Section 1. Information technology

Zolotukhina Maria Alexandrovna,
Postgraduate student
of the Department of Digital Data Processing Technologies
MIREA – Russian Technological University

Scientific supervisor: Zykov Sergey Viktorovich,
Doctor of Technical Sciences, Associate Professor,
Professor of the Department of Software Engineering
National Research University Higher School of Economics,
MIREA – Russian Technological University

APPLICATION OF EXPERT ASSESSMENT OF VULNERABILITIES IN ENTERPRISE DATA

Abstract. The article focuses on the current goals of research in the field of cybersecurity. The decision-making system used and examples of classification of expert assessments are given. The assessment of the relevance of the knowledge base for machine learning algorithms is determined. Examples of building diagnostic images to identify vulnerability incidents are given. The necessity of establishing the data structure of diagnostic information is revealed.

Keywords: data mining, expert systems, machine learning, security, diagnostic rules, knowledge base, datasets.

Золотухина Мария Александровна,
Аспирант кафедры цифровые технологии обработки данных МИРЭА
Российский технологический университет

Научный руководитель: Зыков Сергей Викторович,
д.т.н., доцент, профессор департамента программной инженерии
Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЕ),
МИРЭА – Российский технологический университет

ПРИМЕНЕНИЕ ЭКСПЕРТНОЙ ОЦЕНКИ УЯЗВИМОСТЕЙ В ДАННЫХ НА ПРЕДПРИЯТИИ

Аннотация. В статье внимание обращено на актуальные цели исследования в области кибербезопасности. Приведена используемая система принятия решений и примеры классификации экспертных оценок. Определена оценка востребованности базы знаний для алгоритмов машинного обучения. Приведены примеры построения диагностических образов для выявления инцидентов уязвимостей. Выявлена необходимость установления структуры данных диагностической информации.

Ключевые слова: интеллектуальный анализ данных, экспертные системы, машинное обучение, безопасность, диагностические правила, база знаний, датасеты.

Введение

Каждое предприятие работает с данными, которые требуют повышенного контроля. Из-за уязвимостей, на предприятиях повышается степень угрозы скрытых атак на данные, следовательно, свою работоспособность теряют все структуры отделов. Применение экспертной системы как аналитико-прогностического метода позволит предотвратить утечку данных, благодаря интеллектуальному анализу диагностической информации.

Аналитическая обработка данных предполагает интеллектуальный разбор метрик с применением алгоритмов машинного обучения. Требуется создание экспертной системы, чтобы осуществить качественный анализа диагностической информации [1]. База знаний формируется из аномалий обнаруженных при сканировании скрытых угроз с применением оценки узкопрофильного специалиста в области кибербезопасности и управления информационными системами. Исследование диагностической информации уязвимостей можно начать с рассмотрения распространенных источников таких как Kaggle, Data world, Datasetsearch и т.д. Датасеты являются источниками знаний и правил.

Выбор набора данных зависит от цели применения разрабатываемой системы. Также стоит учесть признаки, которые зависят от специфики узкопрофильного эксперта в области обнаружения инцидентов кибербезопасности.

Постановка задачи

Создание единой базы знаний для структурирования признаков скрытых атак и их описаний в информационных системах предприятий является главным критерием повышения точности модели машинного обучения. Первая и наиболее важная задача – это осуществление подбора критериев внешних состояний. Следовательно, разработка и внедрение информационного образа и системы представления знаний в информационные системы обеспечения безопасности данных на предприятиях. Для этого следует организовать процедуру получения качественных параметров при реализации ИО. Степень устойчивости алгоритма к наличию в данных аномальных значений будет являться робастностью системы, что благотворно влияет на достоверность модели машинного обучения и предполагает до 98,9% качества. При построении

структуры модели машинного обучения все внимание отдают показателю качества выполнения алгоритма. Если при выполнении задачи функция имеет значение ниже 95%, то система обработки данных идет на модернизацию алгоритма [2]. Построение пайплайна потребует достаточно много времени из-за большого количества данных и их обработки.

Целью создания пайплайна представляет собой разработку многофункциональной системы управления безопасностью данных на предприятии. Определение новых параметров и критериев отбора при анализе датасетов позволит конкретизировать объект исследования. Для дальнейшего развития системы поиска нужно определить место в организации корпорации. Учет всех факторов индивидуальных состояний структур предприятий обуславливает качественную защиту данных и внутренних процессов. Исследование данного вопроса позволит без потерь внедрять систему поиска уязвимостей. Качественный анализ данных завершит глубокое обучение на стадии поиска оптимального алгоритма [3].

