ИСПОЛЬЗОВАНИЕ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ RANDOM FOREST ДЛЯ ВЫЯВЛЕНИЯ СЛОЖНЫХ КОМПЬЮТЕРНЫХ ИНЦИДЕНТОВ

Павлычев А.В.[1], Стародубов М.И.[2], Галимов А.Д.[3]

Цель работы: *разработка способа выявления сложных компьютерных инцидентов, осуществляемых злоумышленниками путем эксплуатации уязвимостей информационных систем.*

Метод исследования: *анализ записей в системных журналах операционной системы Microsoft Windows с использованием алгоритма машинного обучения Random Forest («Случайный лес»).*

Полученный результат: *несмотря на большое разнообразие различных видов вредоносного программного обеспечения, используемого злоумышленниками при проведении компьютерных атак, все они оставляют следы своего функционирования в сетевой инфраструктуре, подвергшейся несанкционированному воздействию. Одним из способов выявления компьютерных инцидентов является исследование файлов журналов различных информационных систем, в том числе системных журналов операционной системы на предмет выявления скрытых закономерностей и различных аномалий. Функционирование любой компьютерной программы можно представить в виде уникального набора записей в системных журналах операционной системы, которые можно рассматривать в качестве признаков объекта. В работе проведен анализ журнала Security («Безопасность») операционной системы после эксплуатации различных уязвимостей, популярных в хакерской среде. На сформированном таким образом наборе данных с использованием алгоритма машинного обучения построена модель, позволяющая в дальнейшем выявлять объекты, подвергшиеся несанкционированному воздействию.*

Научная новизна *состоит в создании способа выявления сложных компьютерных инцидентов, основанного на результатах изучения журналов операционной системы с использованием алгоритма машинного обучения.*

Ключевые слова: *компьютерные атаки, несанкционированное воздействие, анализ системных журналов, журнал Security, алгоритмы машинного обучения.*

DOI:10.21681/2311-3456-2022-5-74-81

Введение

В условиях фактически объявленной России кибервойны наибольшую опасность для цифрового суверенитета страны представляют злоумышленники с высоким потенциалом – профессиональные политически мотивированные хакеры из недружественных государств.

Согласно изученным аналитическим отчетам, в первом полугодии 2022 года количество инцидентов информационной безопасности в России выросло практически на четверть в сравнении с аналогичным периодом 2021 года. Наибольшее число инцидентов с разным уровнем критичности связано с применением хакерами вредоносного программного обеспечения (далее – вредоносного ПО). Наблюдается значительный рост инцидентов, связанных с эксплуатацией уязвимостей. Повышается скорость применения эксплойтов для проведения атак на ресурсы российских компаний.

Складывающаяся политическая обстановка говорит о дальнейшем росте хакерских атак на россий-

1 Павлычев Алексей Викторович, директор Центра информационной безопасности ФГАОУ ВО «Дальневосточный федеральный университет» (ДВФУ), г. Владивосток, Россия. E-mail: pavlychev.av@dvfu.ru

2 Стародубов Максим Игоревич, аспирант ФГАОУ ВО «Дальневосточный федеральный университет» (ДВФУ), г. Владивосток, Россия. E-mail: starodubov.mi@dvfu.ru

3 Галимов Александр Дмитриевич, аспирант ФГАОУ ВО «Дальневосточный федеральный университет» (ДВФУ), г. Владивосток, Россия. E-mail: galimov.ad@dvfu.ru

скую информационную инфраструктуру с использованием новых видов кибероружия [1-3; 16].

Указанные обстоятельства свидетельствуют о высокой актуальности исследований, направленных на выявление различных компьютерных инцидентов.

В рамках указанного исследования авторами поставлен ряд задач:

- Проведение анализа сложных компьютерных инцидентов на предмет оставления следов в системных журналах информационной системы, подвергшейся несанкционированному воздействию;
- Разработка автоматизированного метода обработки данных системных журналов информационных систем, формирование набора данных для дальнейшего исследования;
- Краткий обзор основных методов машинного обучения для решения задач в области информационной безопасности, выбор оптимального алгоритма машинного обучения с учетом полученного набора данных;
- Применение к полученным данным выбранного алгоритма машинного обучения, подбор оптимальных параметров для получения максимальной точности модели и оценка эффективности разработанного способа.

