ration schemes based on TSN standards are proposed: centralized and distributed. Having considered these schemes, we will identify their limitations necessary in meeting requirements close to real time and ensuring strict quality of service guarantees, taking into account the restrictions applied to a time-sensitive environment. The work also reveals the need to use additional equipment, a centralized controller, to reallocate priorities.

Keywords: *dynamic priority, reinforcement learning, time-sensitive networks, deadline*

DOI: 10.18469/ikt.2023.21.3.03

Gerasimov Vyacheslav Vasilevich, Povolzhskiy State University of Telecommunications and Informatics, 23, L. Tolstoy Street, Samara, 443010, Russian Federation; Senior Teacher of Networks and Communication Systems Department. Tel. +7 987 434-77-32. E-mail: v.gerasimov@psuti.ru

References

1. Roslyakov A.V. et al. TSN Ethernet time-sensitive networking. *Infocommunicionnye technologii*, 2021, vol. 19, no. 2, pp. 187–201. DOI: 10.18469/ikt.2021.19.2.07 (In Russ.)
2. Meng S., Zhu Q., Xia F. Improvement of the dynamic priority scheduling algorithm based on a heapsort. *IEEE Access*, 2019, vol. 7, pp. 503–510.
3. Bulb N.S., Fischer M. Reinforcement learning assisted routing for time-sensitive networks. *IEEE GLOBECOM Global Communications Conference*, 2022, pp. 3863–3868.
4. Grigorjew A. et al. ML-assisted latency assignments in time-sensitive networking. *IFIP/IEEE International Symposium on Integrated Network Management*. Bordeaux, 2021, pp. 116–124.
5. Lee H. et al. Panda: reinforcement learning-based priority assignment for multi-proce. *IEEE Access*, 2020, no. 8, pp. 185570–185583.
6. Roslyakov A.V. et al. Time-sensitive networking standardization. *Standarty i kachestvo*, 2021, no. 4 (1006), pp. 29–33. DOI: 10.35400/0038-9692-2021-4-48-53 (In Russ.)
7. Roslyakov A.V. *NETWORK 2030: Architecture, Technologies, Services*. Moscow: ICTS «Kolos-s», 2022, 278 p. (In Russ.)
8. Atiq M.K. et al. When IEEE 802.11 and 5G meet time-sensitive networking. *IEEE Open Journal of the Industrial Electronics Society*, 2021, vol. 3, pp. 14–36. DOI: 10.1109/OJIES.2021.3135524
9. Kogan S. Standardization of solutions and segmentation of the transport level of the 5G network. *Pervaya milya*, 2021, no. 2, pp. 40–47. DOI: 10.22184/2070-8963.2021.94.2.40.47 (In Russ.)
10. Gutierrez M. et al. Self-configuration of IEEE 802.1 TSN networks. *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Limassol, 2017, pp. 1–8.

Received 26.01.2024

УДК 004.855.53

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛЬНОГО ПОВЕДЕНИЯ ТРАФИКА

Поздняк И.С., Макаров И.С.

Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ

E-mail: i.pozdnyak@psuti.ru, i.makarov@psuti.ru

В статье рассматривается способ обнаружения аномального поведения трафика с использованием методов машинного обучения. Для этого используется набор данных, содержащий значительный объем трафика, собранный в момент проведения атаки на Web-приложение. Набор содержит три варианта атак: Brute Force, XSS, SQL-инъекция. Отдельно рассмотрен дамп трафика, содержащий атаку Infiltration. Проведен сравнительный анализ моделей машинного обучения с выбором наиболее оптимального. В статье также приводится описание процедуры предобработки данных, которая проводится с целью устранения аномалий и пустот в записях массивов, что может привести к неправильной работе обучаемой модели. Проведено обучение моделей на отобранных данных с целью выявления аномального поведения трафика, указывающего на конкретный тип атаки. Кроме того, проведено исследование на наборе данных, не содержащих сведений об атаках.

