансамблевая реализация Extra Trees; Random Forest; разработанная архитектура ИНС; ИНС типа «многослойный персептрон».* В рамках эксперимента проведено обучение ИНС на первых 120 эпохах. Сравнение реализаций алгоритма АК-1 с аналогами, приведено в табл. 1.

Как видно из таблицы, разработанный алгоритм АК-1 имеет значительное преимущество перед большинством представленных алгоритмов. Выигрыш АК-1 перед однозначными алгоритмами по метрике Φ^{it} составил 16%, по метрике AUC_{ovo} - 13%.

Анализ таблицы позволяет заключить, что предложенная модель табличного представления профилей функционирования КС, учитывающая многозначность целевых атрибутов, связанных с реализацией КА, повышает точность классификации, в среднем, на 5% по интегральной метрике Φ^{it} в сравнении однозначным представлением, для всех алгоритмов МО, участвующих в сравнении.

Результаты работы АК-1 сопоставимы по точности с многозначной реализацией алгоритмов *Random Forest* (выигрыш составил 3%) и бустинговых алгоритмов *Extra Trees, Extra Tree* (выигрыш до 13%).

Апробация разработанного АМК в задаче классификации КА проводилась на наборе данных *SR-BH* 2020, состоящем из веб-запросов. Набор содержал 918.176 записей по 14 видам web-атак, включая web-атаки, направленные на вывод КС из

строя («отказ в обслуживании»); сбор информации о КС; фаззинг. Web-атаки проводились на сервер на базе Wordpress, установленным на виртуальной машине с доступом в сеть Интернет.

Таблица 1 – Сравнение алгоритмов по метрикам Φ^{it} и AUC_{ovo}

Алгоритм	***Все атрибуты КС***		***10 наиболее значимых атрибутов КС***		***5 наиболее значимых атрибутов КС***	
	AUC_{ovo}	Φ^{it}	AUC_{ovo}	Φ^{it}	AUC_{ovo}	Φ^{it}
ИНС с множественным выходом (120 эпох, <**мзн.**>)	**0,95**	**4,51**	**0,93**	**4,38**	**0,90**	**4,04**
ИНС без множественного выхода (120 эпох, <**мзн.**>)	0,93	4,41	0,91	4,25	0,88	3,94
DecisionTree <**мзн.**, **модель**>	0,93	4,39	0,92	4,37	0,87	3,90
ExtraTree <**мзн.**, **модель**>	0,92	4,11	0,89	4,11	0,88	3,86
ExtraTrees <**мзн.**, **модель**>	0,91	4,07	0,90	4,23	0,88	3,89
RandomForest <**мзн., модель**>	0,89	3,99	0,89	3,95	0,87	3,87
DecisionTree <**мзн. игнор.**>	0,84	3,87	0,83	3,83	0,76	3,41
ExtraTree <**мзн. игнор.**>	0,82	3,36	0,75	3,38	0,74	3,10
ExtraTrees <**мзн. игнор.**>	0,78	3,59	0,78	3,58	0,75	3,22
RandomForest **мзн. игнор.**>	0,77	3,82	0,83	3,86	0,76	3,45

Набор данных был предварительно обработан, путем разделения на обучающую и тестовую выборки в пропорции 2:1. Записи равномерно перемешивались и нормировались по всему массиву данных.

Эксперимент заключался в сравнении (в одинаковых условиях) трех методов при работе с многозначными данными (целевыми столбцами) - <мзн.>; <мзн. преобр.>; <мзн. игнор.> для разработанного алгоритма АК-1.

В рамках эксперимента проведено обучение ИНС на первых 120 эпохах. На рис. 2 представлена зависимость AUC_{ovo} от эпохи обучения АК-1.

Маркерами помечено три метода – метод №1 (мзн.), метод №2 (мзн. преобр.) и метод №3 (мзн. игнор.). Для каждого графика нанесена аппроксимирующая линия тренда (тип аппроксимации – степенной, 3-й степени).

