ВОПРОСЫ КИБЕРБЕЗОПАСНОСТИ

№5 2023 (57)

DOI: 10.21681/2311-3456

Безопасный искусственный интеллект

Безопасность киберфизических систем

Безопасность веб-приложений

ВОПРОСЫ КИБЕРБЕЗОПАСНОСТИ

НАУЧНЫЙ РЕЦЕНЗИРУЕМЫЙ ЖУРНАЛ

№5(57) 2023 г.

Выходит 6 раз в год

Журнал выходит с 2013 г. (Свидетельство о регистрации ПИ № ФС77-75239). Перерегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 07.03.2019.

Журнал входит в перечень научных изданий, в которых должны быть опубликованы основные результаты исследований соискателей учёных степеней кандидата и/или доктора наук, а также в российский индекс научного цитирования RSCI на международной платформе научных публикаций Web of Science (WoS)

Учредитель и издатель

АО «Научно-производственное объединение «Эшелон»

Над номером работали:

Г.И. Макаренко – *шеф-редактор* И.Ю.Шахалов – *отв. секретарь*

И.М. Ануфриев – *дизайн*

Подписано к печати 10.08.2023 г.

Общий тираж 120 экз. Цена свободная

Адрес: 107023, Москва, ул. Электрозаводская, д. 24, стр. 1.

E-mail: editor@cyberrus.info, тел.: +7 (985) 939-75-01.

Требования, предъявляемые к рукописям, размещены на сайте: **https://cyberrus.info/**

СОДЕРЖАНИЕ

Подписка на журнал осуществляется в почтовых отделениях по каталогу «Пресса России». Подписной индекс 40707

ИНТЕРВЬЮ С ПРЕЗИДЕНТОМ НАЦИОНАЛЬНОЙ АССОЦИАЦИИ МЕЖДУНАРОДНОЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ВЛАДИСЛАВОМ ПЕТРОВИЧЕМ ШЕРСТЮКОМ[1]

Корреспондент: Уважаемый Владислав Петрович, 25 лет назад наша страна первая из мирового сообщества поставила на обсуждение ООН проблематику международной информационной безопасности. Вы принимали участие в разработке российской инициативы, расскажите с чего все начиналось.

Шерстюк В.П. Да, международной информационной безопасности, как новой сферы международных отношений, в 2023 году исполнится 25 лет!

23 сентября 1998 года Постоянный представитель Российской Федерации при Организации Объединенных Наций Сергей Лавров направил на имя Генерального секретаря ООН письмо Министра иностранных дел Российской Федерации И.С. Иванова, к которому был приложен проект резолюции «Достижения в сфере информатизации и телекоммуникаций в контексте международной безопасности». Этим документом предлагалось всем государствам-членам ООН активизировать рассмотрение на двустороннем и многостороннем уровнях существующих и потенциальных угроз в сфере информационной безопасности. Россией впервые была сформулирована триада угроз в информационной сфере. Тогда она вызвала резкое неприятие американской стороной, особенно ее военно-политический аспект.

Должно было пройти 15 лет для того, чтобы в совместном «Заявлении президентов США и Российской Федерации о новой области сотрудничества и укрепления доверия» от17 июня 2013 года появились такие слова:

«Мы (т.е. Россия и США) признаем, что угрозы в сфере использования ИКТ и самим ИКТ включают военно-политические и криминальные угрозы, а также угрозы террористического характера и относятся к ряду наиболее серьезных проблем национальной и международной безопасности».

Это было славное время надежд. В заявлении сообщалось:

- о создании механизма обмена информацией между Россией и США для обеспечения более эффективной защиты критически важных информационных систем;
- об установлении канала связи по вопросам урегулирования потенциально опасных ситуаций в информационной сфере между должностными лицами высокого уровня, в том числе возможности использования для этой цели существующей линии прямой связи между Центрами по уменьшению ядерной опасности;
- о создании в рамках российско-американской Президентской комиссии двухсторонней рабочей группы по вопросам угроз в сфере использования ИКТ.

Корреспондент: США пошли на такое сотрудничество, чтобы не потерять свое политическое лидерство или ими двигали исключительно прагматичные соображения защиты собственной критической инфраструктуры? Что дало заключение соглашения?