Анализ сложных компьютерных инцидентов

Важной задачей информационной безопасности является обнаружение вредоносного ПО. Последние тенденции развития вредоносных программ свидетельствуют, что все более актуальной становится разработка эффективных способов обнаружения ранее неизвестного вредоносного ПО.

Существует множество видов вредоносного ПО, при этом все они оставляют следы своего функционирования в операционной системе. Одним из способов выявления компьютерных инцидентов является исследование файлов журналов различных информационных систем, в том числе системных журналов операционной системы на предмет выявления скрытых закономерностей и различных аномалий.

В операционной системе Microsoft Windows ведутся журналы, которые регистрируют пользовательские события и работу системных и прикладных программ на компьютере. Журнал событий представляет собой бинарный файл специального формата (с расширением EVTX), схожий с файлом базы данных. Набольший интерес для исследователей информационной безопасности представляет журнал Security (Безопасность) операционной системы Windows.

Согласно документации Microsoft журнал Security содержит 422 возможных события, имеющих уникальный код (EventID), который может иметь значения в интервале от 1100 до 8191 [4-5].

При помощи Банка данных угроз ФСТЭК России в рамках исследования выбраны и в дальнейшем проэксплуатированы уязвимости операционных систем Windows, которые широко используются для компрометации целевых информационных систем различными хакерскими группировками:

- CVE-2017-0144 / BDU:2017-01099 (CVSS v3.0 Base Score 9.8 HIGH) - уязвимость протокола SMBv1 операционной системы Microsoft Windows, позволяющая нарушителю выполнить произвольный код («EternalBlue»);
- CVE-2020-0796 / BDU:2020-01005 (CVSS v3.0 Base Score 10 HIGH) - уязвимость реализации сетевого протокола Server Message Block (SMBv3) операционных систем Windows, позволяющая нарушителю выполнить произвольный код;
- CVE-2021-1675 / BDU:2021-03322 (CVSS v3.0 Base Score 9.8 HIGH) – уязвимость операционных систем Windows, связанная с небезопасным управлением привилегиями, позволяющая нарушителю повысить свои привилегии («PrintNightmare»);
- CVE-2021-24084 / BDU:2021-00932 (CVSS v3.0 Base Score 5.5 MEDIUM) - уязвимость диспетчера мобильных устройств Windows Mobile Device Management операционных систем Windows, позволяющая нарушителю получить несанкционированный доступ к защищаемой информации;
- CVE-2021-36934 / BDU:2021-03913 (CVSS v3.0 Base Score 7.8 HIGH) - уязвимость операционных систем Windows, связанная с недостатками разграничения доступа, позволяющая нарушителю повысить свои привилегии («SeriousSAM» или «HiveNightmare»);
- CVE-2021-40444 / BDU:2021-04442 (CVSS v3.0 Base Score 7.8 HIGH) - уязвимость механизма MSHTML браузера Internet Explorer, связанная с неверным управлением генерацией кода, позволяющая нарушителю выполнить произвольный код;
- CVE-2021-40449 / BDU:2021-05018 (CVSS v3.0 Base Score 7.8 HIGH) - уязвимость компонента Win32k (Win32k.sys) операционной системы

Windows, связанная с использованием памяти после её освобождения, позволяющая нарушителю повысить свои привилегии;

- CVE-2022-21882 / BDU:2022-00596 (CVSS v3.0 Base Score 7.8 HIGH) - уязвимость компонента Win32k (Win32k.sys) операционных систем Windows, позволяющая нарушителю повысить свои привилегии;
- CVE-2022-29072 / BDU:2022-02366 (CVSS v3.0 Base Score 7.8 HIGH) - уязвимость библиотеки 7z.dll файлового архиватора 7-Zip, позволяющая нарушителю повысить свои привилегии;
- CVE-2022-30190 / BDU:2022-03226 (CVSS v3.0 Base Score 7.8 HIGH) - уязвимость утилиты сбора диагностических данных и устранения неполадок Microsoft Support Diagnostics Tool операционных систем Windows, связанная с ошибками при обработке вызываемого URL-адреса, позволяющая нарушителю выполнить произвольный код с привилегиями вызывающего приложения («Follina»).

