нападения, позволяют более точно построить вектор атак, связанных с навязыванием ложных обучающих данных, а также модифицированием классифицирования. При этом выявляется тенденция: максимально снижать стоимость «отравления данными» и поддерживать относительно низкие затраты на легальное взаимодействие.

Список использованных источников

1. Горбачев И.Е., Глухов А.П. Моделирование процессов нарушения информационной безопасности критической инфраструктуры// Труды СПИИРАН. – Москва, 2015. – Вып. 1(38). – С. 112 – 135.

2. Панченко А.А. Анализ подходов к построению системы защиты информации на базе модели процесса обработки данных / А.А, Панченко, М.В. Аникиенко, В.Н. Пржегорлинский // Вестник Рязанского гос. радиотехнического ун-та. – 2005. – № 16. – С. 120–123.

3. Горюнов М. Н. Синтез модели машинного обучения для обнаружения компьютерных атак на основе набора данных CICIDS2017 / М.Н. Горюнов, А.Г. Мацкевич, Д.А. Рыболовлев // Труды ИСП РАН. – 2020. – Т. 32, вып. 5. – С. 81–94.

4. Горюнов М.Н. Оценка применимости методов машинного обучения для обнаружения компьютерных атак / М.Н. Горюнов, А.А. Рыболовлев, Д.А. Рыболовлев // Информационные системы и технологии (Орел). – 2020. – № 6. – С. 103–111.

5. Кохендерфер М., Уилер Т., Рэй К. Алгоритмы принятия решений / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2023. – 684 с.

6. Методика оценки угроз безопасности информации Методический документ ФСТЭК России: утв. ФСТЭК России России 5 февраля 2021 г.

Подтопельный Владислав Владимирович, Ст. преп. Института цифровых технологий государственного технического университета (КГТУ), ionpvv@mail.ru.

УДК 004.056.52; 004.891.3

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ПОВЕДЕНЧЕСКИХ ПАТТЕРНОВ В DLP-СИСТЕМАХ

Е.А. Марченко, С.В. Жуков
«Самарский национальный исследовательский университет имени академика С.П. Королёва», г. Самара

Ключевые слова: DLP-система, система контроля, рекуррентная нейронная сеть, нейросетевые технологии, поведенческий анализ.

DLP-системы (Data Loss Prevention) - это специализированные программные решения, разработанные для предотвращения утечек конфиденциальных данных, обеспечения соблюдения правил безопасности и защиты информации в организациях. Они играют ключевую роль в предотвращении утечек конфиденциальной информации, соблюдении нормативных требований по защите данных и минимизации рисков

безопасности. DLP-системы обеспечивают контроль за передачей, использованием и хранением конфиденциальных данных, обнаруживают аномальные или несанкционированные действия пользователей и предпринимают меры по их блокированию или предотвращению. В современном цифровом мире, где данные становятся все более ценным активом, DLP-системы являются неотъемлемой частью информационной безопасности и обеспечивают защиту организаций от потенциальных угроз безопасности [1].

Анализ поведенческих паттернов имеет важное значение для повышения эффективности систем защиты конфиденциальных данных. Этот анализ помогает выявлять аномальное или необычное поведение пользователей, которое может свидетельствовать о потенциальных угрозах безопасности [2]. Путем анализа типичного поведения пользователей и выявления отклонений от него DLP-системы могут более точно идентифицировать потенциальные угрозы безопасности данных. Кроме того, анализ поведения помогает принимать решения на основе контекста, учитывая, например, время суток, местоположение пользователя или тип используемого устройства. Это способствует более гибкой адаптации политики безопасности к изменяющимся условиям. Поведенческий анализ также важен для выявления внутренних угроз, таких как злоумышленные действия или утечка данных из-за недосмотра или ошибок [3].

Для модуля поведенческого анализа в DLP-системе были выбраны рекуррентные нейронные сети (RNN).

Архитектура LSTM сети может включать несколько слоев рекуррентных блоков, каждый из которых содержит набор нейронов с активацией, памятью и вентилями, регулирующими поток информации через блок. Такая сеть может быть обучена на больших наборах данных, включающих информацию о действиях пользователей в сети, таких как вход и выход из системы, доступ к файлам, перемещение по файловой системе и другие действия.

Кроме того, в архитектуру включены слои для обработки других типов данных, таких как текстовая информация о действиях пользователя или метаданные файлов. Для них используются сверточные нейронные сети (CNN) для анализа структуры данных, например, изображений или аудиофайлов, если они присутствуют в системе.

Для обучения LSTM-сети в модуле применяются методы глубокого обучения, такие как обратное распространение ошибки (backpropagation) с использованием метода стохастического градиентного спуска или его вариаций. В ходе обучения сеть адаптируется к данным поведенческих шаблонов, путем корректировки весовых коэффициентов между нейронами на каждом временном шаге. Это позволяет сети выявлять сложные зависимости в последовательностях действий пользователей.

Для оценки эффективности и точности модели LSTM-сети в DLP-системах применяются различные методы оценки производительности, включая метрики классификации, такие как точность (accuracy), полнота (recall), точность (precision) и F1-мера. Также используются методы кросс-валидации, разделения данных на обучающий, валидационный и тестовый наборы, для оценки обобщающей способности модели и предотвращения переобучения. Кроме того, для оценки эффективности модели применяются методы анализа кривых обучения и валидации, такие как графики потерь (loss) и метрик производительности в зависимости от числа эпох обучения.

В целом, эти методы обеспечивают возможность оценки и сравнения различных вариантов моделей LSTM-сетей для модуля поведенческого анализа в DLP-системах, а также оптимизацию их параметров для достижения наилучшей производительности и точности в обнаружении аномалий и угроз безопасности данных.

LSTM-сети представляют собой мощный инструмент для модуля поведенческого анализа в DLP-системах, благодаря своей способности адаптироваться к последовательностям данных и выявлять сложные зависимости в поведении пользователей. Их эффективность и точность делают их идеальным выбором для обнаружения аномалий и угроз безопасности данных в реальном времени.

Список использованных источников

1. Morozov, V. DLP Systems as a Modern Information Security Control / V. Morozov, N. Miloslavskaya. — DOI: https://doi.org/10.1007/978-3-319-63940-6_42 // Biologically Inspired Cognitive Architectures (BICA) for Young Scientists. BICA 2017. — Moscow, Russia: Springer Cham, 2017. — P. 296-301 — URL: https://link.springer.com/chapter/10.1007/978-3-319-63940-6_42 (дата обращения: 18.03.2024).

2. Власова, А.В. Анализ принципов работы систем поведенческого анализа поведения пользователей и сущностей / А.В. Власова, В.А. Дударев, Т.И. Новикова // Fundamental and applied approaches to solving scientific problems. — 2023. — С. 232-236. — URL: https://elibrary.ru/item.asp?id=50077027 (дата обращения: 18.03.2024).

3. Савенков, П.А. Использование методов и алгоритмов анализа данных в мобильной UEBA/DSS-системе для решения задач информационной безопасности / П.А. Савенков // Известия ТулГУ. Технические науки. — 2019. — №12. — URL: https://cyberleninka.ru/article/n/ispolzovanie-metodov-i-algoritmov-analiza-dannyh-v-mobilnoy-ueba-dss-sisteme-dlya-resheniya-zadach-informatsionnoy-bezopasnosti (дата обращения: 18.03.2024).

Марченко Екатерина Александровна, студент гр. 6304-010302D, kate_unmatched@mail.ru.

Жуков Семен Викторович, старший преподаватель каф. геоинформатики и информационной безопасности, zhukovsv91@inbox.ru.