Сравнение реализаций алгоритма АК-1 с известными аналогами, основанными на ансамблях (*CatBoost*, *LightGBM*) и разных методах учета многозначности, приведено в табл. 2.

Из табл. 2 видно, что предложенный алгоритм АК-1 и соответствующая ему архитектура ИНС с множественным выходом на 5 … 7% точнее по метрике AUC_{ovo} по сравнению с известными аналогами и до 9% точнее по интегральной метрике Φ^{it}. Наблюдаемый эффект обеспечивается за счет учета нелинейных связей между классовыми метками, а также за счет наличия полносвязного 1-го слоя $\Theta^{(1)}$, преобразующего $A(n,)$ в $\overline{NET_0}$.

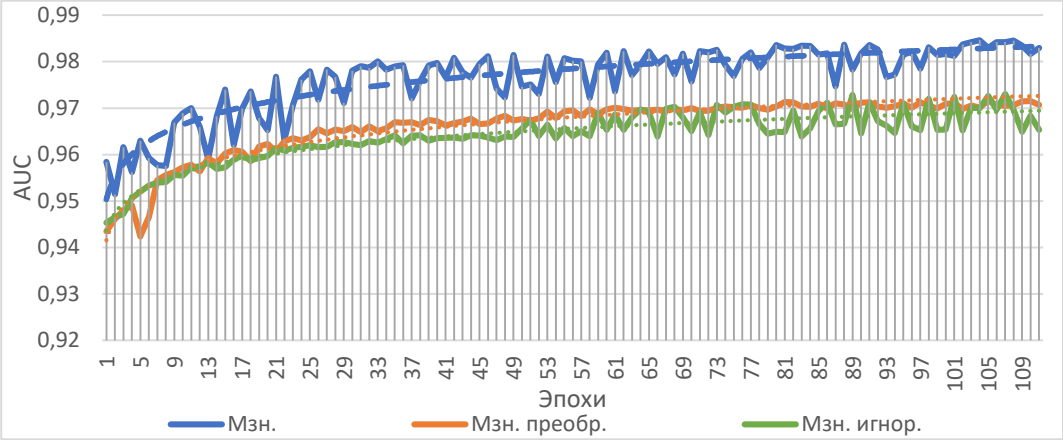

Рисунок 2 – Сравнение AUC_{ovo} от эпохи обучения для трех методов: <мзн.>; <мзн. преобр.>; <мзн. игнор.>.

Таблица 2 – Сравнение алгоритмов по метрикам Φ^{it} и AUC_{ovo} для трех методов: <мзн.>; <мзн. преобр.>; <мзн. игнор.>.

Алгоритм	AUC_{ovo}	Φ^{it}
ИНС с множественным выходом (120 эпох, <**мзн.**>)	**0,983**	**4,86**
ИНС с множественным выходом (120 эпох, <**мзн. преобр.**>)	0,971	**4,76**
ИНС с множественным выходом (120 эпох, <**мзн. игнор.**>)	0,965	**4,73**
CatBoost с множественным выходом	0,913	4,47
LightGBM с множественным выходом	0,911	4,46
Одноступенчатый Clas.Chain LightGBM, учитывающий многозначность методом цепочек классификации	0,904	4,45
Одноступенчатый CatBoost, учитывающий многозначность методом цепочек классификации	0,902	4,39
Одноступенчатый CatBoost, учитывающий многозначность методом бинарной релевантности	0,901	4,39
Двухступенчатый CatBoost, учитывающий многозначность методом бинарной релевантности	0,9	4,38
Одноступенчатый LightGBM, учитывающий многозначность методом бинарной релевантности	0,898	4,37
Двухступенчатый LightGBM, учитывающий многозначность методом бинарной релевантности	0,898	4,36

Достоинствами предложенной архитектуры ИНС и ее основным отличием от имеющихся аналогов являются «ветви», позволяющие учитывать неструктурированные данные и классовые метки более сложной структуры. Выходные слои на окончании «ветвей» могут быть легко преобразованы под решение задач регрессии, кластеризации или их комбинаций.