Ключевые слова: *атаки, аномальный трафик, системы обнаружения вторжений, машинное обучение, атака Infiltration, атаки на веб-приложения, Python*

Введение

Совершенствование систем обнаружения вторжений (IDS), использующих машинное обучение, зависит от доступности наборов данных, но получить надежный набор для обучения – не самая простая задача. Среди факторов, затрудняющих сравнение, – отсутствие надлежащего описания методов обнаружения вторжений, отсутствие методологии сравнения, а также отсутствие маркировки достоверных данных и сложность получения реального трафика [1]. Кроме того, в настоящее время сетевой трафик в основном шифруется в целях безопасности и конфиденциальности, и лишь очень немногие наборы отражают этот факт.

Набор данных является важной частью построения моделей IDS на основе машинного обучения. Процесс начинается с захвата трафика. После этого он компилируется в данные определенного типа, содержащие сетевые функции, включая маркировку. Общая структура систем обнаружения вторжений на основе машинного обучения показана в [2]. Маркировка – важнейший процесс для набора данных. Обработка достоверных данных является настоящей проблемой, особенно когда эксперты не могут определить, содержит ли трафик атаку или нет. Именно по этой причине специалисты используют синтетический трафик. Однако это означает, что генерируемый трафик не всегда является характерным для реальной среды. По этой причине процесс создания набора данных начинается с захвата трафика и заканчивается заключительным этапом предварительной обработки. Конечным результатом этапа предварительной обработки является размеченный набор данных. Каждая точка данных определяется как вредоносный или нормальный трафик. Файл содержит табличные данные в удобочитаемом формате, например файл CSV, или в двоичной форме, например файл IDX. Количество обнаруженных атак или ложных тревог можно использовать для сравнения набора данных при обучении IDS.

Большинство существующих исследований, в которых используется зашифрованный трафик, сосредоточены на различных областях, таких как классификация и анализ трафика. Хотя такие исследования существуют [3], набор данных не является общедоступным из-за конфиденциальности содержимого.

Наборы контрольных данных являются важной основой для оценки и сравнения качества различных IDS. В зависимости от методов обнаружения существует три типа IDS: на основе сигнатур, на основе аномалий (статистические) и комбинированные [4]. Основанные на сигнатурах фокусируются на построении автоматической генерации шаблона, тогда как основанные на аномалиях – на наблюдении отклонений от нормального поведения. Сигнатурные системы основаны на методе сопоставления с образцом для идентификации и попытки сопоставления с базой данных. Когда попытка атаки совпадает с шаблоном, выдается предупреждение. Системы на основе сигнатур имеют высокую точность определения аномалий и низкий уровень ложных срабатываний, но при этом они не могут обнаружить неизвестные атаки. При этом системы, основанные на статистических методах, могут обнаруживать неизвестные атаки путем сравнения аномального трафика с нормальным, соотношение ложных тревог остается достаточно высоким.

Системы, обученные с помощью алгоритмов машинного обучения, могут улучшить свою способность прогнозировать события, используя обратную связь относительно того, насколько хорошо они выполнили предыдущие задачи, и использовать эту информацию для внесения изменений [5].

Сравнение наборов данных

В представленной работе использовались готовые наборы данных с трафиком, содержащие атаки [6–9]. Сравнение наборов представлено в таблице 1.

На основе представленной сравнительной характеристики делается вывод о целесообразности использования набора CSE-CIC-IDS2018 по причине его большей актуальности, а также из-за объема содержащихся данных [7]. CSE-CIC-IDS2018 был получен в результате моделирования крупной компьютерной сети, содержащей 420 хостов и 30 серверов, разделенных на несколько сегментов. При этом трафик передавался по протоколам HTTP/HTTPS, SMTP/POP3/IMAP, SSH и FTP. Были реализованы атаки шести разновидностей: Brute Force, DoS, DDoS, Botnet, внедрение и атаки на веб-приложения.

Для обработки набора данных использовались библиотеки языка программирования Python: Pandas, Scikit-learn, TensorFlow, Matplotlib, Pickle.

