Имитационное моделирование многозначных компьютерных атак

Шелухин Олег Иванович
доктор технических наук профессор, заведующий кафедрой «Информационная безопасность», Московский технический университет связи и информатики, Москва, Россия, sheluhin@mail.ru

Раковский Дмитрий Игоревич
аспирант, ассистент кафедры «Информационная безопасность», Московский технический университет связи и информатики, Москва, Россия, Prophet_alpha@mail.ru

Александров Илья Дмитриевич
студент, Московский технический университет связи и информатики, Москва, Россия

Боков Александр Дмитриевич
студент, Московский технический университет связи и информатики, Москва, Россия

АННОТАЦИЯ

Введение: Рассмотрена структура разработанного и программно-реализованного стенда для сбора телеметрии и статистические характеристики компьютерных атак в компьютерной сети, с несколькими атакуемыми и атакующими хостами с учетом их многозначности на этапе сбора данных. Для апробации разработанного стенда, была произведена многозначная компьютерная атака на хост-«жертву». В качестве примера были выбраны атаки двух типов – «отказ в обслуживании» и «сканирование операционной системы». Показано, что количество информационно значимых атрибутов для компьютерной атаки «сканирование операционной системы» превышает их количество для компьютерной атаки «отказ в обслуживании», что необходимо учитывать при реализации алгоритмов машинного обучения в задачах обнаружения и классификации атак.

КЛЮЧЕВЫЕ СЛОВА: экспериментальные данные; multilabel classification; сетевая атака; компьютерная атака; исследовательский стенд; информационная безопасность.

Введение

Современные компьютерные атаки (КА) на корпоративные компьютерные сети (КС) характеризуются умением обходить статистические и сигнатурные средства защиты информации [1]. Для обеспечения безопасности таких систем внедряются все более сложные механизмы защиты информации [2]. Наиболее актуальными механизмами обнаружения компьютерных атак являются решения на основе использования алгоритмов машинного обучения (МО) [3–6], как в статическом режиме [7], так и в режиме online [8].

Особенностью алгоритмов МО является необходимость в настройке их параметров, таких как весовые коэффициенты искусственных нейронных сетей, решающих правилах в случае древовидных алгоритмов и т.д. Большая часть известных наборов данных, связанных с информационной безопасностью (например, [9–11]), позиционируются как однозначные, когда каждой записи «исторических данных» ставится в соответствие одна классовая метка.

Вместе с тем , как показывают исследования [12,13], практически во всех известных базах данных (БД) обнаруживаются записи, практически полностью идентичные по всем атрибутам за исключением целевого атрибута (классовых меток), что говорит о многозначности таких наборов данных. Природа происхождения таких записей различна.

В ряде случаев к многозначным записям приводят ограниченные возможности сбора данных - округление метрических переменных под *float16 / float32*; малое количество атрибутов, высокая корреляция между атрибутами, ограниченное множество значений, которое может принимать атрибут категориального типа [14].

Известны КА, направленные на «отравление» наборов данных, пригодных для алгоритмов МО [15]. Одной из разновидностей «отравляющих» КА является инъекция записей с неправильной маркировкой, что впоследствии приводит к ошибкам первого и второго родов.

Значительное количество исследований в области контролируемого обучения посвящено анализу данных с одной единственной меткой, когда обучающие примеры связаны с одной меткой λ из набора непересекающихся меток L [16–18]. Однако иногда обучающие примеры связаны с целым набором меток $Y \subseteq L$. Такие данные называются данными с использованием нескольких меток.

Иная природа многозначности проявляется при одновременной маркировке временных рядов несколькими маркерами событий, например, такими, как возникновение КА; аномалии; появление определенных записей в системных журналах. С ростом количества меток, одновременно проставляемых временному ряду, растет вероятность маркировки одной записи несколькими классовыми метками. В результате такая запись становится многозначной. Примером такого набора данных является многозначный набор SR-BH 2020 [19,20].

Известно несколько методов работы с подобными многозначными данными [21]:

– игнорирование, когда многозначный набор данных считается однозначным;

– дополнительное преобразование многозначного набора данных и приведение его к однозначному виду;

– учет многозначности данных в архитектуре алгоритмов МО [22,23].

Недостатками первого метода являются сопутствие ошибок, связанных с невозможностью одновременной маркировки записи несколькими классовыми метками. Недостатком второго метода является экспоненциальный рост однозначного алфавита комбинаций классовых меток и накладные расходы на преобразование данных [24].

Для реализации третьего подхода требуется создание алгоритмов, учитывающих много-

значность данных. Необходим инструментарий для порождения многозначных наборов «исторических данных» в контролируемых условиях (прогнозируемая доля многозначных классовых меток; возможность задания их количества и т.д.).