Шерстюк В.П. В результате развития и улучшения российско-американских отношений спустя два года Группе правительственных экспертов ООН удалось договориться о начале обсуждения добровольных и необязательных правилах ответственного поведения

1 Интервью состоялось 5 апреля 2023 года

государств в ИКТ-среде, включающих одиннадцать пунктов, при этом констатировалось, что для поддержания открытой, безопасной, стабильной, доступной и мирной ИКТ-среды нормы и принципы международного права должны быть дополнены нормами ответственного поведения государств в ИКТ-среде. Далее пошло все не так - обозначились два возможных подхода к реализации предложений Группы правительственных экспертов ООН (2015), которые условно могут быть названы американским и российским.

Американский подход базируется на том, что нормы носят необязательный характер, а применение норм является добровольным. Это принципиальное соображение. Отметим, что несмотря на отсутствие очевидных препятствий, до настоящего времени ни одно государство не приступило к реализации такого подхода. При экспертном анализе становится ясным, что различия в толковании и применении норм государствами могут явиться поводом для возникновения международных споров и дальнейшего обострения международной ситуации.

Подписанная 1 марта 2023 года президентом США Байденом Национальная стратегия кибербезопасности США, наряду с голословными обвинениями России и её союзников в кибератаках, предусматривает усиление «колонизации» глобального информационного пространства по киберправилам коллективного Запада. США хотят и пользуются своим технологическом превосходством в ИКТ-среде для достижения геополитических целей.

Корреспондент: какой подход продвигает Российская Федерация и какова в этой деятельности роль Национальной Ассоциации международной информационной безопасности?

Шерстюк В.П. Российский подход закреплен в Основах государственной политики Российской Федерации в области международной информационной безопасности, утвержденных Указом Президента Российской Федерации от 12 апреля 2021 г. № 213.

Этот подход подтверждён и в обновленной Концепции внешней политики Российской Федерации (Указ Президента РФ от 31 марта 2023 г. № 229). В иерархии приоритетов обеспечения международной информационной безопасности первое место отведено укреплению и совершенствованию международно-правового режима предотвращения и разрешения межгосударственных конфликтов и регулирования деятельности в глобальном информационном пространстве.

По мнению российских экспертов, нормы должны применяться «на основе общепризнанных принципов и норм международного права, на условиях равноправного партнерства и обеспечивать поддержание международного мира, безопасности и стабильности», т.е. иметь императивный (обязательный) характер и устанавливаться международным договором.

Определенное время, после их появления в 2015 году, тема применения норм и правил ответственного поведения государств в информационной сфере была центральной на официальных и неофициальных международных встречах, конференциях и форумах.

Однако, на фоне президентских выборов 2016 года в США и общего ухудшения международной обстановки, возможности для обсуждения этих подходов, да и вообще вопросов международного сотрудничества в области противодействия угрозам в сфере информационных и коммуникационных технологий на уровне межгосударственных консультаций к 2018 году оказались практически исчерпаны.

В этих условиях, с тем, чтобы убрать преграды исключительно политического характера для продолжения продвижения инициатив Российской Федерации в области международной информационной безопасности, выход виделся в привлечении к решению этих проблем соответствующей российской негосударственной структуры, в состав которой вошли бы авторитетные профильные эксперты и организации.

В апреле 2018 г. такая структура была создана - Национальная Ассоциация международной информационной безопасности. Учредителями Ассоциации выступили:

- Московский государственный университет имени М.В. Ломоносова;
- Московский государственный институт международных отношений (университет) МИД России;
- Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации;
- Дипломатическая академия МИД России;
- Редакция журнала «Международная жизнь» МИД России;
- АНО «Клуб безопасности информации в промышленности» «ГМК «Норильский никель».

Корреспондент: каковы цели и задачи вашей организации?