В четвертой главе описывается созданный ПАК для сбора телеметрии и имитационного моделирования КА, обладающих свойством многозначности целевых атрибутов, маркирующих тип КА. Топология КС *T* исследуемой КС состоит из двух множеств хостов:

$$T = \left\{VH_i; i = \overline{1,I}\right\} \cup \left\{AH_j; j = \overline{1,J}\right\} \cup DAS \cup Router\ , \quad (3)$$

где VH_i – *i*-й хост, имитирующий жертву (далее – атакуемый хост, от англ. *Victim host*); AH_j – *j*-й хост, имитирующий машину злоумышленника (далее – атакующий хост, от англ. *Attack host*), проводящую контролируемую КА на VH_i; *DAS* – сервер агрегации данных (англ. *data aggregation server*), аккумулирующий телеметрию с VH_i и AH_j, а также содержащий конфигуратор КА; *Router* – маршрутизатор (группа маршрутизаторов, или фрагмент сети Интернет), соединяющий множество атакуемых и атакующих хостов.

Вводится перечень КА, которые атакующие хосты AH_j реализуют на атакуемые хосты VH_i - $AL = \{attack_k; k = \overline{1,K}\}$. Каждая КА описывается рядом статичных $attack_k$ и варьируемых $vattack_k$ параметров - $AoI_k : attack_k \cup vattack_k$. Статичные параметры $attack_k$ являются общими для каждой реализации КА: $attack_k : \langle params_{l_k} \rangle; pl_k = \overline{1, PL_k}$. Общее число параметров атаки PL_k и их содержательное наполнение варьируется в зависимости от специфики КА.

Параметры AoI_k задаются либо фиксированными числами, либо законами распределения случайной величины, выбираемыми из библиотеки распределений $FL = \{F_{lf}(\alpha_p); p = \overline{1, P_{lf}}, lf = \overline{1, LenF}\}$, где α_p - p-й параметр lf-го закона распределения.

Взаимодействие между элементами КС *DAS*, VH_i и AH_j (3) осуществляется через программные агенты 1-го и 2-го типов, распространяемые на соответствующие хосты: $PA = \{prograg_{1,i}; i = \overline{1,I}\} \cup \{prograg_{2,j}; j = \overline{1,J}\}$. Программные агенты обоих типов связаны с сервером агрегации данных *DAS*. Программные агенты 1-го типа $prograg_{1,i}$ осуществляют сбор телеметрической информации с атакуемых хостов VH_i и их передачу на *DAS*. Программные агенты 2-го типа - $prograg_{2,j}$ - осуществляют сбор телеметрической информации с атакующих хостов AH_j и реализуют КА, связанные с AH_j, согласно управляющим командам, поступающим с *DAS*.

Взаимодействие между атакующими хостами AH_j и хостами-жертвами VH_i, описывается вектором AoI_k: $\vec{V}_k = (AoI_{kw}; w = \overline{1, W_k})$, где W_k – количество случаев,

когда k-я КА $attack_k$ реализуется в течение эксперимента. Конфигурация воздействий по каждой КА представляется в виде итогового множества $CoA = (\vec{V}_k; k = \overline{1,K})$.

Сценарий использования разработанного ПАК, при известном перечне КА $AL = \{attack_k; k = \overline{1,K}\}$ и их статичных параметрах $\langle params_{l_k} \rangle$, включает:

1. *настройку взаимодействия атакуемых хостов VH_i между собой в рамках эксперимента (определение роли каждого хоста в рамках моделируемого бизнес-процесса; актуализация программного обеспечения; сетевой топологии на хостах и на Router);*

2. *развертывание подсети атакующих хостов AH_j;*

3. *создание «расписания КА»: задание векторов AoI_k по каждой из K КА при помощи разработанного конфигуратора КА;*