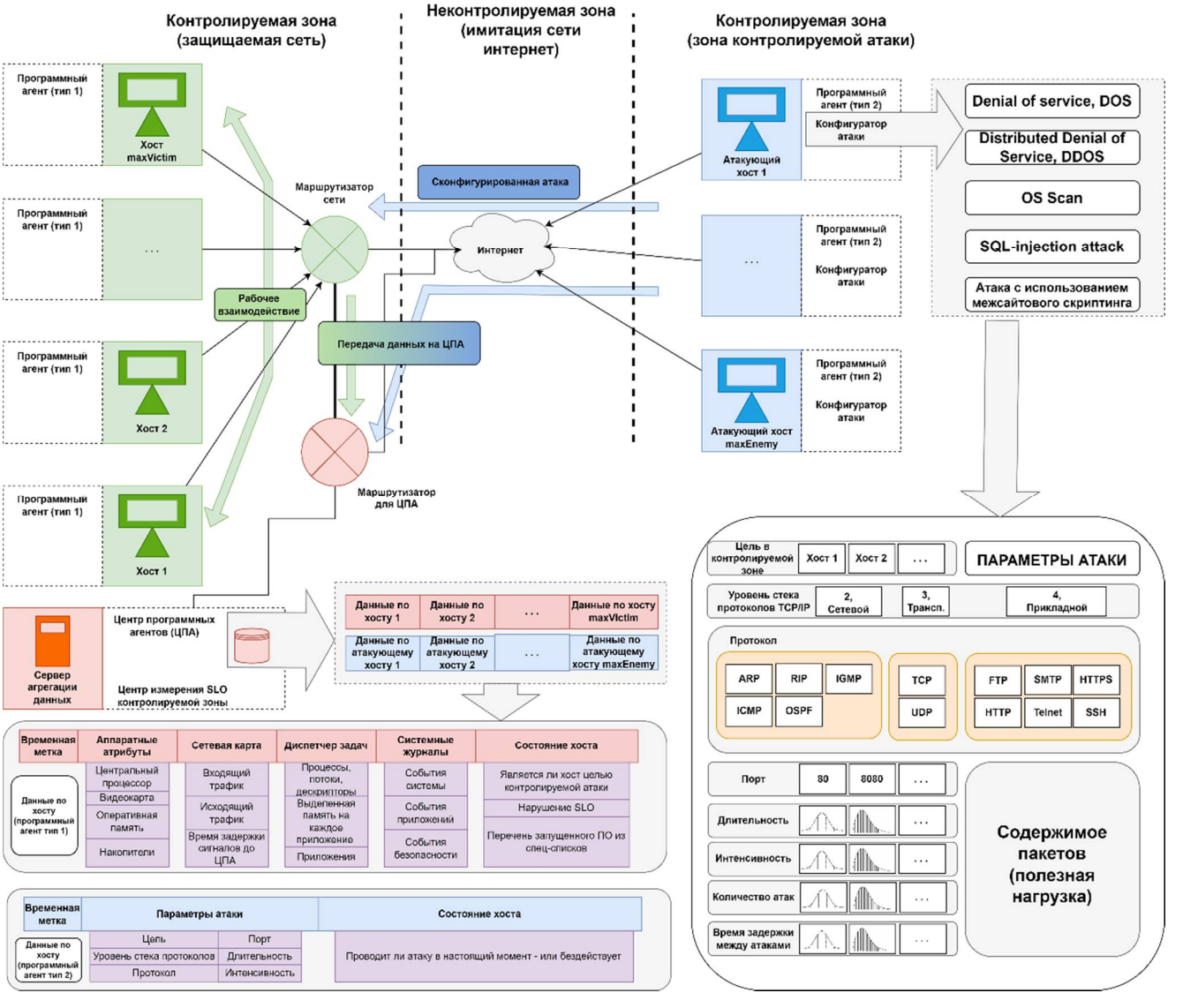

Рис. 1. Структурная и схема функционирования стенда

Учитывая важность проблемы многозначности в экспериментальных данных в области информационной безопасности, необходимо решить задачу разработки и программной реализации стенда для исследования подобных многозначных наборов данных возникающих при воздействии нескольких КА на КС.

Целью работы является анализ результатов имитационного моделирования многозначных КА, полученных с помощью реализованного экспериментального стенда для сбора многозначной телеметрии КС в условиях одновременного воздействия нескольких контролируемых КА [25,26].

Разработанный стенд представляет собой совокупность трех компонентов: программные агенты первого (ПА-1) и второго (ПА-2) типов, конфигуратор компьютерных атак (ККА) и база данных (БД). ККА и БД распложены на сервере агрегации данных (САД).

ПА-1 осуществляет сбор и передачу данных с хостов, на которые совершается контролируемая КА, в БД, расположенную на САД. Каждый установленный интервал времени ПА-1, находясь на атакуемом хосте, агрегирует информацию с аппаратных датчиков и низкоуровневого API (аппаратные атрибуты); извлекает информацию из системных журналов (логов) операционной системы; собирает информацию с сетевой карты и снимает перечень процессов, запущенных в диспетчере задач. За этот временной интервал агент успевает выполнить системный запрос, создать строку данных и передать ее на сервер для агрегации информации. ПА-1, реализованный на языке Python, расположен на хосте и работает под управлением операционной системы Windows.

Визуализация механизма работы стенда представлена на рисунке 1.

Для последующей обработки полученных данных необходимо их разметить. С этой целью ПА-2 собирает данные о реализуемых атаках (номер атакуемого хоста, частота атаки, ее длина и плотность), которые используются для разметки данных. Если атака осуществлялась, то в результирующей базе данных на момент атаки добавляется тип данной атаки, а также метка наличия атаки: символ 1- если атака была и символ 0 - если атаки не было.

Каждую секунду в файл формата *DB Databased*, используя язык SQL запросов сервер по агрегации данных записывает строки, которые ему отправляет рабочая станция, управляемая ПА 1-го типа.

На заключительном этапе все таблицы, ассоциирующиеся с хостами, объединяются в одну. К заголовкам названий столбцов приписывается номер хоста и данные записываются по времени и хостам в единую базу данных как это показано на рис. 2. На рисунке 2 зелеными стрелками обозначена передача данных на файловый сервер, а красными – вектор компьютерной атаки.

САД (сервер на рис. 2), получая информацию от клиентов, идентифицирует номер хоста и сохраняет полученные данные в соответствующей таблице.

Для удобства последующего анализа данных с помощью алгоритмов МО при экспорте информации с САД предусмотрено сведение всех БД в общую (итоговую) таблицу. Пример структуры итоговой таблицы атакуемого хоста КС, подвергаемой КА двух типов – «отказ в обслуживании» (1) и «сканирование операционной системы» (2) - приведен в таблице 1. Таблица содержит атрибуты, связанные с аппаратными компонентами атакуемого хоста КС и данными, снимаемыми с сетевой карты.

Из таблицы 1 видно, что КА могут совершаться одновременно. Представленная таблица адаптирована для обработки существующими многозначными алгоритмами МО. С этой целью целевые атрибуты КА представлены методом бинарной релевантности [27].

Разработанный стенд позволяет формировать итоговые таблицы различных вариаций среди которых [25,26]:

– Извлечение информации по конкретному хосту КС. В этом случае из БД извлекаются атрибуты по выбранному хосту, а также ряд целевых меток, связанных с проведением КА.

– Извлечение информации по группе хостов КС. В этом случае из БД извлекаются атрибуты по каждому выбранному хосту, затем объединяются в единую, «широкую», таблицу

посредством операции конкатенации.