Шерстюк В.П. Основными целями Ассоциации, закрепленными в Уставе, были определены следующие:

DOI:10.21681/2311-3456-2023-5-2-8

— содействие продвижению российских инициатив в области обеспечения МИБ;
— содействие федеральным органам законодательной и исполнительной власти Российской Федерации в их деятельности по реализации государственной политики в области МИБ, а также содействие российским коммерческим и некоммерческим организациям, и гражданам, участвующим в соответствии с законодательством Российской Федерации в реализации государственной политики в указанной области;
— содействие объективному информированию и разъяснению организациям гражданского общества Российской Федерации и зарубежных государств основных положений государственной политики Российской Федерации в области МИБ;
— содействие формированию системы обеспечения устойчивого функционирования глобальной и национальной информационных инфраструктур, безопасного использования информационных и коммуникационных технологий во всех сферах жизни общества и управления государством.

Корреспондент: каким потенциалом для достижения этих целей обладает Ассоциация, в каких форматах осуществляется ее деятельность?

Шерстюк В.П. Ассоциация была создана не на пустом месте. Ее костяк составили эксперты, которые с момента появления проблематики международной информационной безопасности были вовлечены в эту интереснейшую сферу научной деятельности вначале на семинарах и конференциях Московского государственного университета имени М.В. Ломоносова и МГИМО.

Экспертный потенциал Ассоциации год от года растет, например, в 2022 году нашими членами опубликовано более сорока научных статей по ключевым проблемам обеспечения международной информационной безопасности. В марте этого года Ассоциация подготовила сборник материалов, содержащих обзор развития систем информационной безопасности национальных сегментов ИКТ-среды Лиги арабских государств.

Но следует признать, что форматы работы складывались постепенно. В далеком 2006 году Московский университет впервые инициировал Международный Форум «Партнерство государства, бизнеса и гражданского общества, при обеспечении международной информационной безопасности», который с тех пор проводился ежегодно, вначале в Гармиш-Партенкирхене (Германия), затем в Москве. Этот Форум явился одной из основных, если не основной, международной площадкой для обсуждения российскими и зарубежными экспертами наиболее актуальных проблем формирования системы международной информационной безопасности. За 17 лет через «Форум Гармиш-Партенкирхена» прошло несколько сотен международных экспертов. Обсуждены многие проблемы безопасности информационной сферы. Актуальность Форума подтверждается его включением в План реализации Основ государственной политики Российской Федерации в области международной информационной безопасности на 2021-2025 годы.

Следует отметить, что несмотря на не слишком благоприятные внешние факторы, Ассоциации удалось на должном уровне поддерживать интерес к Форуму. Так, если в 2018 году в его работе приняли участие 103 эксперта из 17 стран, то в 2022 их было 250 из 43 стран.

Организация и проведение ежегодных Форумов позволяет нам решить многие задачи, в том числе:
— продвижения в среде зарубежных экспертов инициатив Российской Федерации по принятию в качестве универсальных международных соглашений, в том числе концепции конвенции международной информационной безопасности и проекта Конвенции ООН по противодействию использования информационно-коммуникационных технологий в преступных целях;
— разъяснения позиции Российской Федерации по вопросу об обеспечении безопасности объектов критической информационной инфраструктуры и о необходимости расширения сотрудничества в области формирования системы международной информационной безопасности;
— формирования представления о НАМИБ как об авторитетной российской негосударственной организации, которая выражает позицию политического руководства Российской Федерации по вопросам международной информационной безопасности.

Корреспондент: Вы отметили высокий экспертный потенциал Ассоциации. Поделитесь с широкой аудиторией, какую тематику исследований вы считаете приоритетной, привлекаете ли вы к ней зарубежных специалистов?

Шерстюк В. П. На наш взгляд, обсуждение проблемных вопросов на Форумах должно способствовать

постепенному переносу в центр внимания международных экспертов практических вопросов формирования системы международной безопасности. Для этого Ассоциации следует уделить большее внимание переводу общей тематики докладов, представляемых на Форум, с обоснования необходимости создания системы международной информационной безопасности (здесь все слова уже сказаны) на обсуждение возможных подходов к решению этой задачи.