Предобработанные данные выбранного набора отображаются в CSV-формате. Для представ-

Таблица 1. Сравнение наборов данных

Критерий	Набор данных			
	CIC-IDS2017	CSE-CIC-IDS2018	HIKARI-2021	USB-IDS-1
Объем	Средний	Большой	Малый	Малый
Источник	Синтетический, высокая вариация трафика и атак	Синтетический, высокая вариация трафика и атак	Синтетический, низкая вариация трафика и атак	Частично синтетический, низкая вариация трафика и атак
Доверие	Высокий уровень	Высокий уровень	Высокий уровень	Высокий уровень
Описание	Достаточное	Достаточное	Достаточное	Достаточное
Масштаб сети	Офисная сеть средних размеров	Большая корпоративная сеть	Небольшая локальная сеть	Простейшая архитектура

Таблица 2. Характеристики выборок для атак на Web-приложения

Характеристика	Выборка		
	Обучающая	Валидационная	Тестовая
Общее число записей	29 696	3712	3712
Число записей с меткой вредоносного трафика	14 848 (50,0%)	1890 (50,9%)	1890 (50,9%)
Число записей с меткой легального трафика	14 848 (50,0%)	1822 (49,1%)	1822 (49,1%)

ления данных в необходимом формате они подлежат определенной обработке: первичный анализ данных и предварительная обработка данных. На выходе обработки получается тот вид данных, который в дальнейшем будет использоваться для обучения моделей классификаторов. Для этого будут выделены три выборки: тестовая, обучающая и валидационная.

Так как в выбранном наборе данных приводятся более одной разновидности атаки одного из представленных типов, то они будут в итоге объединены в один класс. В итоге при обучении будет приниматься решение об отнесении потока данных к типовой атаке, не учитывая ее подвид. Кроме того, в наборе данных представлены разные соотношения трафика с атакой и нормального трафика. Для лучшего обучения классификаторов весь трафик будет приводиться к идеальной классовой сбалансированности. При этом будем использовать некоторые стратегии сэмплирования.

Составление выборок для атак на Web-приложения

Атаки на Web-приложения в наборе данных CSE-CIC-IDS2018 представлены тремя разными вариантами: XSS (Cross Site Scripting), Brute Force и SQL-инъекция. Объединим таблицы для последующей совместной обработки и произведем замену меток классов, выполнив уже известные преобразования. Получаем критический дисбаланс классов. Объем меньшего из классов мал, что подводит нас к использованию подхода пересемплирования балансировки классов. При этом будут дублироваться записи меньшего из классов, пока объемы классов не сравняются.

Сбалансированные данные пропустим через конвейер предобработки, сохраним параметры нормализации и разобьем данные на выборки. Характеристики полученных выборок сведены в таблицу 2.

Обучим классификатор на основе логистической регрессии со стандартными параметрами. В результате обучена модель со следующими значениями метрик качества [10] по тестовой выборке: аккуратность: 0,945043; точность: 0,905288; полнота: 0,996296; F-мера: 0,948615.

При попытке подобрать оптимальные гиперпараметры классификатора была получена более качественная модель. Сохраним подогнанную модель средствами библиотеки Pickle. Метрики качества классификатора с базовыми и оптимальными гиперпараметрами сведены в таблицу 3.

Обучим классификатор на основе решающего дерева со стандартными параметрами. В результате получена модель со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,999461; точность: 0,998943; полнота: 1,0; F-мера: 0,999471.

Осуществим перебор оптимальных параметров. При этом за счет меньшего размера выборок появилась возможность расширить сетку параметров без потерь вычислительной скорости. Полученный в результате перебора классификатор имеет те же значения метрик качества, что и базовый. Делаем вывод о том, что базовый вариант справлялся с задачей оптимальным образом.

Рассмотрим базовый вариант классификатора на основе нейронной сети. В результате получена нейронная сеть со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,987607; точность: 0,983735; полнота: 0,992063; F-мера: 0,987882.

При получении более аккуратной модели была уменьшена скорость обучения с 0,001 до 0,0005. Метрики классификаторов сведены в таблице 4. Оптимальную модель сохраним средствами библиотеки Pickle.