На специально подготовленной инфраструктуре проведена эксплуатация указанных уязвимостей. После успешного применения вредоносного ПО осуществлен сбор и анализ записей журналов Security скомпрометированных рабочих станций. Установлено, что каждый эксплойт оставил уникальный набор записей в системных журналах. По результатам эксперимента сформирован набор данных (датасет), содержащий уникальные наборы значений EventID журналов Security (Рис. 1)

Ввиду значительных объемов событий в системных журналах необходимо применение автоматизированных методов обработки данных. Для решения данной задачи возможно использование методов машинного обучения.

Краткий обзор методов машинного обучения

Машинное обучение представляет собой метод анализа данных, который автоматизирует построение аналитических моделей. Это ветвь искусственного интеллекта, основанная на идее, что системы могут учиться на данных, выявлять закономерности и принимать решения с минимальным вмешательством человека.

Машинное обучение доказало свою эффективность в решении различных аналитических задач и все чаще используется для обнаружения угроз и автоматического устранения их, прежде чем они смогут нанести ущерб. Разработанные алгоритмы быстро сканирует большие объемы данных и анализирует их с помощью статистики [6-8].

В рамках проводимого исследовании основным подходом является Data Mining - выявление скрытых закономерностей и взаимозависимостей в больших наборах данных для поддержки принятия решений.

В технологии Data Mining гармонично объединены строго формализованные методы и методы неформального анализа, т.е. количественный и качественный анализ данных.

Большинство аналитических методов, используемые в технологии Data Mining – это известные математические алгоритмы и методы. Ввиду развития современных технологий данные методы широко применяются не только в теоретических исследованиях, но и для решения практических задач [9].

1100	1101	1102	1104	1105	1108	4608	4609	4610	4611	...	6416	6417	6418	6419	6420	6421	6422	6423	6424	8191
0	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

Рис. 1. Пример уникальных наборов значений EventID журналов Security, полученных по результатам проведенного эксперимента

В рамках исследования изучен ряд прикладных способов обнаружения компьютерных инцидентов с использованием алгоритмов машинного обучения.

1. Деревья решений

Дерево решений - логический алгоритм классификации, решающий задачи классификации и регрессии. Дерево решений – это метод машинного обучения, основанный на рекурсивной древовидной структуре. Дерево решений состоит из ряда элементов: корневого и промежуточного узла, пути и конечного узла. Корневой и промежуточный узел дерева представляет объект или атрибут. Каждый путь расхождения дерева представляет возможные значения родительского узла (объекта). Конечный узел соответствует прогнозируемой категории или классифицированному атрибуту. Результирующее дерево далее представляется в виде правил «если-то».

Данный метод используется в системе обнаружения и предотвращения вторжений SNORT. Авторы статьи «Using decision trees to improve signature- based intrusion detection» в своем эксперименте заменили правила SNORT с использованием модели дерева решений. Хотя авторы не приводят никаких количественных показателей, исследование выявило существенное ускорение с использованием алгоритма дерева решений, а время обработки правил резко сократилось [10].

2. Случайный лес

Случайный лес – один из примеров объединения классификаторов в ансамбль. Для задачи классификации выбирается решение по большинству результатов, выданных классификаторами, а в задаче регрессии - по их среднему значению.

Таким образом, случайный лес представляет собой усреднение над решающими деревьями, при обучении которых для каждого разбиения признаки выбираются из некоторого случайного подмножества признаков.