Именно эта тема в качестве научной была вынесена на рассмотрение Международного исследовательского консорциума информационной безопасности (МИКИБ). Несколько слов о Консорциуме. Он был создан по инициативе российской стороны в рамках Гармиш-процесса. Для того, чтобы не растерять экспертов в промежутке времени между ежегодными Форумами, мы предложили в это время заниматься исследованиями по совместным согласованным планами и периодически встречаться на сессиях. В Консорциум вступило 28 научных центров и организаций, изучающих проблемы международной информационной безопасности, из 18 стран. Мы провели ряд исследований, приобрели некоторый опыт совместной работы, состоялись сессии в Баку, Пекине, Астане, Москве и Софии.

В 2018 году Ассоциация предложила Консорциуму рассмотреть вопрос «Методологические вопросы применения норм, правил и принципов ответственного поведения государств, призванных способствовать обеспечению открытой, безопасной, стабильной, доступной и мирной ИКТ-среды». В совместном выполнении проекта приняли участие эксперты Московского государственного университета имени М. В. Ломоносова (Россия), Института Восток-Запад (США), Института киберполитики (Эстония, Финляндия) и Фонда «ИКТ для мира» (Швейцария).

В результате проведенных исследований были сформулированы методологические подходы к изучению проблем и сформулированы предложения по направлениям дальнейших исследований. По многим вопросам было достигнуто согласие, однако было подтверждено и существование принципиальных методологических разногласий между российскими и западными экспертами по подходам к применению норм ответственного поведения государств в ИКТ-среде. Эти расхождения были вынесены в отдельный раздел отчета, который нами опубликован в виде брошюры, а его электронная версия представлена на сайте Ассоциации.

Основные результаты исследований были доложены на Форуме в Москве в 2020 году, вошли в доклады членов Ассоциации на международных конференциях в г. Гватемале (Республика Гватемала), г. Гаване (Республика Куба), г. Пекине (КНР), 12-м Конвенте РАМИ в г. Москве, использованы на переговорах в Центре Гуманитарного диалога в Женеве и МИД Швейцарии (г. Берн), а также на семинарах с участием российских и зарубежных экспертов в г. Москве и г. Санкт-Петербурге. С сожалением следует отметить, что сближения западной и российской позиции достичь не удалось. Проблематика международной информационной безопасности с годами стала все более политизированной, что в конечном итоге привело к прекращению формата Группы правительственных экспертов ООН по международной информационной безопасности.

Корреспондент: влияет ли на деятельность Ассоциации переход к многополярному мироустройству и как себя проявляют ваши западные партнеры? Расскажите о том, какие задачи вы ставите перед Ассоциацией в новых исторических условиях для достижения более справедливых условий использования глобального информационного пространства развивающимися странами?

Шерстюк В. П. Как я отмечал выше, изучение проблемы практического применения норм ответственного поведения государств в ИКТ-среде для Ассоциации является приоритетной.

Конечная цель заключается в подготовке научно обоснованных рекомендаций по вопросам международного сотрудничества в области безопасного использования информационно-коммуникационных технологий (ИКТ) в формирующемся глобальном информационном обществе. В этой связи Ассоциацией была поставлена научно-исследовательская работа, к выполнению которой удалось привлечь лучших российских юристов- международников, а также авторитетных экспертов технического профиля.

Научно-исследовательская работа, которая в некотором смысле продолжила начатый МИКИБом анализ, была направлена на «содействие выработке, с учетом специфики информационно-коммуникационных технологий, новых принципов и норм международного права, регулирующих деятельность государств в глобальном информационном пространстве». НИР была нацелена на определение условий, при которых в сфере использования информационно-коммуникационных технологий может быть «установлен международно-правовой режим обеспечения безопасности».

Фактическую основу исследования составили «нормы, правила и принципы ответственного поведения государств в ИКТ-среде», применение которых может «снизить риск нарушения международного мира, безопасности и стабильности», сформулированные с учетом предложений Китая, России, Таджикистана и Узбекистана и представленные в докладах Групп правительственных экспертов ООН (2015, 2021 гг.).

В рамках выполнения НИР показано, что для практического применения норм ответственного поведения государств в ИКТ-среде прежде необходимо решение целого ряда принципиальных проблем. К ним относятся:

- делимитация и демаркация зон ответственного поведения государств в ИКТ-среде;
- имплементация норм в национальное законодательство;
- развитие механизмов мирного разрешения споров, возникающих в связи с инцидентами в ИКТ-среде.