В конце таблицы дописывается ряд целевых меток, связанных с проведением КА.

При формировании табличной структуры может быть указан тип представления категориальной информации (в качестве отдельных атрибутов таблицы или в качестве отдельных файлов) и номера хостов, атрибуты которых участвуют в формировании таблицы.

Таблица 1

Итоговая таблица по всем БД

Date	**Time**	**GPU_usage**	**…**	**Code_exec_TIME**	**Ip**	**Normal**	**Атака$_{(1)}$**	**Атака$_{(2)}$**
29-09-2023	12:30:00	0.0	…	1.215	10.10.10.85	1	0	0
29-09-2023	12:30:03	0.4	…	1.203	10.10.10.85	0	1	0
29-09-2023	12:30:06	0.0	…	1.256	10.10.10.85	0	1	1
29-09-2023	12:30:09	0.4	…	1.216	10.10.10.85	0	1	1
29-09-2023	12:30:12	0.0	…	1.190	10.10.10.85	0	1	1

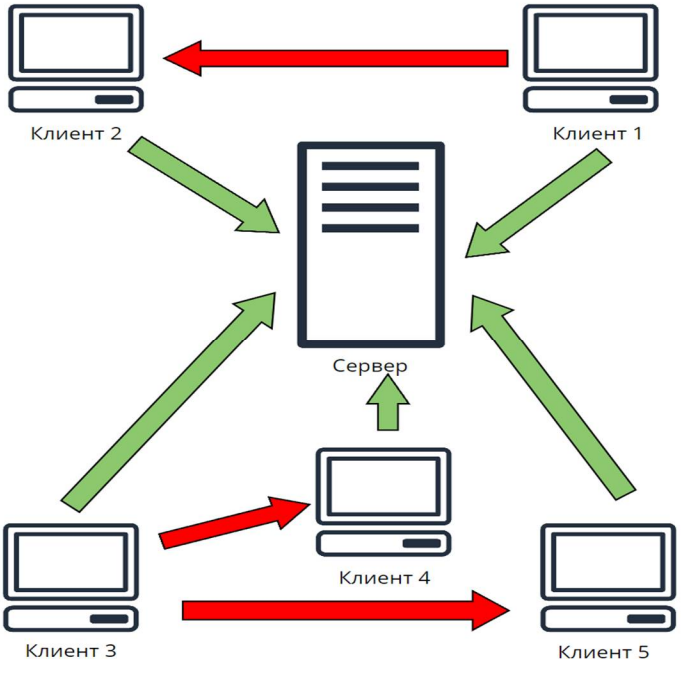

Рис. 2. Обобщённая схема формирования единой БД

Особенности имитационного моделирования многозначных КА

Для апробации разработанного стенда, была произведена многозначная КА на один хост-жертву. Выбраны атаки двух типов – «отказ в обслуживании» и «сканирование операционной системы». Атака типа «отказ в обслуживании» реализована посредством утилиты *hping3*, «сканирование операционной системы» – *nmap*. Интервалы проведения КА сконфигурированы таким образом, чтобы часть атак различных категорий происходила одновременно.

Перечень всех атрибутов итоговой сводной таблицы, имеющей формат согласно табл.1, приведен в табл. 2.

Таблица 2

Перечень и описание атрибутов БД

Индекс	**Атрибут БД**	**Краткое описание**
0	Date	День и месяц снятия атрибутов
1	Time	Время снятия атрибутов
2	CPU_usage	Процент использования CPU за последнюю секунду
3	Available_RAM_memory	Доступная оперативная память для процессов
4	RAM_memory_used	Процент используемой оперативной памяти
5	RAM_Used	Используемая оперативная память в ГБ
6	Free_RAM	Свободная оперативная память
7	System_swap_usage	Процент использования системной подкачки
8	Read_count	Количество операций чтения дискового ввода-вывода
9	Write_count	Количество записей дискового ввода-вывода
10	Read_bytes	Количество прочитанных байтов дискового ввода-вывода
11	Write_bytes	Количество записанных байтов дискового ввода-вывода
12	Read_time	Время, затраченное на чтение с диска (в миллисекундах)
13	Write_time	Время, затраченное на запись на диск (в миллисекундах)
14	bytes_sent	Количество отправленных байтов
15	bytes_recv	Количество полученных байтов
16	packets_sent	Количество отправленных пакетов
17	packets_recv	Количество полученных пакетов
18	Errin	Общее количество ошибок при получении пакетов
19	Errout	Общее количество ошибок при отправке
20	Dropin	Общее количество входящих пакетов, которые были отброшены
21	Dropou	Общее количество исходящих пакетов, которые были отброшены (всегда 0 на macOS и BSD)
22	Ctx_switches	Количество переключений контекста с момента загрузки ЦП. Контекстное переключение происходит, Когда один поток выполнения сменяется другим
23	Interrupts	Количество прерываний с момента загрузки ЦП
24	Syscalls	Количество системных вызовов с момента загрузки ЦП
25	User_time	Время, затрачиваемое на выполнение задач в пользовательском режиме
26	System_time	Время, затрачиваемое процессами, выполняющимися в режиме ядра
27	Idle	Время, потраченное на бездействие ЦП

28	Interrupt_time	Время, затрачиваемое на обслуживание аппаратных прерываний (аналогично “irq” в UNIX)
29	DPS_time	Время, затрачиваемое на обслуживание вызовов отложенных процедур (DPC)
30	Processor_frequency	Частота процессора
31	Processor_frequency_MIN	Минимальная зафиксированная частота процессора за период фиксации данных
32	Processor_frequency_MAX	Максимальная зафиксированная частота процессора за период фиксации данных
33	Code_exec_time	Скорость обработки низкоуровневого кода процессором
34	Ip	Уникальный идентификатор устройства
35	Token	Затокенизированные данные диспетчера задач
36	Token_logs	Затокенизированные данные системных журналов
Классовые метки		
37	Отсутствие КА	Бинарная величина, маркер отсутствия КА
38	КА типа «отказ в обслуживании»	Бинарная величина, маркер наличия КА типа «отказ в обслуживании»
39	КА типа «сканирование операционной системы»	Бинарная величина, маркер наличия КА типа «сканирование операционной системы»

Файловый сервер агрегации данных, получая информацию от ПА-1, идентифицирует номер хоста и сохраняет полученные данные в соответствующей таблице. По завершении приема строк начинается процесс формирования конечной таблицы. Информация о всех рабочих станциях записывается последовательно, сохраняя временной порядок, что обеспечивает более удобную и точную обработку данных.

