А. М. Каимов, К. В. Таран, Т. И. Паюсова
Тюменский государственный университет, г. Тюмень
УДК 004.056.53

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ В DLP-СИСТЕМАХ ДЛЯ ЗАЩИТЫ ОТ УТЕЧЕК ИНФОРМАЦИИ В ФИНАНСОВОМ СЕКТОРЕ

Аннотация. В данной статье рассматривается ситуация с утечками из финансового сектора. Была подобрана статистика по утечкам, методам защиты информации, а также рассмотрены будущие исследования в этой области.

Ключевые слова: DLP-системы, машинное обучение, утечка информации, конфиденциальная информация.

Введение. На момент 2020 г. доля утечек в результате действий внешних сил в России выросла с 8,7 до 17,9% [1] как следует из аналитического отчета от InfoWatch, далее ситуация изменилась следующим образом: с 17,9 до 21,4% и с 21,4 до 75% [2] на момент 2021 и 2022 гг. соответственно. Для наглядности нами был сделан график данной статистики (рис. 1). Так же мы использовали и научные статьи для улучшения качества нашей статьи, так, например мы изучили книгу, в одной из ее глав обсуждаются определение утечки данных, последствия и причины утечки данных, а также многочисленные подходы к предотвращению и обнаружению утечки данных [3]. Так же, в исследовании рассматривается использование DLP-системы для обеспечения безопасности конфиденциальной информации в организации. Предлагается методика адаптации DLP-системы к особенностям деятельности компании [4]. Далее мы, используя популярный журнал нашли в нем подходящие для нашей работы статьи, в одном из них представлены результаты использования методов машинного обучения для выявления нетипичного поведения сотрудников банка при использовании электронной почты [5]. Рассматривается актуальная проблема деструктивной социальной инженерии, которая представляет угрозу экономической безопасности. Были предложены меры по ее предотвращению, включая обучение сотрудников, использование технических средств защиты

информации и контроль доступа к конфиденциальным данным [6]. С технической точки зрения искусственный интеллект в DLP-системы можно внедрить несколькими способами, для каждой организации рекомендуется использовать более подходящий метод, в зависимости от специфики самой организации, которой требуется такая система [7].

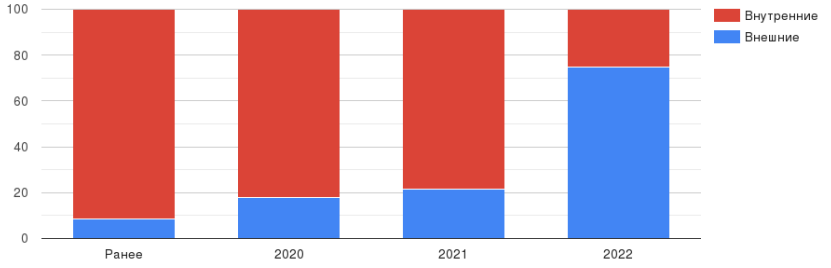

Рис. 1. Распределение утечек в зависимости от внешнего и внутреннего воздействия

Проблема исследования. Практика показывает, что причины утечки и утраты информации могут быть самыми разнообразными. Более важная информация всегда имеет приоритет по защите. Очевидно, что любая кража или модификация информации может привести к серьезным убыткам, однако не стоит забывать, что существует возможность потерять управление объектом. Однако какими методами организации могут пользоваться для защиты от утечек? Мы нашли много способов для реализации, но лишь один привлек наибольшее внимание — это DLP-системы (Data Leak Prevention).

Материалы и методы. DLP-системы являются неотъемлемой частью обеспечения безопасности конфиденциальной информации, DLP непрерывно отслеживают и анализируют поток данных, передаваемых по различным каналам связи, включая корпоративную и электронную почту, интернет-сервисы и системы обмена мгновенными сообщениями. Схема стандартной DLP-системы представлена на рис. 2.

Рис. 2. Схема стандартной DLP-системы

Стандартные DLP-системы имеют минусы, описанные ниже:

Один из недостатков заключается в огромном количестве ложных срабатываний, которые возникают из-за загруженных категорий и классов конфиденциальности. Это увеличивает количество детектируемых документов и требует дополнительных ресурсов для их анализа.

Еще одним недостатком стандартных DLP-систем является то, что они работают в режиме копирования, а не блокировки утечек. Это означает, что многие случаи утечки информации не могут быть предотвращены и расследуются только после того, как произошла утечка. Такой подход может приводить к серьезным потерям для организации, поскольку важная конфиденциальная информация может быть скомпрометирована до того, как ее утечка будет обнаружена и пресечена.

Для преодоления этих недостатков можно использовать более продвинутые методы анализа данных, такие как машинное обучение и искусственный интеллект. Эти методы могут помочь уменьшить количество ложных срабатываний и повысить точность обнаружения утечек. Кроме того, они могут позволить работать в режиме

блокировки утечек, что повышает эффективность защиты конфиденциальной информации.