По результатам выполнения НИР в 2023 году выпущена коллективная монография «Международная безопасность в среде информационно-коммуникационных технологий».

В ней представлены:

- предложения в проект позиции Российской Федерации по вопросам применения норм ответственного поведения государств в ИКТ-среде;
- проект универсального международного соглашения о применении норм ответственного поведения государств в ИКТ-среде;
- проект международного стандарта технического регулирования в области менеджмента международной информационной безопасности.

Считаем необходимым и целесообразным организовать и провести широкое экспертное обсуждение полученных результатов.

Корреспондент: расскажите о планах Ассоциации на ближайшие годы. Изменится ли вектор ее международного сотрудничества?

Шерстюк В. П. Ассоциация предполагает активно развивать сотрудничество по вопросам изучения проблем практического применения норм ответственного поведения государств в ИКТ-среде с негосударственными организациями конструктивно настроенных государств.

Вектор внешней политики России, как вам известно, теперь направлен на Восток. С этой целью соответствующие предложения об установлении более тесного взаимодействия были направлены Государственному секретарю Союзного государства, Генеральному секретарю Шанхайской Организации Сотрудничества, Генеральному Секретарю Организации Договора о коллективной безопасности и Главе представительства Лиги арабских государств.

В рамках деятельности по налаживанию сотрудничества с зарубежными негосударственными организациями для продвижения российских подходов к формированию системы международной информационный безопасности Ассоциация установила устойчивое взаимодействие с неправительственными организациями Китайской Народной Республики.

В рамках этого взаимодействия Ассоциация приняла участие в работе Конференции по информационной безопасности в Университете Циньхуа (г.Пекин, КНР, 2018), Шестой всекитайской конференции по безопасности Интернета (г.Пекин, КНР, 2018), Пекинской конференции по кибербезопасности и Конференции по Интернет-безопасности (г.Пекин, КНР, 2019), Конференции по Безопасности Данных в рамках первой сессии Форума по Глобальному сотрудничеству в области общественной безопасности (г.Ляньюньган, КНР, 2022).

В рамках XVI Международного форума в прошлом году состоялись плодотворные консультации представителей Ассоциации с делегацией Республики Иран во главе с Заместителем Министра информации и цифрового развития. Договорились об установлении прочных отношений. Ассоциация получила приглашение посетить Иран для развития контактов.

Активно укрепляются отношения с нашими ближайшими соседями. В развитие российско-белорусских межведомственных консультаций в области обеспечения информационной безопасности, состоявшихся 13 апреля 2022 г. в Минске (российскую делегацию возглавлял О.В. Храмов), Ассоциация приняла участие в конференции Союзного государства по данной проблеме в Минске. Продолжением данного направления сотрудничества стало участие делегации Белорусского института стратегических исследований во главе с его директором О. С. Макаровым в работе XVI Международного форума. Рассчитываем, что наши партнерские связи с Белоруссией будут активно развиваться в рамках реализации Концепции информационной безопасности Союзного государства, утвержденной в феврале 2023 г.

Также мы неустанно ищем новых партнеров и новые площадки для сотрудничества, например, уста-

новлено взаимодействие с Центром гуманитарного диалога (Швейцария). В 2018 году и 2020 году в Москве и Женеве были проведены экспертные консультации с представителями Центра по возможным направлениям сотрудничества в области международной информационной безопасности. По этой же теме была проведена встреча представителей Ассоциации с представителями МИД Швейцарии (г. Берн, 2020). Во взаимодействии с Центром гуманитарного диалога проведен семинар экспертов США, России и КНР (2021) по проблемам применения норм ответственного поведения государств.

Важность развития международного сотрудничества была отмечена в сообщениях представителей Ассоциации в ходе участия в работе межведомственных делегаций под руководством аппарата Совета Безопасности Российской Федерации и МИД России на Конференции «Кибер- и Информационная безопасность» с участием представителей Центральноамериканского парламента (г. Гватемала, Гватемала, 2019), на Сингапурской кибернеделе и российско-сингапурских межведомственных консультациях по международной информационной безопасности (Сингапур, 2019), на Международной конференции по международной информационной безопасности (г. Гавана, Республика Куба, 2019).