Многозначность собранных данных, в первую очередь, обусловливается контролируемым проведением двух КА одновременно. Данные, поступающие в САД в момент одновременного проведения КА, маркируются как меткой «КА типа «отказ в обслуживании» (№38 в табл. 2), так и меткой «КА типа «сканирование операционной системы» (№39 в табл. 2).

Результаты имитационного моделирования многозначных КА

Эксперимент был проведен в период с 29.09.2023 12:30:00 по 04.10.2023 11:39:57. За это время собрано 61,806 записей по 36 атрибутам.

В течение эксперимента зарегистрировано 49 атак типа «отказ в обслуживании» (совокупное количество записей в экспериментальных данных – 3877 шт.) и 189 атак типа «сканирование операционной системы» (совокупное количество записей в экспериментальных данных – 36081 шт.). Распределение классовых меток приведено на рисунке 2. Визуализация экспериментальных записей представлена на рисунке 3.

Количество многозначных записей в данных составляет 2709 шт. (4,3%). Доля многозначных записей КА «Отказ в обслуживании» составляет 7%. Доля многозначных записей КА «Сканирование операционной системы» составляет 23%.

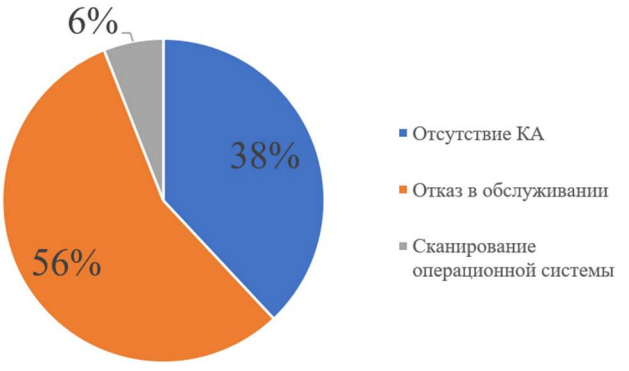

Рис. 3. Распределение КА в собранных экспериментальных данных по типу

Высокая волатильность многозначных записей по разным категориям КА обусловлено дисбалансом классов. Дисбаланс классов вызван механизмом реализации КА: атака типа «отказ в обслуживании» считается завершенной по истечение установленного промежутка времени (как правило – минуты, часы) или при успешном вводе КС в состояние «отказ в обслуживании»; атака типа «сканирование операционной системы» считается успешной при получении информации о КС (как правило, секунды).

Дисбаланс классов (в контексте многозначной классификации) актуализирует выбор метрик оценки качества классификации, поскольку в контексте решения задач информационной безопасности последствия от реализации разных КА различаются на порядки. Следовательно, стоимость ложноотрицательных и ложноположительных ошибок для каждой КА различается. Как правило, наиболее разрушительные КА являются комплексом действий (комбинацией более простых КА, разнесенных во времени); их совокупная доля в общем массиве данных незначительна, а урон, наносимый КС – максимальный. Учет многозначности КА в архитектуре алгоритмов МО повышает эффективность классификации КА за счет учета случаев одновременной реализации нескольких КА.

Отметим, что доля многозначных записей в множестве аномальных записей возрастает с ростом количества целевых атрибутов, поскольку многозначной считается любая запись, ассоциированная с более, чем одной КА.

Для разработки или выбора алгоритмов МО, эффективно решающих задачу многозначной классификации, необходимо оценить информативность атрибутов КС (табл. 2). Воспользовавшись методом вычисления индекса Джини [28,29], оценим информативность каждого типа КА, построив гистограмму распределения важности каждого атрибута КС из таблице 2. На рисунке 4 представлены гистограммы важности атрибутов для КА «отказ в обслуживании» (4а) и «сканирование операционной системы» (4.б).

Анализ представленных данных на рисунке 4а позволяет сделать вывод, что наиболее информативными являются первые 8 атрибутов. В отличие от КА «отказ в обслуживании», количество значимых атрибутов КА «сканирование операционной системы» (рис. 4б) значительно больше и достигает 20, что позволяет сделать вывод о том, что использование алгоритмов машинного обучения для анализа данных атаки «сканирование операционной системы» может оказаться более эффективным. Большее число значимых атрибутов предоставляет более обширную информацию, что может способствовать более точному обнаружению и

классификации данного типа атак.