Характеристики лучших моделей каждого вида сведены в таблицу 5. В связи с порядком значений метрик качества полученных моделей валидационная выборка оказалась невостребованной во всех трех случаях.

На основании данных представленной выше таблицы можно утверждать, что оптимальной моделью для выявления атак на Web-приложения является модель на основе решающего дерева. Использование нейронной сети избыточно.

Составление выборок для атаки типа Infiltration

Для части набора данных, содержащих атаку типа Infiltration, произведем замену меток классов. В списке новых меток наблюдаем ярко выраженный дисбаланс классов. Объемы выборок достаточны для использования недосемплирования. При этом дублируются записи меньшего из классов, пока объемы классов не сравняются. В этом случае используются все данные.

Пропустим сбалансированные данные через конвейер предобработки, сохраним параметры нормализации и разобьем данные на выборки. Реализация данных преобразований аналогична

Таблица 3. Значения метрик качества классификаторов на основе логистической регрессии для атак на Web-приложения

Классификатор	Метрика качества			
	Аккуратность	Точность	Полнота	F-мера
Классификатор с базовыми параметрами	0,945043	0,905288	0,996296	0,948615
Классификатор с оптимальными параметрами	0,951509	0,913043	1,0	0,954545

Таблица 4. Значения метрик качества классификаторов на основе нейронной сети для атак на Web-приложения

Классификатор	Метрика качества			
	Аккуратность	Точность	Полнота	F-мера
Классификатор с базовыми параметрами	0,987607	0,983735	0,992063	0,987882
Классификатор с оптимальными параметрами	0,992457	0,989485	0,995767	0,992616

Таблица 5. Характеристики моделей для атак на Web-приложения

Модель	Метрика качества				Размер модели, кб
	Аккуратность	Точность	Полнота	F-мера	
Логистическая регрессия	0,951509	0,913043	1,0	0,954545	7
Решающее дерево	0,999461	0,998943	1,0	0,999471	16
Нейронная сеть	0,992457	0,989485	0,995767	0,992616	110

представленной выше. Характеристики полученных выборок сведены в таблицу 6.

Рассмотрим классификатор на основе логистической регрессии со стандартными параметрами. В результате обучения получена модель со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,555470; точность: 0,594545; полнота: 0,355796; F-мера: 0,445180.

Модель не справляется со своей задачей. Качество классификации близко к уровню случайного угадывания. При попытке подобрать лучшие гиперпараметры классификатора качество работы классификатора не улучшилось.

В результате обучения классификатора на основе решающего дерева со стандартными параметрами получена модель со следующими значениями метрик качества по тестовой выборке: аккуратность: 0,568221; точность: 0,570853; полнота: 0,558334; F-мера: 0,564524. Полученная модель не превосходит предыдущую по качеству.

Попробуем подобрать более осмысленную сетку параметров. Для этого определяются зависимости метрик качества классификатора от ключевых гиперпараметров модели на тестовой и валидационной выборках: в зависимости от глубины дерева, в зависимости от предела разветвления, в зависимости от критерия разветвления. На основе этих сведений составляется новая сетка и таким образом находится лучшая модель, которая обладает немного лучшими свойствами. Характеристики моделей сведены в таблицу 7.

При обучении классификатора на основе нейронной сети усложним ее архитектуру относительно представленной ранее путем добавления второго скрытого полносвязного слоя. Число нейронов в обоих скрытых слоях устанавливается равным 100. Базовая модель после обучения обладает значениями метрик качества по тестовой выборке: аккуратность: 0,571124; точность: 0,694168; полнота: 0,258100; F-мера: 0,376291.

Попробуем получить более аккуратную модель при помощи подхода, аналогичному представленному в обучении с решающим деревом, с определением метрик качества для ключевых параметров модели. Произведем перебор таких параметров. Лучшая из рассмотренных моделей была получена при использовании гиперболического тангенса как функции активации скрытых слоев с сохранением прочих гиперпараметров базового классификатора. Характеристики моделей сведены в таблицу 8.