4. *инициализацию эксперимента: запуск программных агентов 1-го и 2-го типов для сбора данных; проверка корректности их взаимодействия с DAS; запуск стороннего программного обеспечения для сбора дополнительной телеметрии (при необходимости) на VH_i;*

5. *проведение эксперимента: КА реализуются атакующими хостами AH_j согласно управляющим командам, посылаемым с DAS на программные агенты 2-го типа;*

6. *завершение эксперимента и формирование выходных данных.*

После завершения эксперимента, ПАК формирует многозначный набор данных, содержащий диагностическую информацию о сети, подвергаемой КА из перечня $AL = \{attack_k; k = \overline{1,K}\}$ согласно п. 3 сценария использования. ПАК включает в

себя ряд скриптов, написанных на языке *python*, позволяющих объединить диагностическую информацию, собранную с программных агентов и сторонних сборщиков телеметрии: *Wireshark, MSI Afterburner, Windows Perfmon*.

С использованием разработанного ПАК сформирован новый многозначный набор данных, представляющий собой диагностическую информацию о сети, подвергаемой 6 типам КА, совершаемым параллельно. Перечень КА AL_{exp} включает в себя: 3 атаки типа «отказ в обслуживании», осуществляемые по протоколам *ICMP*, *UDP* и *TCP*; атаку типа «сканирование портов»; атаку типа «сканирование операционной системы» и атаку типа «фаззинг».

Конфигурация ПАК а T_{exp} включает два набора хостов, объединенных по подсетям: $T_{\exp} = \{VH_1\} \cup \{AH_1, AH_2, AH_3, AH_4, AH_5\} \cup DAS \cup Router_{victim} \cup Router_{internet} \cup Router_{reserve}$. Первая подсеть являлась атакуемым сегментом; в подсеть включено два хоста: VH_1 и *DAS*. Вторая подсеть имитирует сегмент внешней сети, с которого совершались контролируемые КА на VH_1. Подсети соединены маршрутизатором $Router_{victim}$, через который осуществлялось взаимодействие между VH_1 и $\{AH_1, AH_2, AH_3, AH_4, AH_5\}$. Сбор данных с подсетей осуществляется с программных агентов $PA_{\exp} = \{prograg_{1,1}\} \cup \{prograg_{2,1}, prograg_{2,2}, prograg_{2,3}, prograg_{2,4}, prograg_{2,5}\}$ через резервный маршрутизатор $Router_{reserve}$. Для имитации рабочего процесса на VH_1 настроен дополнительный сетевой интерфейс, связывающий атакуемый хост с сетью Интернет через $Router_{internet}$. Спецификация ПАК приведены в табл. 3.

Таблица 3 – Спецификация ПАК для проведения эксперимента

Элемент топологии	**Программ-ный агент**	**Назначение**	**Кол-во уникальных наборов параметров КА**	**Кол-во КА, направленных на VH_1**
VH_1	$progag_{1,1}$	Имитация рабочей деятельности жертвы	-	-
DAS	-	Сервер агрегации данных	-	0
AH_1	$progag_{2,1}$	Реализация атак «отказ в доступе» по протоколам TCP, UDP.	2	6
AH_2	$progag_{2,2}$	Реализация атаки «отказ в доступе» по протоколу ISMP.	1	11
AH_3	$progag_{2,3}$	Реализация атаки «сканирование портов в тихом режиме».	1	6
AH_4	$progag_{2,4}$	Реализация атаки «сканирование портов в агрессивном режиме»	1	11
AH_5	$progag_{2,5}$	Реализация атак «сканирование операционной системы», «фаззинг».	1	9

Эксперимент проводился в период апреля 2024 г. с шагом сбора информации 1 секунда. Во время эксперимента на VH_1 совершено 43 контролируемых КА трех типов, включающих в себя: DOS по протоколам ISMP; UDP; TCP; сканирование портов; сканирование ОС; фаззинг сетевого программного обеспечения. Количество записей экспериментальных данных в итоговом наборе составило 263.388 ед. Итоговый набор данных содержал 125 атрибутов метрического типа. Набор данных содержит избыточный набор целевых атрибутов включая информацию о количестве одновременно задействованных в атаке хостов и типах КА. Информация, собранная

с сетевой карты и системных журналов, преобразовывалась в метрические атрибуты посредством вычисления количественных характеристик в рамках окна размером 1 секунда. Доля многозначных записей в данных по количеству хостов, полученных при помощи ПАК в рамках проведенного эксперимента составила 55%; по типу КА - 46%.

