5. *Nah S., Baik S., Hong S., et al.* NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study / 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – IEEE: Long Beach, CA, USA, 2019. – DOI: 10.1109/CVPRW.2019.00251.

6. *Soomro K., Zamir A.R., Shah M.* UCF101: A Dataset of 101 Human Action Classes from Videos in the Wild // arXiv:1212.0402. – 2012.

7. *Tancik M., Mildenhall B.; Ng R.* StegaStamp: Invisible Hyperlinks in Physical Photographs. – URL: https://ieeexplore.ieee.org/document/9156548 (дата обращения 1.09.2025).

8. *Hayes J., Danezis G.* Generating Steganographic Images via Adversarial Training // NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. – NY: Curran Associates Inc., 2017. – P. 1951-1960.

9. *Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P.* Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Transactions on Image Processing. – 2004. – Volume 13, Issue 4. – P. 600-612. – DOI: 10.1109/TIP.2003.819861.

10. *Goodfellow I., Pouget-Abadie J., Mirza M., et al.* Generative Adversarial Nets. – URL: https://archive.org/details/arxiv-1406.2661 (дата обращения 10.09.2025).

Михалевич И.Ф., Пчелинцев Д.И.

Использование регулярных выражений для управления информационной безопасностью интеллектуальных транспортных систем

Аннотация: Статья посвящена проблеме неполноты информации об инцидентах информационной безопасности интеллектуальных транспортных систем. Рассмотрены методы обработки и анализа данных из открытых источников, содержащих фрагментарную информацию о возможных угрозах функционирования

компьютеризированных систем интеллектуальных транспортных систем.

Ключевые слова: автоматизация, анализ угроз, валидация, информационная безопасность, компьютерная атака, нормализация, обработка данных, открытые регулярные выражения, уязвимости

Введение

Интегрированные системы корпоративного и технологического управления интеллектуальными транспортными системами функционируют в условиях постоянного роста количества и сложности компьютерных угроз [1, 2]. Ключевым элементом обеспечения информационной безопасности таких систем является оперативное выявление и устранение уязвимостей. Однако решение данной задачи затрудняется неполнотой информации об инцидентах компьютерной безопасности, имевших место в отрасли. Для преодоления данной трудности предлагается использовать методы сбора информации по косвенным признакам, коррелированным с соответствующими угрозами.

Гибридная система управления безопасностью интеллектуальных транспортных систем

В [3] предложена система гибридного управления рисками информационной безопасности интеллектуальных систем транспорта.

В базе данных уязвимостей и угроз накапливается информация, получаемая из системы мониторинга.

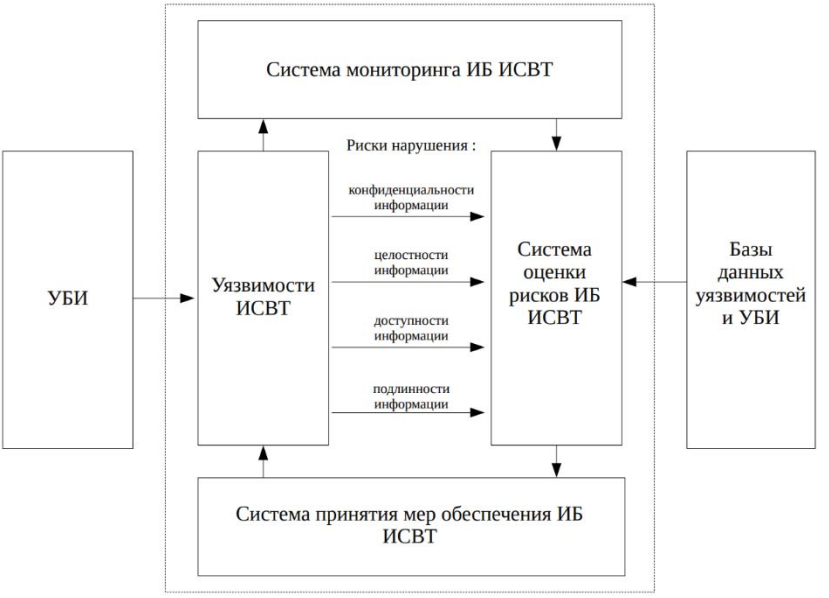

Рисунок 1 – Схема системы управления рисками информационной безопасности интеллектуальных систем транспорта

Полноту мониторинга угроз компьютерной безопасности, выявление информации об инцидентах, не получивших широкой огласки и не отраженных в официальных базах данных угроз и уязвимостей, позволяет использование регулярных выражений.