Машинное обучение является ключевой технологией, которая позволяет улучшить эффективность и точность DLP-систем, а также сделать их более автоматизированными. Машинное обучение — это метод искусственного интеллекта, который позволяет компьютеру обучаться на основе опыта и данных. В процессе обучения система анализирует большие объемы данных и находит в них закономерности и шаблоны, которые в дальнейшем используются для принятия решений и решения задач. Преимущества модели DLP-системы с использованием машинного обучения включают повышение точности обнаружения утечек данных, адаптацию к новым типам угроз и методам слива информации, снижение нагрузки на пользователей и оптимизацию использования ресурсов.

Интеграция искусственного интеллекта (ИИ) с системой защиты от потери данных (DLP) может быть выполнена несколькими способами:

Первый способ — это интеграция через API, которая позволяет двум системам обмениваться данными и выполнять действия. Например, система DLP может использовать API для отправки данных на платформу ИИ для анализа и получения результатов в режиме реального времени.

Второй способ — это интеграция моделей машинного обучения в систему DLP для классификации конфиденциальных данных и обнаружения угроз. Модели могут быть обучены на больших наборах данных и использоваться для анализа сетевого трафика в реальном времени.

Третий способ — это интеграция с «озером» данных, централизованным хранилищем для хранения больших объемов данных, что позволяет алгоритмам ИИ получить необходимые данные для анализа.

Четвертый способ — это интеграция с облачными платформами ИИ, что позволяет организациям использовать возможности ИИ без необходимости в специализированном оборудовании или опыте.

Цель интеграции ИИ с DLP заключается в том, чтобы предоставить системе DLP интеллектуальные возможности и возможности

для лучшей защиты конфиденциальных данных. Конкретный метод интеграции будет зависеть от конкретных требований организации и возможностей используемых решений DLP и ИИ.

Результаты. Таким образом, DLP-системы с машинным обучением являются более эффективными инструментами защиты конфиденциальной информации, которые позволяют автоматизировать процесс мониторинга и анализа данных, что уменьшает нагрузку на сотрудников и повышает эффективность работы системы.

Заключение. В свете растущей доли утечек конфиденциальной информации в результате действий внешних сил, использование DLP-систем с машинным обучением и интеграцией искусственного интеллекта становится все более необходимым. Более продвинутые методы анализа данных позволяют уменьшить количество ложных срабатываний и повысить точность обнаружения утечек, а также работать в режиме блокировки утечек, что повышает эффективность защиты конфиденциальной информации. Интеграция машинного обучения с DLP-системами может быть выполнена через API, интеграцию моделей машинного обучения в систему DLP, интеграцию с озером данных или облачными платформами искусственного интеллекта. DLP-системы с машинным обучением являются более эффективными инструментами защиты конфиденциальной информации, что уменьшает нагрузку на сотрудников и повышает эффективность работы системы.

СПИСОК ЛИТЕРАТУРЫ

1. Исследование утечек конфиденциальной информации из организаций финансового сегмента в 2020 году. — Текст : электронный // InfoWatch : [сайт]. — URL: https://www.infowatch.ru/sites/default/files/analytics/files/InfoWatch_%D1%84%D0%B8%D0%BD%D0%B0%D0%BD%D1%81%D1%8B_2020_%D0%BE%D1%82%D1%87%D0%B5%D1%82.pdf (дата обращения: 31.05.2023).
2. Исследование утечек конфиденциальной информации в финансовом секторе Мир-Россия, 2022 г.. — Текст : электронный // InfoWatch : [сайт]. — URL: https://www.infowatch.ru/sites/default/files/analytics/files/finansovaya-sfera-utechki-informatsii-za-2022-god.pdf (дата обращения: 31.05.2023).

3. Sustainable Computing / V. Singh, M. Raj, I. Gupta, M. A. Sayeed. — Berlin: Springer, 2023. — 352 p. — Text : direct.
4. Андриянова Т. А. DLP: снижение риска утечки конфиденциальной информации Банка / Т. А. Андриянова, С. Б. Саломатин. — Текст : непосредственный // Системный анализ и прикладная информатика. — 2017. — № 4. — С. 76-81.
5. Domashova, J. Procedia Computer Science / J. Domashova, E. Bystrova. — 213. — Moscow : Elsevier, 2022. — 832 p. — Text : direct.
6. Деструктивная социальная инженерия как угроза экономической безопасности: масштабы явления и меры предотвращения / Л. В. Санина, О. А. Чепинога, Э. А. Ржепка, О. Ю. Палкин. — Текст : непосредственный // Baikal Research Journal. — 2021. — № 12. — С. 1-15.
7. Мартиросян, В. В. Применение искусственного интеллекта в dlp-системах / В. В. Мартиросян, Н. В. Медведев. — Текст : непосредственный // 16-я Международная научно-техническая конференция молодых ученых и студентов. — 2023. — № 1. — С. 299-302.