Данные, порождаемые ПАК, могут быть использованы для противодействия КА при разработке СОВ, учитывающих многозначность записей и интерпретироваться как средство обнаружения КА в КС. Комплекс порождает реальные данные, соответствующие задачам информационной безопасности. За счет большого количества настраиваемых параметров моделирования КА, возможна «тонкая» настройка распределения классовых меток и соотношения доли однозначных и многозначных записей в данных, формируемых ПАК. Новизна разработанного ПАК заключается автоматизированной параллельной маркировке всех КА, осуществляемых на КС, что позволяет учесть многозначность уже на этапе сбора данных.

В **заключении** отражены основные результаты проведенного диссертационного исследования.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В диссертационном исследовании предложен подход к обеспечению ИБ, заключающийся в повышении точности классификации сетевых атак в КС в условиях многозначности целевых атрибутов, маркирующих их тип.

В главе №1 получены следующие результаты:

1. Проведен анализ специфических особенностей многозначной классификации в задачах обнаружения и классификации КА.

В главе №2 диссертационного исследования получены следующие результаты:

2. Разработана модель табличного представления профилей функционирования КС, учитывающую многозначность целевых атрибутов, связанных с реализацией КА. Алгоритмом поиска полных дубликатов по атрибутному пространству проанализирован ряд наборов данных, связанных с ИБ. По результатам сравнения выявлена многозначность целевых атрибутов во всех анализируемых наборах данных. Доля многозначных целевых атрибутов составила от 1,3% до 86%.

В главе №3 диссертационного исследования получены следующие результаты:

3. Разработан многозначный метод обнаружения и классификации КА в КС, заключающийся в многозначном отображении пространства атрибутов в множество классовых меток, и алгоритм на его основе, выигрыш в точности от которого достигает 16% по интегральной метрике Φ^{it} и 13% по метрике AUC_{ovo} перед однозначными алгоритмами классификации, и до 13% по интегральной метрике Φ^{it}; 6% по метрике AUC_{ovo} перед многозначными алгоритмами классификации.

4. Предложенная модель табличного представления профилей функционирования КС, учитывающая многозначность целевых атрибутов, связанных с реализацией КА, повышает точность классификации, в среднем, на 5% по интегральной метрике $Ф^{it}$ в сравнении однозначным представлением, для всех алгоритмов МО, участвующих в сравнении.

В главе №4 диссертационного исследования получен следующий результат:

5. Разработан и программно реализован ПАК для сбора телеметрии КА в КС, позволяющий в контролируемых условиях автоматизировать сбор телеметрии и имитационного моделирования КА, обладающих свойством многозначности. Доля многозначных записей в данных по количеству хостов, полученных при помощи ПАК в рамках проведенного эксперимента, составляет 55%. Доля многозначных записей в данных по типу КА, составляет 46%.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендуемых ВАК РФ:

1. ***Раковский, Д. И.*** *Влияние проблемы многозначности меток классов системных журналов на защищенность компьютерных сетей / Д. И. Раковский // Наукоемкие технологии в космических исследованиях Земли. – 2023. – Т. 15, № 1. – С. 48-56.*

2. *Шелухин, О. И. Прогнозирование профиля функционирования компьютерной системы на основе многозначных закономерностей / О. И. Шелухин,* ***Д. И. Раковский*** *// Вопросы кибербезопасности. – 2022. – № 6(52). – С. 53-70.*