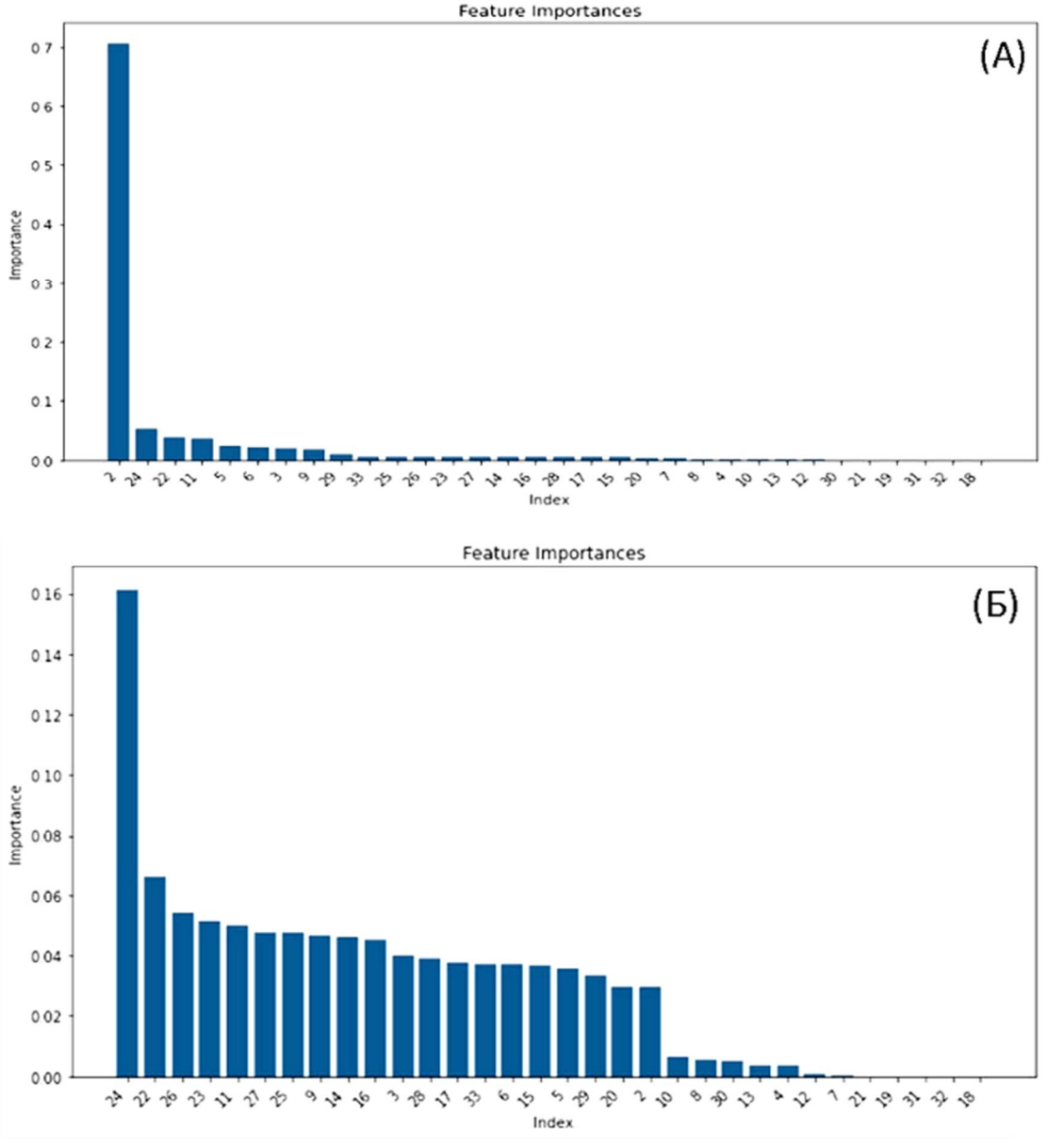

Рис. 4. Гистограммы важности атрибутов КА: 4.а – типа «отказ в обслуживании»; 4.б – типа «сканирование операционной системы»

Для сравнения распределений информационной значимости атрибутов для двух КА. объединим их на одной гистограмме, как это показано на рисунке 5.

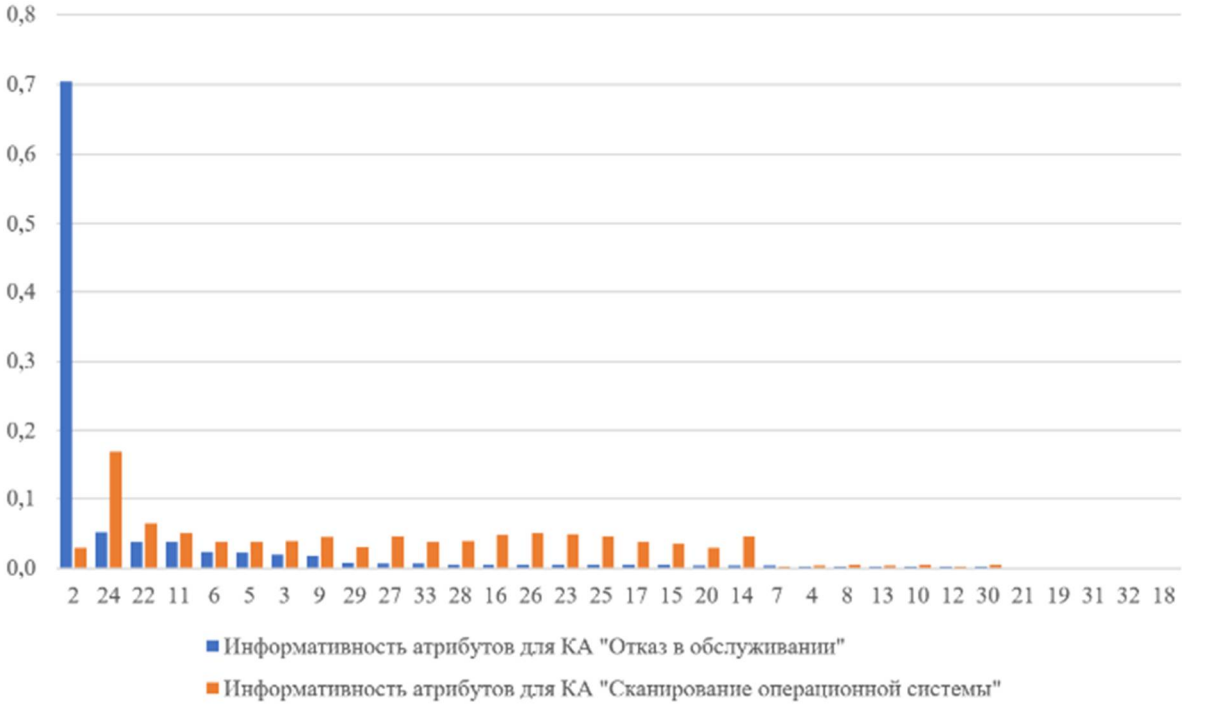

Рис. 5. Объединенная гистограмма распределения информационной значимости атрибутов для двух КА

При большом количестве комбинаций целевых атрибутов анализ становится затруднительным, что актуализирует применение и разработку новых методов анализа многозначных экспериментальных данных в задаче отбора атрибутов под задачи МО.