Использование регулярных выражений для управления информационной безопасностью

Одной из значительных проблем является обработка большого объема разнородных данных об уязвимостях, поступающих из различных источников: сканеров безопасности, логов приложений, баз данных (таких как NVD – National Vulnerability Database) и отчетов об инцидентах. Нестандартизированные, избыточные или некорректно оформленные записи об уязвимостях затрудняют их автоматизированную обработку, повышают риск пропуска критических угроз и снижают эффективность систем защиты интеллектуальных систем водного транспорта [9, 12].

Проблема усугубляется тем, что разнородность форматов представления данных об уязвимостях (например, различные варианты написания идентификаторов CVE, упоминания CWE в свободной текстовой форме) затрудняет их автоматическую корреляцию и анализ. Например, некорректно извлеченный идентификатор CVE может привести к неверной оценке критичности уязвимости или несрабатыванию правил корреляции в SIEM-системе. Таким образом, отсутствие строгой нормализации и валидации на этапе предобработки данных создает системную уязвимость, повышая риски реализации атак, связанных с несвоевременным обнаружением и устранением уязвимостей в интеллектуальных системах водного транспорта.

В области информационной безопасности интеллектуальных систем водного транспорта вопросам валидации и санации входных данных об уязвимостях уделяется первостепенное значение, поскольку они являются первым рубежом защиты от некорректной оценки угроз и несвоевременного принятия мер. Известны подходы, основанные на применении строгих шаблонов ввода (белый список), синтаксического анализа и, что наиболее распространено, на использовании формальных языков описания шаблонов – регулярных выражений [5]. Существующие работы в основном сосредоточены на защите от эксплуатации известных уязвимостей в ПО транспортных систем, однако вопрос комплексной нормализации разнородных данных об уязвимостях (CVE, CWE, CVSS) из различных источников – логов бортовых систем, отчетов сканеров безопасности, баз данных уязвимостей – с точки зрения минимизации рисков их некорректной обработки и корреляции остается недостаточно изученным. Особенно остро ощущается недостаток практических решений, которые сочетали бы в себе приведение данных к стандартному виду и механизмы их верификации, направленные на выявление и исключение потенциально ложных или некорректных записей, что крайне важно для оперативного реагирования на инциденты в реальном времени в системе управления рисками информационной безопасностью интеллектуальных систем водного транспорта.

В рамках решения обозначенных проблем настоящее исследование сфокусировано на разработке методов и алгоритмов предварительной обработки данных об уязвимостях, которые

позволяют снизить риски, связанные с их некорректной интерпретацией и потенциальным пропуском критических угроз. Предлагаемый подход предусматривает создание унифицированного этапа проверки и преобразования данных, обеспечивающего их соответствие строгим паттернам и стандартам (таким как форматы CVE и CWE), что в конечном итоге способствует повышению общей эффективности систем анализа угроз и управления уязвимостями.

Представленные в статье методы позволяют усилить защиту информационных систем интеллектуальных систем водного транспорта за счет обеспечения целостности и корректности обрабатываемых данных об уязвимостях, что способствует соблюдению требований регуляторов и отраслевых стандартов и снижению рисков, связанных с нарушениями в работе систем, обрабатывающих критически важную информацию.

Нормализация данных – это процесс приведения информации к единому, стандартному формату, устраняющему избыточность, противоречивость и аномалии ввода [13]. В контексте обработки данных об уязвимостях нормализация решает задачи унификации, очистки, валидации и структурирования поступающей информации. Теоретической основой нормализации является необходимость обеспечения непротиворечивости данных, что напрямую влияет на их пригодность для автоматизированного анализа и корреляции [8].