3. *Шелухин, О. И. Бинарная классификация многоатрибутных размеченных аномальных событий компьютерных систем с помощью алгоритма SVDD / О. И. Шелухин,* ***Д. И. Раковский*** *// Наукоемкие технологии в космических исследованиях Земли. – 2021. – Т. 13, № 2. – С. 74-84.*

4. *Шелухин, О. И. Многозначная классификация меток классов системных журналов компьютерных сетей. Формализация задачи / О. И. Шелухин,* ***Д. И. Раковский*** *// Проблемы информационной безопасности. Компьютерные системы. – 2023. – № 3(56). – С. 154-169.*

5. *Шелухин, О. И. Многозначная классификация компьютерных атак с использованием искусственных нейронных сетей с множественным выходом / О. И. Шелухин,* ***Д. И. Раковский*** *// Труды учебных заведений связи. – 2023. – Т. 9, № 4. – С. 97-113.*

6. *Шелухин, О. И. Многозначная классификация меток классов системных журналов компьютерных сетей. Сравнительный анализ эффективности классификаторов / О. И. Шелухин,* ***Д. И. Раковский*** *// Вопросы кибербезопасности. – 2023. – № 3(55). – С. 62-77.*

7. *Шелухин, О. И. Влияние многозначности баз данных на результаты многоклассовой классификации компьютерных атак / О. И. Шелухин,* ***Д. И. Раковский*** *// Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. – 2023. – № 3. – С. 111-119.*

8. *Шелухин, О. И. Разработка программно-аппаратного комплекса моделирования многозначных компьютерных атак / О. И. Шелухин,* ***Д. И. Раковский*** *// Вопросы кибербезопасности. – 2024. – № 4(62). – С. 116-130.*

*9. Шелухин, О. И. Имитационное моделирование многозначных компьютерных атак / О. И. Шелухин, **Д. И. Раковский**, И. Д. Александров, А. Д. Боков // I-methods. – 2023. – Т. 15, № 4.*

Публикации в научных изданиях, индексируемых в базах Scopus и Web of Science:

10. Sheluhin, O. I. New Algorithm for Predicting the States of a Computer Network Using Multivalued Dependencies / O. I. Sheluhin, A. V. Osin, D. I. Rakovsky // Automatic Control and Computer Sciences. – 2023. – Vol. 57, No. 1. – P. 48-60.

Свидетельства о государственной регистрации программ для ЭВМ:

*11. Программа для ЭВМ 2024618269 Российская Федерация. Программное обеспечение Ива-1. Искусственная нейронная сеть с множественным выходом для решения задач многозначной классификации [текст] / Шелухин, О.И., **Раковский, Д.И. –** № 2024617387 заявл. 10.04.2024; опубл. 10.04.2024.*

*12. Программа для ЭВМ 2022615309 Российская Федерация. Multivalued dependencies prognosis algorithm (MDPA) [текст] / Шелухин, О.И., **Раковский, Д.И. –** № 2022614460 заявл. 22.03.2022; опубл. 30.03.2022.*

*13. Программа для ЭВМ 2022610329 Российская Федерация. Программа для прогнозирования временного ряда на основе многозначных зависимостей [текст] / Шелухин О.И., Осин А.В., **Раковский Д.И.** № 2021682196 заявл. 27.12.2021; опубл. 11.01.2022.*

*14. Программа для ЭВМ 2024665452 Российская Федерация. Программный комплекс для сбора многозначных данных в условиях проведения контролируемых компьютерных атак v.1 [текст] / О. И. Шелухин, **Д. И. Раковский**, И. Д. Александров, А. Д. Боков – № 2024664308 заявл. 24.06.2024; опубл. 02.07.2024.*

*15. Программа для ЭВМ 2020667645 Российская Федерация. Программный модуль для выбора эффективного метода защиты информации с использованием многокритериального анализа [текст] / А. С. Большаков, **Д. И. Раковский** – № 2020666754 заявл. 16.12.2020; опубл. 25.12.2020.*