Решение проблемы

Чтобы систематизировать диагностические правила для базы знаний потребуется введение новых значений уязвимостей и привлечение экспертной оценки. Для этого все внимание будет нацелено на исследование датасетов и ПО идентификации атак, показано в таблице 1. Приведение новых параметров к единому стандарту записи и усовершенствование правил, введение новых алгоритмов и аналитико-прогностических систем позволит качественно и быстро осуществить внедрение в структуры предприятия для обнаружения скрытых атак. Для определенных параметров выделим информативные критерии и информативные диапазоны критерий. Найденные значения примут вид качественных показателей, как показано в таблице 1.

Таблица 1.– Пример динамических диагностических правил

№	P_0	P_1	P_2	P_3	Оценка	Вероятность оценки	Примечания
1	11:20	+	++	+	Уменьшение производительности	0,91	задача 1
2	12:55	++	++	++	Изменения в диагностических данных	0,98	задача 2
3	14:20	++	++	+	Увеличение производительности	0,95	задача 3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
m					n		L

Пример экспертной модели демонстрирует наличие 3 задач, где оценивается состав уязвимости, время появление и вероятность определения. В зависимости от задачи присутствуют следующие показатели: уменьшение производительности, изменения в диагностических данных, увеличение производительности. Для создания полных и неполных информационных образов каждого узла уязвимости, применяется техника составление диагностических правил с использованием разностороннего определения сочетаний.

Актуальность создания новых методов и алгоритмов решения задач обнаружения угроз подтверждается растущим количеством данных в промышленности, требующей постоянной диагностики скрытых уязвимостей. Применение методов искусственного интеллекта в связке с аналитико – прогности-

ческой системой для реорганизации данных в знания – это перспектива многофункциональной системы обработки диагностической информации объекта исследования. Следовательно, необходимо осуществить классификацию структуры данных и применить конструкцию данных: пространственных (cross section data), панельных (panel data) или временных рядов (time series). Также при внедрении системы защиты и идентификации уязвимостей на предприятии стоит учесть риски внутренних обвалов в системах [4]. Для этого будет сконцентрировано внимание на круглосуточный мониторинге и анализе обнаруженных признаков.

Так как, диагностическая информация имеет изменчивый характер, весьма действенен метод регрессионного анализа, что приведет к определению новых зависимостей и состояний. Это уменьшит ошибки в последующей обработке модели алгоритмами машинного обучения. Чтобы увеличить производительность системы поиска требуется применить параллельно обучение.

Заключение

Рассмотрение вопроса защищенности данных на предприятии показало уязвимости, которые трудно обнаружить и без точечного анализа не обойтись. Это угрозы сформированные нейронной сеть и алгоритмами машинного обучения. Также применение методов искусственного интеллекта злоумышленниками позволяет им долгое время быть необнаруженными в структурах предприятия.

Для точечного анализа уязвимостей в системе определения инцидентов кибербезопасности требуется параллельное применение методов и алгоритмов искусственного интеллекта. Построенная модель ориентированных диагностических правил для составления экспертной оценки, будет применена в модели машинного обучения, как ветвь анализа идентифицируемых

признаков. Чтобы реализовать качественную метрику данных потребуется использовать знания узкопрофильного специалиста. Специалист должен иметь опыт устранения точечных инцидентов на предприятиях. Быстрое внедрение экспертной системы – будет возможно при наличии ее в составной части информационной системы. Также использование инновационной модели поиска сбалансирует производительность предприятия и уменьшит потерю данных.

Благодарность

Выражаем благодарность «Российскому технологическому университету – МИРЭА», за предоставленную возможность проводить исследования в области искусственных нейронных сетей, интеллектуального анализа данных и безопасности.

Список литературы:

1. Хасти Тревор и Тибширани Роберт. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование. [2-е изд.] – Springer. 2020. – 770 с.
2. Henrik Brink joseph Mark W. Rlchards Fetherolf. Real-World Machine Learning. MANNING SHEL TER ISLAND 2017. – 338p.
3. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. 2018. – 653 с.
4. Зыков С. В. Семантическая интеграция данных для безопасности и целостности корпоративных систем // Безопасность информационных технологий. 2009. – № 3. – С. 16–19.