Заключение

Новизна разработанного стенда заключается автоматизированной параллельной маркировке всех КА в КС, что позволяет учесть многозначность на этапе сбора данных.

При проведении эксперимента зарегистрировано 49 атак типа «отказ в обслуживании» (совокупное количество записей в экспериментальных данных – 3877 шт.) и 189 атак типа «сканирование операционной системы» (совокупное количество записей в экспериментальных данных – 36081 шт.

Количество многозначных записей в данных составляет 2709 шт. (4,3%). Доля многозначных записей КА «Отказ в обслуживании» составляет 7%. Доля многозначных записей КА «Сканирование операционной системы» составляет 23%.

Показано, что поскольку многозначной считается любая запись, ассоциированная с более, чем одной КА доля многозначных записей в множестве аномальных записей возрастает с ростом количества целевых атрибутов.

Проведенный анализ показал, что высокая волатильность многозначных записей по разным категориям КА обусловлена дисбалансом классов вызванном механизмом реализации КА. Показано, что учет многозначности КА в архитектуре алгоритмов МО может повысить эффективность классификации за счет учета случаев одновременной реализации нескольких КА.

Данное исследование выполнено при финансовой поддержке Минцифры России (грант ИБ) в рамках научного проекта, Соглашение №. 40 469–21/23-К от 30.06.2023 г.

Литература

1. *Sheluhin O.I., Osin A.V., Rakovsky D.I.* New Algorithm for Predicting the States of a Computer Network Using Multivalued Dependencies // Aut. Control Comp. Sci. 2023. Т. 57, № 1. С. 48–60. DOI: 10.3103/S0146411623010091.

2. *Pavlov S.V., Dokuchaev V.A., Mytenkov S.S.* Model of a fuzzy dynamic decision support system // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14, № 9. С. 43–47. DOI: 10.36724/2072-8735-2020-14-9-43-47.

3. *Шелухин О.И., Барков В.В., Симонян А.Г.* Обнаружение дрейфа концепта при классификации мобильных приложений с использованием автокодировщиков // Наукоемкие технологии в космических исследованиях Земли. Т. 15, № 15. С. 20–29. DOI: 10.36724/2409-5419-2023-15-3-20-29.

4. *Большаков А.С., Осин А.В., Жила А.И.* Управление информационной безопасностью персональных данных с использованием нечеткой логики // Наукоемкие технологии в космических исследованиях Земли. Т. 13, № 4. С. 37–47. DOI: 10.36724/2409-5419-2021-13-4-37-47.

5. *Большаков А.С., Емец Л.В.* Обнаружение фишингового сайта методами машинного обучения // Телекоммуникации и информационные технологии. 2020. Т. 10, № 1. С. 36–43.

6. *Шелухин О.И., Канаев С.Д.* Скрытие водяных знаков в цветных изображениях с использованием алгебраических фракталов методами 2D вейвлет преобразования // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12, № 5. С. 46–50. DOI: 10.24411/2072-8735-2018-10107.

7. *Шелухин О.И., Шариков А.Ю.* Имитация поведения компьютерной системы с помощью искусственных нейронных сетей // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15, № 5. С. 29–37. DOI: 10.36724/2072-8735-2021-15-5-29-37.

8. *Шелухин О.И., Барков В.В.* Влияние фонового трафика на эффективность классификации трафика мобильных приложений методами интеллектуального анализа данных // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12, № 10. С. 52–57. DOI: 10.24411/2072-8735-2018-10157.

9. *Gharaibeh M., Papadopoulos C.* DARPA-2009 Intrusion Detection Dataset Report [Электронный ресурс]. 2014. URL: https://www.semanticscholar.org/paper/DARPA-2009-Intrusion-Detection-Dataset-Report-Gharaibeh-Papadopoulos/c324e8525581f5c9dc38d943e222beb0050b8e6e (дата обращения: 21.08.2023).

10.NSL-KDD | Datasets | Research | Canadian Institute for Cybersecurity | UNB [Электронный ресурс]. URL: https://www.unb.ca/cic/datasets/nsl.html (дата обращения: 20.08.2023).

11.IDS 2012 | Datasets | Research | Canadian Institute for Cybersecurity | UNB [Электронный ресурс]. URL: https://www.unb.ca/cic/datasets/ids.html (дата обращения: 20.08.2023).

12.*Шелухин О.И., Раковский Д.И.* Влияние многозначности баз данных на результаты многоклассовой классификации компьютерных атак // Вестник СПГУТД. Серия 1. Т. 3, № 3. С. 111–119. DOI: 10.46418/2079-8199_2023_3_18.

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

14.*Шелухин О.И., Раковский Д.И.* Выбор метрических атрибутов редких аномальных событий компьютерной системы методами интеллектуального анализа данных // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15, № 6. С. 40–47. DOI: 10.36724/2072-8735-2021-15-6-40-47.

15.*Фомичева С.Г., Беззатеев С.В.* Механизмы зашиты моделей машинного обучения от состязательных атак // T-Comm: Телекоммуникации и транспорт. 2023. Т. 17, № 10. С. 28–42с. DOI: 10.36724/2072-8735-2023-17-10-28-42.

16.*Шелухин О.И., Осин А.В., Костин Д.В.* Диагностика «здоровья» компьютерной сети на основе секвенциального анализа последовательностных паттернов // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14, № 2. С. 9–16. DOI: 10.36724/2072-8735-2020-14-2-9-16.

17.*Шелухин О.И., Осин А.В., Костин Д.В.* Мониторинг и диагностика аномальных состояний компьютерной сети на основе изучения «исторических данных» // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14, № 4. С. 23–30. DOI: 10.36724/2072-8735-2020-14-4-23-30.

18.*Sheluhin O.I., Barkov V.V., Sekretarev S.A.* The online classification of the mobile applications traffic using data mining techniques // T-Comm. 2019. Т. 13, № 10. С. 60–67.

19.*Sureda Riera T., Bermejo Higuera J.R., Bermejo Higuera J., Sicilia Montalvo J.A., Martínez Herráiz J.J.* SR-BH 2020 multi-label dataset. Harvard Dataverse, 2022. с. DOI: 10.7910/DVN/OGOIXX.