Следует отметить, что в новых политических условиях Ассоциация продолжает свою многолетнюю деятельность и на ведущих международных площадках, где нашими экспертами продвигаются российские подходы и результаты наших исследований. Доклады по тематике формирования системы международной информационной безопасности на основе применения норм ответственного поведения государств в ИКТ-среде Ассоциацией были представлены:

- на конференции Парламентской ассамблеи ОБСЕ по МИБ (г.Лиссабон, 2018);
- в ходе работы Круглого стола по вопросам военной киберстабильности (г.Париж, Франция, 2018);
- на встрече с экспертами Комитета по международной безопасности и контролю над вооружениями Национальной академии наук США (2018);
- в ходе консультаций по проблеме применения международного права в ИКТ-среде с представителями Международного Комитета Красного Креста (Москва, 2022).

Проведенная на данном направлении работа, на наш взгляд, позволила Ассоциации занять достойное место в системе российских негосударственных организаций, содействующих повышению эффективности реализации государственной политики в области обеспечения международной информационной безопасности. Свидетельством этому, в частности, является получение Ассоциацией при поддержке аппарата Совета Безопасности Российской Федерации и МИД России в 2022 году аккредитации при Рабочей группе ООН открытого состава (РГОС ООН) в качестве негосударственной организации.

Корреспондент: расскажите поподробнее об этом новом формате обсуждения проблематики международной информационной безопасности на площадке ООН, который, в отличие от ГПЭ, стал доступен всем без исключения заинтересованным странам.

Шерстюк В.П. РГОС ООН является новым форматом сотрудничества, созданным по инициативе России, что говорит о сохраняющемся лидерстве нашей страны в проблематике международной информационной безопасности.

Предполагаем использовать открывающиеся в связи с этим возможности для продвижения предложений, способствующих приданию добровольным нормам ответственного поведения государств в ИКТ-среде обязательного характера. В качестве первого шага в августе 2022 года при содействии МИД России Председателю РГОС ООН был представлен краткий результат первого этапа НИР, о которой я говорил выше. После перевода полагаем направить и коллективную монографию по итогам исследований.

В планах также предложить на РГОС ООН обсуждение актуальной тематики, порожденной взрывным развитием технологий искусственного интеллекта и распространением сгенерированного с его помощью контента. Предварительная проработка вопроса с Председателем группы господином Гафуром уже состоялась.

У Ассоциации есть наработки по этим вопросам. Так, в рамках развития взаимодействия с негосударственными организациями зарубежных стран Ассоциация приняла активное участие в обсуждении проблем противодействия угрозам использования ИКТ-среды для распространения ложной информации, идеологии нацизма и терроризма в ходе работы Международного форума «Свободная журналистика в контексте прав человека, новых технологий и международная информационная безопасность» (г.Братислава, Словакия, 2018; г.Прага, Чехия, 2019; г.Н.Новгород, 2020;

DOI:10.21681/2311-3456-2023-5-2-8

г.Суздаль, 2021; г.Ярославль, 2022).

Представители Ассоциации также приняли участие в работе конференции Института ООН по проблемам разоружения, посвященной обсуждению темы: «Диалог по инновациям: дипфейки, доверие и международная безопасность» (2021), в российско-норвежской конференции «Сотрудничество на Севере в контексте международной безопасности» (г.Киркенес, Норвегия, 2019), а также в международной конференции по кибербезопасности в г.Тель-Авиве (Израиль, 2022).

Корреспондент: Вы перечислили так много направлений по которым работает Ассоциация. Как вам удается осуществлять такой объем работы?

Шерстюк В.П. Ассоциация постоянно прирастает новыми членами. Мы предпринимаем усилия по расширению круга российских экспертов, участвующих в изучении и обсуждении проблем обеспечения международной информационной безопасности.

С этой целью мы подписали Договор о взаимодействии с Уральской государственной юридической академией имени В.Ф.Яковлева.

Кроме того, мы все активнее работаем с российской молодежью, участвуем в разработке учебных программ и пособий. Полагаем перспективным расширение контактов в университетской среде. Примером тому может служить продвинутая нами через японский университет Токай и опубликованная в марте этого года в журнале Human Security объемная статья “Актуальные проблемы международной информационной безопасности”.