Ключевыми критериями нормализации являются:

1. Унификация формата: Приведение всех записей к единому шаблону (например, CVE к формату CVE-YYYY-NNNNN, CWE к CWE-NNN).

2. Очистка от избыточности: Удаление лишних пробелов, символов, не несущих смысловой нагрузки (дефисы, скобки, точки), а также дублирующих записей.

3. Валидация: Проверка данных на соответствие логическим и синтаксическим правилам (наличие корректного года и идентификатора в CVE, цифрового кода в CWE).

4. Структурирование: Разделение составных данных на логические компоненты (например, разделение ФИО на отдельные поля «Фамилия», «Имя», «Отчество»).

Эффективность информационной системы в значительной степени определяется двумя критически важными критериями:

оперативностью (скоростью обработки и получения информации) и достоверностью (точностью и корректностью данных).

До нормализации данные находятся в неструктурированном состоянии, что снижает оперативность обработки. Алгоритмам анализа приходится учитывать множество форматов. Например, поиск уязвимости «CVE-2021-44228» не найдет запись «CVE202144228» или «CVE - 2021 – 44228», что потребует дополнительных преобразований. Также достоверность результатов ставится под угрозу. Некорректные данные повышают риск ошибок. Система может пропустить критическую уязвимость из-за опечатки в идентификаторе.

После применения процедуры нормализации оперативность возрастает. Стандартизированные данные позволяют использовать простые и быстрые алгоритмы поиска и сопоставления. Обеспечивается достоверность. Валидация на этапе нормализации отсеивает заведомо некорректные записи, а унификация формата предотвращает ошибки логического уровня [4, 6].

Таким образом, нормализация выступает не только как метод приведения данных к единому виду, но и как ключевой механизм обеспечения высокой скорости обработки и гарантии достоверности информации в интеллектуальных системах водного транспорта.

Обеспечение безопасности обработки персональных данных требует не только теоретического обоснования, но и практической реализации конкретных механизмов контроля целостности и корректности информации. Предлагаемый подход основывается на применении регулярных выражений как универсального инструмента валидации и нормализации разнородных данных, поступающих из различных источников [5, 7]. Использование данного метода позволяет создать единый конвейер предобработки, обеспечивающий стандартизацию формата и отсев потенциально опасных или некорректных записей на ранних этапах работы с информацией.

В качестве ключевого примера выбран стандарт CWE (Common Weakness Enumeration), поскольку он обеспечивает систематизированное описание распространенных уязвимостей программного обеспечения и потенциальных векторов атак, что особенно критично для комплексной оценки угроз в интеллектуальных транспортных системах. В различных источниках

данные об уязвимостях интеллектуальных транспортных систем могут иметь вариативные форматы записи, такие как «CWE-79», «CWE79» или «CWE 79», что осложняет их автоматическую обработку и корреляцию. Для обеспечения достоверности анализа и предотвращения ошибок идентификации необходима нормализация записей CWE к стандартному формату.

Листинг функции нормализации идентификаторов CWE из текста представлен на рисунке 2 и выполнен на языке Python.

```
# Функция нормализации CWE
import re

def extract_and_normalize_cwe(text):
    pattern = r'\bCWE[^a-zA-Z0-9]*(\d+)\b'
    matches = re.findall(pattern, text, re.IGNORECASE)
    normalized_cwes = []
    for num_str in matches:
        # Убираем ведущие нули и формируем стандартный формат
        normalized_id = f"CWE-{int(num_str)}"
        normalized_cwes.append(normalized_id)

    # Удаляем дубликаты, сохраняя порядок
    seen = set()
    unique_cwes = []
    for cwe in normalized_cwes:
        if cwe not in seen:
            seen.add(cwe)
            unique_cwes.append(cwe)
    return unique_cwes

# Пример использования
text = """
Уязвимости в системе:
CWE-  1,
CWE   -   305,
CWE  -3,
CWE44,
CWE.51,
CWE/61,
CWE#71,
CWE-1 (дубликат),
CWE_83
CWE94
CWE=10
"""
print(extract_and_normalize_cwe(text))# ['CWE-1', 'CWE-305', 'CWE
-3', 'CWE-4', 'CWE-5', 'CWE-6', 'CWE-7', 'CWE-8', 'CWE-9', 'CWE-1
0']
```