Таблица 6. Характеристики выборок для атаки типа Infiltration

Характеристика	Выборка		
	Обучающая	Валидационная	Тестовая
Общее число записей	259 094	32 387	32 387
Число записей с меткой вредоносного трафика	129 622 (50,0%)	16 228 (50,1%)	16 234 (50,1%)
Число записей с меткой легального трафика	129 472 (50,0%)	16 159 (49,9%)	16 153 (49,9%)

Таблица 7. Значения метрик качества классификаторов на основе нейронной сети для атаки типа Infiltration

Классификатор	Метрика качества			
	Аккуратность	Точность	Полнота	F-мера
Классификатор с базовыми параметрами	0,568221	0,570853	0,558334	0,564524
Классификатор с подобранными параметрами	0,585821	0,588601	0,576999	0,582742

Таблица 8. Значения метрик качества классификаторов на основе нейронной сети для атаки типа Infiltration

Классификатор	Метрика качества			
	Аккуратность	Точность	Полнота	F-мера
Классификатор с базовыми параметрами	0,572266	0,709484	0,248367	0,367933
Классификатор с подобранными параметрами	0,550252	0,539627	0,699581	0,609281

Таблица 9. Характеристики моделей для атаки типа Infiltration

Модель	Метрика качества				Размер модели, кб
	Аккуратность	Точность	Полнота	F-мера	
Логистическая регрессия	0,555470	0,594545	0,355796	0,445180	7
Решающее дерево	0,585821	0,588601	0,576999	0,582742	489
Нейронная сеть	0,550252	0,539627	0,699581	0,609281	130

Таблица 10. Лучшие полученные модели машинного обучения для классификации атак

Атака	Тип классификатора	Метрика качества				Размер модели, кб
		Аккуратность	Точность	Полнота	F-мера	
Web	Решающее дерево	0,999461	0,998943	1,0	0,999471	16
Infiltration	Нейронная сеть	0,550252	0,539627	0,699581	0,609281	130

Характеристики лучших моделей каждого вида сведены в таблицу 9. На основании этих данных можно утверждать, что оптимальной моделью для выявления атак типа Infiltration является нейронная сеть. При этом результаты классификации даже для лучшей модели являются неудовлетворительными.

Лучшие модели для классификации рассмотренных атак, которые были получены в ходе работы, и значения их метрик качества по тестовым выборкам сведены в таблицу 10.

Для атак на Web-приложения были получены результаты, близкие к оптимальным. Атаку типа Infiltration удовлетворительно классифицировать на представленных данных и при помощи рассматриваемых алгоритмов с фиксированной архитектурой не удалось.

Также было проведено дополнительное исследование, при котором производилась попытка классификации данных в наборе, в котором отсутствовали какие-либо атаки. Все модели справились с задачей. И рассматриваемые атаки не были там обнаружены. При попытке обнаружения неизвестной классификатору атаки (например, Brute Force для атак на Web-приложения) модели ожидаемо не справились с задачей.

Заключение

Для улучшения результатов обучения можно использовать предварительную обработку данных иного вида, где будет осуществляться упор на решение задачи классификации конкретного вида атаки. При этом возможно значительное усложнение архитектуры моделей (особенно перспективны нейронные сети). Также возможен более тщательный перебор гиперпараметров моделей, связанный с повышением требований к вычислительным мощностям или времени обучения.

В дальнейшем предполагается использовать предложенные способы улучшения результатов. Кроме того, предполагается рассмотреть метрики качества для других видов атак с использованием тех же типов классификаторов: решающее дерево, логистическая регрессия, нейронная сеть.