Можно отметить, что Ассоциации удалось создать определенный организационный и научный потенциал для подготовки научно обоснованных предложений по переговорным позициям России в области формирования системы международной информационной безопасности. Созданный потенциал может быть использован для продолжения исследований проблем применения международного права к отношениям в области использования ИКТ-среды, обеспечения устойчивого функционирования глобальной и национальной информационных инфраструктур, безопасного использования информационных и коммуникационных технологий во всех сферах жизни общества и управления государством.

АНАЛИЗ УГРОЗ ЗЛОУМЫШЛЕННОЙ МОДИФИКАЦИИ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ СИСТЕМ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

Костогрызов А.И.[1], Нистратов А.А.[2]

Цель: *предложить методический аппарат для вероятностного анализа корректности обучаемых программных средств (ПС) в системах с искусственным интеллектом (СИИ) при их разработке и эксплуатации в условиях потенциальных угроз злоумышленной модификации модели машинного обучения (ММО).*

Методы исследования включают: *методы теории вероятностей, методы системного анализа. Подход основан на адаптации разработанных ранее авторских вероятностных моделей, которые доведены до уровня реализации в ГОСТ Р 59341-2021 «Системная инженерия. Защита информации в процессе управления информацией системы».*

Результат: *в условиях принятых предположений и допущений разработаны вероятностные модели для оценки частных рисков невыявления некорректностей в машинном обучении (дообучении) при разработке и эксплуатации ПС, а также метод оценки интегрального риска нарушения корректности машинного обучения в течение задаваемого периода прогноза. Проанализированы актуальные угрозы подмены ММО и модификации ММО путем искажения («отравления») обучающих данных. Разработаны предложения по формированию исходных данных для прогнозирования рисков с использованием предложенных моделей. Подход проиллюстрирован расчетными примерами с количественными оценками, зависимостями рисков от исходных данных и обоснованием рекомендаций.*

Научная новизна: *впервые для условий потенциальных угроз злоумышленной модификации ММО предложены модели и методы количественной оценки частных рисков невыявления некорректностей в машинном обучении при разработке и эксплуатации ПС и интегрального риска нарушения корректности машинного обучения для СИИ в течение задаваемого периода прогноза.*

Ключевые слова: *вероятность, искажение обучающих данных, модель, риск, система, угрозы.*

DOI:10.21681/2311-3456-2023-5-9-24

1. Введение

Системы с искусственным интеллектом (СИИ) все глубже проникают в повседневную жизнь человека. И это далеко не только голосовые помощники в наших персональных телефонах, навигаторы, онлайн карты и иные удобные сервисы. СИИ все чаще используется в системах обеспечения безопасности на основе интеллектуальной обработки огромных потоков разнородной информации, поступающей от различных камер, сенсоров, устройств телеметрии. Программные средства (ПС) СИИ, обновляемые с помощью моделей машинного обучения, помогают соответствующим службам в распознавании лиц и документов, строений и сооружений и их местоположений, в идентификации предпосылок к нарушению информационной, промышленной, транспортной, экологической безопасности, в геологоразведке, медицине, фармацевтике и биологии, в мониторинге соблюдения правил дорожного движения, распознавая условия нарушения и государственные номера транспортных средств нарушителей и др. Эти примеры далеко не исчерпывают практических возможностей СИИ – см., например, [1-3].

В основе эффектов от применения СИИ лежат обучаемые нейронные сети. Искусственные нейронные сети основаны на наборе персептронов, называемых нейронами. Каждый нейрон сопоставляет набор входных данных с выходными, используя функцию активации. Машинное обучение управляет весами

1 Костогрызов Андрей Иванович, заслуженный деятель науки РФ, доктор технических наук, профессор, главный научный сотрудник, Федеральный исследовательский центр «Информатика и управление» Российской академии наук. Москва, Россия. E-mail: Akostogr@gmail.com

2 Нистратов Андрей Андреевич, кандидат технических наук, старший научный сотрудник, Федеральный исследовательский центр «Информатика и управление» Российской академии наук. Москва, Россия. E mail: andrey.nistratov.job@yandex.ru