20.*Sheluhin O.I., Ivannikova V.P.* Comparative analysis of informative features quantity and composition selection methods for the computer attacks classification using the UNSW-NB15 dataset // T-Comm. 2020. Т. 14, № 10. С. 53–60с. DOI: 10.36724/2072-8735-2020-14-10-53-60.

21.*Gibaja E., Ventura S. A.* Tutorial on Multilabel Learning // ACM Comput. Surv. 2015. Т. 47, № 3. С. 1–38с. DOI: 10.1145/2716262.

22.*Раковский Д.И.* Влияние проблемы многозначности меток классов системных журналов на защищенность компьютерных сетей // Наукоемкие технологии в космических исследованиях Земли. 2023. Т. 15, № 1. С. 48–56. DOI: 10.36724/2409-5419-2023-15-1-48-56.

23.*Шелухин О.И., Раковский Д.И.* Многозначная классификация компьютерных атак с использованием искусственных нейронных сетей с множественным выходом // Труды учебных заведений связи. Т. 9, № 4. С. 95–111. DOI: 10.31854/1813-324X-2023-9-4-95-111.

24.*Sheluhin O.I., Rakovskiy D.I.* Multi-Label Learning in Computer Networks // 2023 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russian Federation: IEEE, 2023. С. 1–5. DOI: 10.1109/IEEECONF56737.2023.10092157.

25.*Раковский Д.И., Александров И.Д.* Разработка экспериментального стенда для моделирования сетевых атак на компьютерную систему в контролируемых условиях // Сборник трудов научно-технической конференции «Управление и безопасность информации в киберфизических системах». М.: МТУСИ. 2023. С. 111–117.

26.*Раковский Д.И., Александров И.Д., Боков А.Д.* Стенд для сбора телеметрии многозначных компьютерных атак // Сборник трудов III Всероссийской научно-практической конференции «Теория и практика обеспечения информационной безопасности», Москва, Россия. 2023. С. 26–33.

27.*Zhang M.-L., Li Y.-K., Liu X.-Y., Geng X.* Binary relevance for multi-label learning: an overview // Front. Comput. Sci. 2018. Т. 12, № 2. С. 191–202с. DOI: 10.1007/s11704-017-7031-7.

28.*Behdani Z., Darehmiraki M.* An Alternative Approach to Rank Efficient DMUs in DEA via Cross-Efficiency Evaluation, Gini Coefficient, and Bonferroni Mean // J. Oper. Res. Soc. China. 2022. Т. 10, № 4. С. 763–783. DOI: 10.1007/s40305-019-00264-x.

29.*Шелухин О.И., Раковский Д.И.* Многозначная классификация меток классов системных журналов компьютерных сетей. Сравнительный анализ эффективности классификаторов // Вопросы кибербезопасности. 2023. Т. 55, № 3. С. 62–77. DOI: 10.21681/2311-3456-3-62-77.

SIMULATION MODELING OF MULTI-VALUED COMPUTER ATTACKS

Oleg I. Shelukhin
Moscow Technical University of Communications and Informatics, Doctor of Technical Sciences, Head of the Department of Information Security, Professor, Moscow, Russia. E-mail: sheluhin@mail.ru.

Dmitry I. Rakovsky
Moscow Technical University of Communications and Informatics, graduate student, assistant at the Department of Information Security, Moscow, Russia. E-mail: Prophet_alpha@mail.ru.

Ilya D. Aleksandrov
Moscow Technical University of Communications and Informatics, student, Moscow, Russia.

Alexander D. Bokov
Moscow Technical University of Communications and Informatics, student, Moscow, Russia.

ABSTRACT

Introduction: The structure of the developed and a software-implemented stand for collecting telemetry and statistical characteristics of computer attacks (CA), in a computer network (CN), with several attacked and attacking hosts, taking into account their ambiguity at the data collection stage. To test the developed stand, a multivalued spacecraft was carried out on the “victim” host. As an example, two types of attacks were chosen: “denial of service” and “operating system scanning.” It is shown that the number of informationally significant attributes for the “operating system scan” CA exceeds their number for the “denial of service” CA, which must be taken into account when implementing machine learning algorithms in problems of attack detection and classification.

Keywords: experimental data; multi-label classification; network attack; computer attack; research stand; information security.