Случайный лес имеет применение, например, для измерения объема спама и для обнаружения вторжений. Для тестирования метода по обнаружению вторжений использовался набор данных NSL-KDD. На данном наборе метод показал точность обнаружения вторжений 69,7% и 80,2% [11].

3. Байесовская сеть

Байесовская сеть – это направленный ациклический граф, каждой вершине которого соответствует случайная переменная, а дуги графа кодируют отношения условной независимости между этими переменными. Узлы, представляющие потомка, зависят от родительских узлов, и каждый узел поддерживает состояния формы условной вероятности и случайной величины.

Байесовскую сеть можно использовать для обнаружения аномалий, а известные сигнатуры и шаблоны атак также можно сравнивать с потоковыми данными для известных атак. В статье «A framework for an adaptive intrusion detection system using Bayesian network» авторы описывают создание системы обнаружения вторжений с использованием байесовской сети. Для моделирования системы использовался набор данных KDD с девятью его атрибутами. Модель обеспечила уровень обнаружения атак на уровне 99% [12].

4. Кластеризация

Это метод обучения без учителя, в котором мера сходства используется для группировки данных. Алгоритмы кластеризации могут обучаться на данных, полученных в результате аудита, при этом оператору не требуется явное описание различных классов атак.

Авторы статьи «Intrusion signature creation via clustering anomalies» демонстрируют применение обнаружения сигнатур в реальном времени с использованием алгоритма кластеризации. Нормальный и аномальный сетевой трафик был создан схемой кластеризации на основе плотности, известной как Simple Logfile Clustering Tool (SLCT). Используются две схемы кластеризации: во-первых, схема для обнаружения обычных сценариев и сценариев атаки, во-вторых, другая схема используется для контролируемого определения нормального трафика. Для проверки модели использовался набор данных KDD. Для неизвестных атак (в том числе с использованием уязвимостей «нулевого дня») достигнута точность от 70% до 80% [13].

5. Нейронные сети

Нейронная сеть (или искусственная нейронная сеть) строится по аналогии работы человеческого мозга. Сеть имеет структуру слоев, ввод данных активирует нейрон второго слоя сети, что, в свою очередь, выводит на следующий уровень иерархии и так далее. Вывод производится последним слоем сети. Функция, которая преобразует несколько входных параметров в один выходной называется искусственным нейроном. Одним из основных недостатков нейронной сети является большое количество времени, необходимого для обучения.

Автор статьи «Artificial neural networks for misuse detection» описывает модель нейронной сети, которая использует многокатегориальный классификатор для обнаружения аномалий. Данные классифицируются либо как обычный трафик, либо как вредоносный трафик. Для генерации данных использовался сетевой монитор RealSecure со встроенными сценариями атак. Предварительная обработка данных выполнялась с использованием девяти выбранных параметров: код ICMP, тип ICMP, адрес источника, адрес назначения, номер протокола, порт источника, порт назначения, длина необработанных данных и тип необработанных данных. Исследователи сообщают о частоте ошибок 0,058 и 0,070 во время сценариев обучения и тестирования. Точность алгоритма составила 93% на этапе тестирования [14].

6. Метод опорных векторов

Метод опорных векторов считается наиболее часто используемым и успешным методом машинного обучения для задач кибербезопасности, особенно для средств обнаружения вторжений. Метод опорных векторов классифицирует и разделяет два класса данных по обе стороны от гиперплоскости. Точность классификации точек данных может быть повышена за счет увеличения расстояний между гиперплоскостями. Точки данных, лежащие на границе гиперплоскости называются опорными векторными точками. Метод опорных векторов, как и нейронные сети, требует много памяти для обработки и времени для обучения.

Подробное сравнение метода опорных векторов и искусственной нейронной сети провели авторы статьи «Application of SVM and ANN for intrusion detection». Сравнение проводилось на наборе данных KDD, на котором метод опорных векторов показал лучшие результаты [15, 16].