Литература

1. Performance analysis of machine learning algorithms in intrusion detection system: a review / T. Saranya [et al.] // Procedia Computer Science. 2020. Vol. 171. P. 1251–1260.
2. Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic / A. Ferriyan [et al.] // Applied Sciences. 2021. Vol. 11, no. 17. URL: https://www.mdpi.com/2076-3417/11/17/7868 (дата обращения: 15.01.2024).
3. De Lucia M.J., Cotton C. Identifying and detecting applications within TLS traffic // Proceedings Cyber Sensing. 2018. Vol. 10630. P. 106300U.
4. Шелухин О.И., Сакалема Д.Ж., Филинова А.С. Обнаружение вторжений в компьютерные сети (сетевые аномалии): учебное пособие для вузов М.: Горячая линия – Телеком, 2022. 220 с.
5. Sharma S., Krishna C.R., Sahay S.K. Detection of advanced malware by machine learning techniques // Springer, Advances in Intelligent Systems and Computing. 2018. Vol. 742. P. 332–342.
6. CSE-CIC-IDS2018 on AWS. URL: http://www.unb.ca/cic/datasets/ids-2018.html (дата обращения: 18.12.2023).
7. Intrusion Detection Evaluation Dataset (CIC-IDS2017). URL: https://www.unb.ca/cic/datasets/ids-2017.html (дата обращения: 18.12.2023).

8. HIKARI-2021: Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic / A. Ferriyan [et al.]. URL: https://zenodo.org/records/5199540 (дата обращения: 20.01.2024).
9. USB-IDS-1: A public multilayer dataset of labeled network flows for IDS evaluation / M. Catillo [et al.] // 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W). 2021. P. 1–6. DOI: 10.1109/DSN-W52860.2021.00012
10. Харрисон М. Машинное обучение: карманный справочник. Краткое руководство по методам структурированного машинного обучения на Python / пер. В.А. Коваленко. СПб.: Диалектика, 2020. 320 с.

Получено 22.01.2024

Поздняк Ирина Сергеевна, к.т.н., доцент, доцент кафедры информационной безопасности (ИБ) Поволжского государственного университета телекоммуникаций и информатики (ПГУТИ). 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 927 657-24-27. E-mail: i.pozdnyak@psuti.ru

Макаров Игорь Сергеевич, к.т.н., доцент кафедры ИБ ПГУТИ. 443090, Российская Федерация, г. Самара, Московское шоссе, 77. Тел. +7 937 208-80-66. E-mail: i.makarov@psuti.ru

USING MACHINE LEARNING ALGORITHMS TO DETECT ANOMALOUS TRAFFIC BEHAVIOR

Pozdnyak I.S., Makarov I.S.

Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation

E-mail: i.pozdnyak@psuti.ru, i.makarov@psuti.ru

The article describes a method of using machine learning for detecting anomalous traffic behavior. For this purpose, a data set containing a significant amount of traffic collected at the time of the attack on the Web application is used. The set contains three attack options: Brute Force, XSS, SQL injection. A traffic dump containing an Infiltration attack is considered separately. A comparative analysis of machine learning models was carried out with the selection of the most optimal one. The article also provides a description of the data preprocessing procedure, which is carried out in order to eliminate anomalies and voids in array records, which can lead to incorrect operation of the trained model. Models were trained on selected data in order to identify anomalous traffic behavior indicating a specific type of attack. In addition, a study was conducted on a data set that does not contain information about attacks.

Keywords: *attacks, anomalous traffic, intrusion detection systems, machine learning, Infiltration attack, attacks on web applications, Python*

DOI: 10.18469/ikt.2023.21.3.04

Pozdnyak Irina Sergeevna, Povolzhskiy State University of Telecommunications and Informatics, 77, Moscovskoye shosse, Samara, 443090, Russian Federation; Associated Professor of Information Security Department, PhD in Technical Science. Tel. +7 927 657-24-27. E-mail: i.pozdnyak@psuti.ru

Makarov Igor Sergeevich, Povolzhskiy State University of Telecommunications and Informatics, 77, Moscovskoye shosse, Samara, 443090, Russian Federation; Associated Professor of Information Security Department. PhD in Technical Science. Tel. +7 937 208-80-66. E-mail: i.makarov@psuti.ru

References

1. Saranya T. et al. Performance analysis of machine learning algorithms in intrusion detection system: a review. *Procedia Computer Science*, 2020, vol. 171, pp. 1251–1260.
2. Ferriyan A. et al. Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. *Applied Sciences*, 2021, vol. 11, no. 17. URL: https://www.mdpi.com/2076-3417/11/17/7868 (accessed: 15.01.2024).