Публикации в сборниках трудов научно-технических конференций и форумов:

*16. Sheluhin, O. I. Multi-Label Learning in Computer Networks / O. I. Sheluhin, **D. I. Rakovskiy** // Systems of Signals Generating and Processing in the Field of on Board Communications. – 2023. – Vol. 6, No. 1. – P. 412-416.*

*17. **Раковский, Д. И.** Разработка стенда для сбора телеметрии и имитационного моделирования многозначных компьютерных атак / **Д. И. Раковский**, И. Д. Александров, А. Д. Боков // Технологии информационного общества: Сборник трудов XVIII Международной отраслевой научно-технической конференции, Москва, 27–28 февраля 2024 года. – Москва: Московский технический университет связи и информатики, 2024. – С. 125-127.*

*18. Шелухин, О.И. Обнаружение компьютерных атак на основе многозначных закономерностей / Шелухин О.И., **Раковский Д.И.** // Материалы 33-й всероссийской научно-технической конференции «Методы и технические средства обеспечения безопасности информации». – СПб. – 2024. – С. 36 – 38.*

*19. **Раковский, Д.И.** Многозначное прогнозирование профиля функционирования компьютерной сети для обеспечения информационной безопасности / Раковский, Д.И. // сборник статей X Всероссийской научно-технической конференции «Фундаментальные и прикладные аспекты компьютерных технологий и информационной безопасности». – Таганрог. – 2024. – С. 57 – 59.*

*20. **Раковский, Д.И.** Стенд для сбора телеметрии многозначных компьютерных атак / Раковский, Д.И. Александров И.Д., Боков А.Д. // Сборник трудов III Всероссийской*

научно –практической конференции «Теория и Практика Обеспечения Информационной Безопасности». – М. – 2023. – С. 26 – 33.

21. ***Раковский, Д.И.*** *Обнаружение компьютерных атак и предупреждение нарушений функционирования компьютерных сетей на основе многозначных закономерностей / Раковский, Д.И. // Сборник трудов III Всероссийской научной школы-семинара «Современные тенденции развития методов и технологий защиты информации». – М. – 2023. – С. 307-311.*

22. ***Раковский, Д.И.*** *Прогнозирование профиля функционирования компьютерной системы с применением аппарата точечно-множественных отображений / Раковский, Д.И. // Сборник трудов II Всероссийской научно-практической конференции «Теория и Практика Обеспечения Информационной Безопасности». – М. – 2022. – С. 222 – 231.*

23. Шелухин, О.И. Выбор категориальных атрибутов редких аномальных событий компьютерной системы методами символьного анализа / Шелухин О.И., ***Раковский Д.И.*** *// Сборник трудов XV Международной отраслевой научно-технической конференции «Технологии информационного общества». – М. – 2021. – С. 179-181.*

<u>Публикации в других изданиях</u>:

24. ***Rakovskiy, D. I.*** *Analysis of the problem of multivalued of class labels on the security of computer networks / D. I. Rakovskiy // Synchroinfo Journal. – 2022. – Vol. 8, No. 6. – P. 10-17.*

25. Шелухин, О. И. Визуализация аномальных событий при прогнозировании состояний компьютерных систем на основе "исторических данных" / О. И. Шелухин, ***Д. И. Раковский*** *// REDS: Телекоммуникационные устройства и системы. – 2022. – Т. 12, № 2. – С. 53-58.*

26. Шелухин, О. И. Выбор метрических атрибутов редких аномальных событий компьютерной системы методами интеллектуального анализа данных / О. И. Шелухин, ***Д. И. Раковский*** *// T-Comm: Телекоммуникации и транспорт. – 2021. – Т. 15, № 6. – С. 40-47.*

27. Большаков, А. С. Эффективный метод многокритериального анализа в области информационной безопасности / А. С. Большаков, ***Д. И. Раковский*** *// Правовая информатика. – 2020. – № 4. – С. 55-66.*