В рамках исследования объектом изучения с помощью методов машинного обучения является набор данных, содержащий булевые функции, соответствующие наступлению определенного события и записанные в журнал Security. Построенная модель должна предсказывать, является ли программа с заданным набором признаков вредоносной (значение 1), либо легитимной (значение 0).

С учетом набора изучаемых данных и специфики решаемой задачи алгоритмом машинного обучения для анализа в рамках исследования выбран алгоритм Random Forest («Случайный лес»).

Применение алгоритма, оценка эффективности

В качестве инструмента использовалась библиотека Scikit-learn, содержащая класс RandomForestClassifier.

Указанный алгоритм содержит следующие основные входные параметры:

- n_estimators - число деревьев в «лесу»;
- max_features - число признаков для ветвления;
- max_depth - максимальная глубина дерева;
- min_samples_split - минимальное число объектов, необходимое для того, чтобы узел дерева мог разделиться;
- min_samples_leaf - минимальное число объектов в листьях;
- bootstrap - использование для построения деревьев подвыборки с возвращением.

В рамках очередной задачи исследования осуществлен перебор различных параметров модели. Результаты сравнительного анализа приведены на Рис. 2.:

С использованием полученных результатов осуществлен поиск значений параметров, при которых эффективность модели является максимальной:

n_estimators	param_min_samples_split	param_min_samples_leaf	param_max_features	param_max_depth	param_bootstrap	mean_test_score	rank_test_score
700	2	2	log2	11	True	0.777778	1
300	18	39	log2	2	True	0.444444	2
800	39	44	sqrt	4	True	0.444444	2
200	39	44	sqrt	1	True	0.444444	2
1000	50	28	sqrt	4	True	0.444444	2
700	50	23	log2	2	True	0.444444	2
600	39	34	log2	13	True	0.444444	2
800	50	39	log2	1	True	0.444444	2
1000	7	12	sqrt	2	True	0.444444	2
700	23	12	sqrt	8	True	0.444444	2

Рис. 2. Результаты работы алгоритма с различным набором параметров

```
{'bootstrap': True,
  'max_depth': 2,
  'max_features': 'log2',
  'min_samples_leaf': 2,
  'min_saples_split': 2,
  'n_estimators': 100}
classifier_best.score (x_train, y_train)
1.0
classifier_best.score (x_test, y_test)
0.8888888888888
```

Итоговая эффективность модели на тренировочной выборке составила 100%, а на тестовой выборке – 89%. Измерение эффективности модели производилось с помощью F-меры, которая гармонично учитывает как ложноположительные, так и ложноотрицательные значения классификатора (ошибки первого и второго рода). Такие значения как точность (recall) и полнота (precision), составили 80% и 100% соответственно:

```
'Hyperparameter Tuned Random Forest
recall score'
0.8
'Hyperparameter Tuned Random Forest
precision score'
1.0
'Hyperparameter Tuned Random Forest f1
score'
0.8888888888889
```

Заключение

В рамках работы разработан классификатор, позволяющий с эффективностью 89% выявлять признаки сложных компьютерных инцидентов на основании анализа записей журнала Security операционной системы Microsoft Windows. Задачи, поставленные авторами, решены в полном объеме.

В ходе дальнейших исследований планируется проведение дополнительных экспериментов, направленных на эксплуатацию более широкого набора уязвимостей, а также применение к полученному набору данных новых алгоритмов машинного обучения, с целью повышения эффективности разработанного способа выявления компьютерных инцидентов.

Исследование проведено при финансовой поддержке Минобрнауки России («Грант ИБ МТУСИ») № 40469-23-2021-К.