REFERENCES

1. Sheluhin O.I., Osin A.V., Rakovsky D.I. New Algorithm for Predicting the States of a Computer Network Using Multivalued Dependencies. Aut. Control Comp. Sci. 2023. T. 57, No. 1. Pp. 48–60. DOI: 10.3103/S0146411623010091.
2. Pavlov S.V., Dokuchaev V.A., Mytenkov S.S. Model of a fuzzy dynamic decision support system. T-Comm: Telecommunications and Transport. 2020. T. 14, No. 9. Pp. 43–47. DOI: 10.36724/2072-8735-2020-14-9-43-47.
3. Shelukhin O.I., Barkov V.V., Simonyan A.G. Detection of concept drift when classifying mobile applications using autoencoders. Science-intensive technologies in space exploration of the Earth. T. 15, No. 15. Pp. 20–29. DOI: 10.36724/2409-5419-2023-15-3-20-29.
4. Bolshakov A.S., Osin A.V., Zhila A.I. Managing information security of personal data using fuzzy logic. Science-intensive technologies in space exploration of the earth. T. 13, No. 4. Pp. 37–47. DOI: 10.36724/2409-5419-2021-13-4-37-47.
5. Bolshakov A.S., Yemets L.V. Detection of a phishing site using machine learning methods. Telecommunications and information technologies. 2020. Vol. 10, No. 1. Pp. 36–43.
6. Shelukhin O.I., Kanaev S.D. Hiding watermarks in color images using algebraic fractals using 2D wavelet transform methods. T-Comm: Telecommunications and Transport. 2018. T. 12, No. 5. Pp. 46–50. DOI: 10.24411/2072-8735-2018-10107.
7. Shelukhin O.I., Sharikov A.Yu. Simulating the behavior of a computer system using artificial neural networks. T-Comm. 2021. T. 15, No. 5. Pp. 29–37. DOI: 10.36724/2072-8735-2021-15-5-29-37.
8. Shelukhin O.I., Barkov V.V. The influence of background traffic on the efficiency of mobile application traffic classification using data mining methods. T-Comm: Telecommunications and Transport. 2018. T. 12, No. 10. Pp. 52–57. DOI: 10.24411/2072-8735-2018-10157.
9. Gharaibeh M., Papadopoulos C. DARPA-2009 Intrusion Detection Dataset Report [Electronic resource]. 2014. URL: https://www.semanticscholar.org/paper/DARPA-2009-Intrusion-Detection-Dataset-Report-Gharaibeh-Papadopoulos/c324e8525581f5c9dc38d943e222beb0050b8e6e (accessed: 08/21/2023).
10. NSL-KDD | Datasets | Research | Canadian Institute for Cybersecurity | UNB [Electronic resource]. URL: https://www.unb.ca/cic/datasets/nsl.html (accessed 08/20/2023).
11. IDS 2012 | Datasets | Research | Canadian Institute for Cybersecurity | UNB [Electronic resource]. URL: https://www.unb.ca/cic/datasets/ids.html (accessed 08/20/2023).
12. Shelukhin O.I., Rakovsky D.I. The influence of database ambiguity on the results of multi-class classification of computer attacks. Bulletin of SPGUTD. Series 1. T. 3, No. 3. Pp. 111–119. DOI: 10.46418/2079-8199_2023_3_18.
13. Rakovsky D.I. Detection of computer attacks and prevention of disruptions in the functioning of computer networks based on multi-valued patterns. Collection of proceedings of the III All-Russian scientific school-seminar “Modern trends in the development of methods and technologies for information security”. 2023. Pp. 307–311.
14. Shelukhin O.I., Rakovsky D.I. Selection of metric attributes of rare anomalous events of a computer system using data mining methods. T-Comm: Telecommunications and Transport. 2021. T. 15, No. 6. Pp. 40–47. DOI: 10.36724/2072-8735-2021-15-6-40-47.
15. Fomicheva S.G., Bezzateev S.V. Mechanisms for protecting machine learning models from adversarial attacks. T-Comm. 2023. T. 17, No. 10. Pp. 28–42. DOI: 10.36724/2072-8735-2023-17-10-28-42.
16. Shelukhin O.I., Osin A.V., Kostin D.V. Diagnostics of the “health” of a computer network based on sequential analysis of sequential patterns. T-Comm. 2020. T. 14, No. 2. Pp. 9–16. DOI: 10.36724/2072-8735-2020-14-2-9-16.
17. Shelukhin O.I., Osin A.V., Kostin D.V. Monitoring and diagnostics of abnormal states of a computer network based on the study of “historical data”. T-Comm. 2020. T. 14, No. 4. Pp. 23–30. DOI: 10.36724/2072-8735-2020-14-4-23-30.
18. Sheluhin O.I., Barkov V.V., Sekretarev S.A. The online classification of the mobile applications traffic using data mining techniques. T-Comm. 2019. Vol. 13, No. 10, Pp. 60–67.

19. Sureda Riera T., Bermejo Higuera J.R., Bermejo Higuera J., Sicilia Montalvo J.A., Martínez Herráiz J.J. SR-BH 2020 multi-label dataset. Harvard Dataverse, 2022. p. DOI: 10.7910/DVN/OGOIXX.
20. Sheluhin O.I., Ivannikova V.P. Comparative analysis of informative features quantity and composition selection methods for the computer attacks classification using the UNSW-NB15 dataset. T-Comm. 2020. T. 14, No. 10. Pp. 53–60. DOI: 10.36724/2072-8735-2020-14-10-53-60.
21. Gibaja E., Ventura S. A Tutorial on Multilabel Learning. ACM Comput. Surv. 2015. T. 47, No. 3. Pp. 1–38. DOI: 10.1145/2716262.
22. Rakovsky D.I. The influence of the problem of polysemy of system log class labels on the security of computer networks. Science-Intensive Technologies in Space Research of the Earth. 2023. T. 15, No. 1. Pp. 48–56. DOI: 10.36724/2409-5419-2023-15-1-48-56.
23. Shelukhin O.I., Rakovsky D.I. Multivalued classification of computer attacks using artificial neural networks with multiple outputs. Proceedings of educational institutions of communication. T. 9, No. 4. Pp. 95–111. DOI: 10.31854/1813-324X-2023-9-4-95-111.
24. Sheluhin O.I., Rakovskiy D.I. Multi-Label Learning in Computer Networks. 2023 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russian Federation: IEEE, 2023. Pp. 1–5. DOI: 10.1109/IEEECONF56737.2023.10092157.
25. Rakovsky D.I., Aleksandrov I.D. Development of an experimental stand for simulating network attacks on a computer system under controlled conditions. Collection of proceedings of the scientific and technical conference “Information management and security in cyber-physical systems”. M.: MTUSI. 2023. Pp. 111–117.
26. Rakovsky D.I., Aleksandrov I.D., Bokov A.D. Stand for collecting telemetry of multi-valued computer attacks. Collection of proceedings of the III All-Russian Scientific and Practical Conference “Theory and Practice of Ensuring Information Security”, Moscow, Russia. 2023. Pp. 26–33.
27. Zhang M.-L., Li Y.-K., Liu X.-Y., Geng X. Binary relevance for multi-label learning: an overview. Front. Comput. Sci. 2018. T. 12, No. 2. Pp. 191–202. DOI: 10.1007/s11704-017-7031-7.
28. Behdani Z., Darehmiraki M. An Alternative Approach to Rank Efficient DMUs in DEA via Cross-Efficiency Evaluation, Gini Coefficient, and Bonferroni Mean. J. Oper. Res. Soc. China. 2022. T. 10, No. 4. Pp. 763–783. DOI: 10.1007/s40305-019-00264-x.
29. Shelukhin O.I., Rakovsky D.I. Multi-valued classification of class labels of system logs of computer networks. Comparative analysis of the effectiveness of classifiers. Issues of cybersecurity. 2023. T. 55, No. 3. Pp. 62–77. DOI: 10.21681/2311-3456-3-62-77.