Рисунок 2 – Нормализация CWE

Реализованная функция: извлекает и нормализует идентификаторы CWE из текста, приводит их к стандартному формату: CVE-N, обрабатывает различные форматы записи (включая варианты с пробелами), возвращает уникальные значения без дубликатов.

Аналогичным образом, как и в представленных примерах, можно обрабатывать и другие данные о компьютерных угрозах, такие как оценки CVSS и другие.

Использование регулярных выражений позволяет автоматизировать обработку данных об уязвимостях в системе управления рисками информационной безопасности интеллектуальных систем водного транспорта, обеспечивая унификацию их представления и снижение количества ошибок при анализе независимо от источника получения информации. Данный подход значительно ускоряет процесс идентификации критических уязвимостей в системах, что особенно важно для оперативного принятия решений в условиях реального времени. Кроме того, автоматизированная обработка помогает минимизировать человеческий фактор.

Заключение

Рассмотренные методы могут быть использованы в системах управления уязвимостями и SIEM-решениях для автоматизации процесса анализа угроз и снижения рисков информационной безопасности интеллектуальных систем транспорта.

Работа выполнена за счет бюджетного финансирования в рамках государственного задания от 20.03.2025 № 103-00001-25-02

Литература:

1. *Baranov L.A*., *Mikhalevich I.F.* On trust in autonomous shipping systems of critical information infrastructure // T-Comm. – 2025. – Vol. 19, No. 6. – P. 52-59. – DOI: 10.36724/2072-8735-2025-19-6-52-59.

2. *Михалевич И.Ф*. Концептуальные проблемы транспортной безопасности водных интеллектуальных транспортных систем // Надежность. – 2024. – Т. 24, № 2. – С. 72-87. – DOI: 10.21683/1729-2646-2024-24-2-72-87.

3. *Баранов Л.А.*, *Иванова Н.Д., Михалевич И.Ф.* Нечеткая система оценки рисков информационной безопасности интеллектуальных систем водного транспорта // Автоматика на транспорте. – 2024. – Т. 10, № 1. – С. 7-17. – DOI: 10.20295/2412-9186-2024-10-01-7-17.

4. *Мэтиас Э*. Программирование на Python. – М.: ДМК Пресс, 2020. – 432 с.

5. *Фридл Дж*. Регулярные выражения. 3-е изд. – СПб.: Питер, 2018. – 544 с.

6. *Лутц М.* Изучаем Python. 5-е изд. – СПб.: Символ-Плюс, 2022. – 992 с.

7. *Scarfone K*. Guide to Enterprise Patch Management Technologies. – NIST Special Publication 800-40, Rev. 4. – NIST, 2022.

8. *Басов В.А.* Прикладные математические методы предварительного анализа и обработки текстовой образовательной информации // Интеллектуальные информационные системы: теория и практика: Сборник научных статей по материалам IV Всероссийской с международным участием конференции, Курск, 21–23 ноября 2023 года. – Курск: Курский государственный университет, 2023. – С. 7-12.

9. MITRE. CVE® Program Vision / MITRE Corporation. – 2023. – URL: https://cve.mitre.org/ (дата обращения 1.09.2025).

10. *Иванова Г.С., Мартынюк П.А.* Анализ методов предобработки текстовых данных // Искусственный интеллект. Теория и практика. – 2023. – № 4(4). – С. 20-31.

11. *Абрамов П.С.* Извлечение ключевой информации из текста // Новые информационные технологии в автоматизированных системах. – 2018. – № 21. – С. 217-219.

12. ГОСТ Р ИСО/МЭК 27035-1-2023. Информационная безопасность. Управление инцидентами информационной безопасности. Часть 1. Принципы и процессы.