Литература

1. R. Badhwar, The CISO's Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms // Springer. – 2021. – P. 279–285.
2. N. Dutta, N. Jadav, S. Tanwar, Cyber Security: Issues and Current Trends // Springer. – 2021. – P. 129–141.
3. S. James, Carbanak Threatens Critical Infrastructure: Cybercriminal APTs Merit Significant Investigation and Discussion / S. James. – Washington, DC, USA: ICIT, 2017. – 16 p.
4. Markus Ring, Daniel Schlör, Sarah Wunderlich, Dieter Landes, Andreas Hotho, Malware detection on windows audit logs using LSTMs // Computers & Security. – 2021. – Vol. 109. – P. 1-12.
5. Thomas T. Machine learning approaches in cyber security analytics / Tony Thomas, Athira P Vijaya-raghavan, Sabu Emmanuel. – Singapore: Springer, 2020. – 217 p.
6. Zico J. Kolter, Marcus A. Maloof, Learning to Detect Malicious Executables in the Wild // Journal of Machine Learning Research. – 2006. – Vol. 7. – P. 2721-2744.
7. Joseph Rabaiotti, Counter Intrusion Software: Malware Detection using Process Behaviour Classification and Machine Learning [Электронный ресурс]. – Режим доступа: URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.2417 &rep=rep1&type=pdf свободный (дата обращения: 10.08.2022).
8. Shu He, Gene Moo Lee, Sukjin Han, Andrew B. Whinston, how would information disclosure influence organizations' outbound spam volume? Evidence from a field experiment // Journal of Cybersecurity. – 2016. – Vol. 2. – P. 99-118.
9. C. Kruegel, T. Toth, using decision trees to improve signature-based intrusion detection // 6th International Workshop on the Recent Advances in Intrusion Detection, West Lafayette. – 2003. – P. 173–191.
10. Sean Miller, Curtis C.R. Busby-Earle, Multi-Perspective Machine Learning a Classifier Ensemble Method for Intrusion Detection // The 2017 International Conference on Machine Learning and Soft Computing. – 2017. - P. 7-12.
11. Farah Jemili, Montassar Zaghdoud, Mohamed Ben Ahmed, A framework for an adaptive intrusion detection system using Bayesian network // Intelligence and Security Informatics, IEEE. – 2007.
12. Gilbert R. Hendry, Shanchieh Jay Yang, Intrusion signature creation via clustering anomalies // SPIE Defense and Security Symposium, International Society for Optics and Photonics. – 2008.
13. Cannady, Artificial neural networks for misuse detection // Proceedings of the 1998 National Information Systems Security Conference, Arlington, VA. – 1998. - P. 443.
14. Wun-Hwa Chen, Sheng-Hsun Hsu, Hwang-Pin Shen, Application of SVM and ANN for intrusion detection // Computers & Operations Research. – 2005. – Vol. 32. – No. 10. – P. 2617-2634.

15. Bernhard Schölkopf, Robert C. Williamson, Alex Smola, John Shawe-Taylor, John Platt, Support vector method for novelty detection // Advances in Neural Information Processing Systems. – 2000. – P. 582-588.
16. Марков А.С. Техническая защита информации. Курс лекций. М. 2020. 220 с. ISBN 978-5-6045553-0-9

USING THE RANDOM FOREST MACHINE LEARNING ALGORITHM FOR THE EXTRACTION OF COMPLEX COMPUTER INCIDENTS

Pavlychev A.V.[4], Starodubov M.I.[5], Galimov A.D.[6]

The aim of the work *is to develop a way to identify complex computer incidents carried out by attackers by exploiting vulnerabilities of information systems.*

The research method *is the analysis of entries in the system logs of the Microsoft Windows operating system using the Random Forest machine learning algorithm.*

The result obtained: *despite the wide variety of different types of malicious software used by attackers in conducting computer attacks, they all leave traces of their functioning to the network infrastructure that has been exposed to unauthorized effects. One of the ways to identify computer incidents is to examine the log files of various information systems, including the system logs of the operating system for the identification of hidden patterns and various anomalies. The functioning of any computer program can be represented as a unique set of records in the system logs of the operating system, which can be considered as features of an object. The paper analyzes the Security log of the operating system after exploiting various vulnerabilities that are popular in the hacker environment. On the data set formed in this way using a machine learning algorithm, a model is built that allows you to further identify objects that have been exposed to unauthorized effect.*