3. De Lucia M.J., Cotton C. Identifying and detecting applications within TLS traffic. *Proceedings Cyber Sensing*, 2018, vol. 10630, pp. 106300U.
4. Sheluhin O.I., Sakalema D.Zh., Filinova A.S. *Detection of Intrusions into Computer Networks (Network Anomalies): Textbook for Universities*. Moscow: Goryachaya liniya – Telekom, 2022, 220 p. (In Russ.)
5. Sharma S., Krishna C.R., Sahay S.K. Detection of advanced malware by machine learning techniques. *Springer, Advances in Intelligent Systems and Computing*, 2018, vol. 742, pp. 332–342.
6. CSE-CIC-IDS2018 on AWS. URL: http://www.unb.ca/cic/datasets/ids-2018.html (accessed: 18.12.2023).
7. Intrusion Detection Evaluation Dataset (CIC-IDS2017). URL: https://www.unb.ca/cic/datasets/ids-2017.html (accessed: 18.12.2023).
8. Ferriyan A. et al. HIKARI-2021: Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. URL: https://zenodo.org/records/5199540 (accessed: 20.01.20244).
9. Catillo M. et al. USB-IDS-1: A public multilayer dataset of labeled network flows for IDS evaluation. *51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops* (DSN-W), 2021, pp. 1–6. DOI: 10.1109/DSN-W52860.2021.00012
10. Harrison M. *Machine Learning: A Pocket Guide. A Quick Guide to Structured Machine Learning Methods in Python*. Transl. from English by V.A. Kovalenko. Saint Petersburg: Dialektika, 2020, 320 p. (In Russ.)

Received 22.01.2024

ТЕХНОЛОГИИ КОМПЬЮТЕРНЫХ СИСТЕМ И СЕТЕЙ

УДК 004.056

МОДЕЛИ И АЛГОРИТМЫ ОПРЕДЕЛЕНИЯ АКТУАЛЬНЫХ УГРОЗ НА ОСНОВЕ ЭКСПЕРТНОГО ПОДХОДА

Рычкова А.А.[1], Долгушев Н.А.[1,3], Бурькова Е.В.[1], Коннов А.Л.[1,2]
[1] Оренбургский государственный университет, Оренбург, РФ,
[2] Оренбургский филиал Поволжского государственного университета телекоммуникаций и информатики, Оренбург, РФ,
[3] Оренбургский филиал ООО «Уральский центр систем безопасности», Оренбург, РФ
E-mail: rnansy@yandex.ru, dolgushevn1@yandex.ru, tulpan63@bk.ru, andrey_konnov@mail.ru

Задача определения актуальных угроз безопасности приобретает все большее значение, что обусловлено ростом объемов обрабатываемой информации ограниченного доступа, увеличением количества всевозможных угроз, повышением потенциала нарушителей. На предприятиях и в организациях разных сфер экономики специалисты регулярно проводят аудит информационной безопасности с целью выявления уязвимостей и в конечном итоге для предотвращения возможных негативных последствий. Для построения рациональной системы безопасности информации на объекте важно организовать защиту именно от актуальных угроз, так как от всех угроз построить защиту невозможно, да и нецелесообразно. В статье предложены модели и алгоритмы определения актуальных угроз на основе экспертного подхода, которые могут стать базисом для разработки автоматизированной системы принятия достоверных решений в задачах защиты объектов информатизации.

Ключевые слова: *угроза безопасности, метод анализа иерархий, иерархическая модель*

Введение

Для определения актуальных угроз используют различные методы [2; 3; 7; 9], учитывают особенности для конкретных сфер деятельности [5], существуют также автоматизированные системы на основе программных продуктов [12]. Все эти методы основаны на положениях методики оценки угроз безопасности федеральной службы по техническому и экспортному контролю (ФСТЭК) России от 2021 года. Учитывается необходимость проведения основных этапов этого процесса: анализа защищаемых объектов, определения не-