The scientific novelty *consists in creating a way to identify complex computer incidents based on the results of studying the logs of the operating system using a machine learning algorithm.*

Keywords: *computer attacks, unauthorized impact, analysis of system logs, Security log, machine learning algorithms.*

References

1. R. Badhwar, The CISO's Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms // Springer. – 2021. – P. 279–285.
2. N. Dutta, N. Jadav, S. Tanwar, Cyber Security: Issues and Current Trends // Springer. – 2021. – P. 129–141.
3. S. James, Carbanak Threatens Critical Infrastructure: Cybercriminal APTs Merit Significant Investigation and Discussion / S. James. – Washington, DC, USA: ICIT, 2017. – 16 p.
4. Markus Ring, Daniel Schlör, Sarah Wunderlich, Dieter Landes, Andreas Hotho, Malware detection on windows audit logs using LSTMs // Computers & Security. – 2021. – Vol. 109. – P. 1-12.
5. Thomas T. Machine learning approaches in cyber security analytics / Tony Thomas, Athira P Vijaya-raghavan, Sabu Emmanuel. – Singapore: Springer, 2020. – 217 p.
6. Zico J. Kolter, Marcus A. Maloof, Learning to Detect Malicious Executables in the Wild // Journal of Machine Learning Research. – 2006. – Vol. 7. – P. 2721-2744.
7. Joseph Rabaiotti, Counter Intrusion Software: Malware Detection using Process Behaviour Classification and Machine Learning [Online]. – URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.2417 &rep=rep1&type=pdf.
8. Shu He, Gene Moo Lee, Sukjin Han, Andrew B. Whinston, How would information disclosure influence organizations' outbound spam volume? Evidence from a field experiment // Journal of Cybersecurity. – 2016. – Vol. 2. – P. 99-118.

4 Aleksey V. Pavlychev, Director, Cybersecurity Center, Far Eastern Federal University (FEFU), Vladivostok, Russia. E-mail: pavlychev.av@dvfu.ru
5 Maksim I. Starodubov, Ph.D. Student, Far Eastern Federal University (FEFU), Vladivostok, Russia. E-mail: starodubov.mi@dvfu.ru
6 Alexander D. Galimov , Ph.D. Student, Far Eastern Federal University (FEFU), Vladivostok, Russia. E-mail: galimov.ad@dvfu.ru

9. C. Kruegel, T. Toth, Using decision trees to improve signature-based intrusion detection // 6th International Workshop on the Recent Advances in Intrusion Detection, West Lafayette. – 2003. – P. 173–191.
10. Sean Miller, Curtis C.R. Busby-Earle, Multi-Perspective Machine Learning a Classifier Ensemble Method for Intrusion Detection // The 2017 International Conference on Machine Learning and Soft Computing. – 2017. - P. 7-12.
11. Farah Jemili, Montassar Zaghdoud, Mohamed Ben Ahmed, A framework for an adaptive intrusion detection system using Bayesian network // Intelligence and Security Informatics, IEEE. – 2007.
12. Gilbert R. Hendry, Shanchieh Jay Yang, Intrusion signature creation via clustering anomalies // SPIE Defense and Security Symposium, International Society for Optics and Photonics. – 2008.
13. Cannady, Artificial neural networks for misuse detection // Proceedings of the 1998 National Information Systems Security Conference, Arlington, VA. – 1998. - P. 443.
14. Wun-Hwa Chen, Sheng-Hsun Hsu, Hwang-Pin Shen, Application of SVM and ANN for intrusion detection // Computers & Operations Research. – 2005. – Vol. 32. – No. 10. – P. 2617-2634.
15. Bernhard Schölkopf, Robert C. Williamson, Alex Smola, John Shawe-Taylor, John Platt, Support vector method for novelty detection // Advances in Neural Information Processing Systems. – 2000. – P. 582-588.
16. Markov A.S. Tehnicheskaja zashhita informacii. Kurs lekcij. M. 2020. 220 s. ISBN 978-5-6045553-0-9