и функцией активации таким образом, чтобы иметь возможность правильно определять выходные данные. В то время, как однослойная нейронная сеть (или персептрон) - это подход к разработке объектов, глубокая нейронная сеть позволяет изучать объекты, используя необработанные данные в качестве входных данных. За счет этого достигается существенное увеличение производительности СИИ по сравнению с обычным человеческим интеллектом при решении многих практических задач. При этом обеспечение безопасности информации СИИ должно предусматривать возможность противодействия злоумышленным угрозам подмены и модификации ММО. Однако сегодня системная зависимость нарушения нормального функционирования СИИ от этих угроз является не только далеко не прозрачной, но и на количественном уровне не анализируется. Не представляя всей «внутренней кухни» машинного обучения, заказчик и пользователи системы могут вполне воспринимать нарушения ее нормального функционирования за обычное техническое несовершенство, не устанавливая прямой связи со злоумышленными действиями «умного» нарушителя по модификации ММО. Опасность в том, что нарушитель пытается целенаправленно подменить ММО или исказить обучающие данные, вводя тщательно разработанные ложные образцы так, чтобы в конечном итоге скомпрометировать весь процесс машинного обучения.

В рамках настоящей работы для систем, использующих СИИ, из множества различных угроз выделены следующие актуальные угрозы[3]: угроза подмены ММО (УБИ.222) и угроза модификации ММО путем искажения («отравления») обучающих данных (УБИ.221). Это обусловлено следующими соображениями. В наше время нередко разработчики ПС, осуществляющие машинное обучение (дообучение), принадлежат сторонним организациям относительно разработчика систем, использующих СИИ. Они являются основными владельцами ММО, не хотят раскрывать и передавать заказчику и головному разработчику системы исходные тексты, находятся на субконтракте, сами разрабатывают ПС, в которых содержатся результаты машинного обучения, и контролируют его корректность. Обученные и дообученные ПС передаются заказчику и головному разработчику систем, использующих СИИ, для функционального тестирования, после чего оттестированные ПС принимаются в эксплуатацию в системе. Сертификация дообучаемых ПС по требованиям безопасности может оказаться нецелесообразной из-за длительности и дороговизны ее проведения для заказчика, а также из-за возможного нежелания владельцев ММО раскрывать все исходные тексты программ и методы обучения. В этом случае угрозы, связанные со злоумышленной модификацией ММО, становятся остро актуальными и требуют системного анализа.

Применение предлагаемого подхода к решению различных прямых и обратных задач для обеспечения эффективного целевого применения СИИ позволит прогнозировать риски и количественно обосновывать принимаемые решения о стратегии и мерах противодействия рассматриваемым угрозам. При этом под риском понимается сочетание вероятности нанесения ущерба и тяжести этого ущерба (по ГОСТ Р 51898). В работе основное внимание сосредоточено на анализе вероятностного выражения риска, полагается, что возможный ущерб (чаще - репутационный) противопоставляется расчетным значениям рисков и соответствующим условиям моделирования.

Подход учитывает последние взгляды Национального института стандартизации США на таксономию внедрения в СИИ вредоносного машинного обучения, а также основы управления рисками для СИИ[4],[5] и не противоречит им.

2. Характеристика возможных угроз и сценариев их реализации

Краткая характеристика угроз УБИ.222 и УБИ.221, а также возможных злоумышленных действий нарушителей, именуемых атаками (Attacks), приведена со ссылками на обобщенные взгляды в России и международном сообществе, анализирующем риски в СИИ[6] – см. также [1-4] и сноски 3–5.

Угроза УБИ.222 заключается в возможности подмены ММО внутренним нарушителем (с высоким потенциалом). Угроза обусловлена слабостями разграничения доступа в СИИ, реализация угрозы возможна при наличии у нарушителя непосредственного доступа к ММО.

Угроза УБИ.221 заключается в возможности модификации ММО внешним нарушителем (с высоким по-

3 см. сайт ФСТЭК России https://bdu.fstec.ru/ - Банк данных угроз безопасности информации. ФАУ «ГНИИИ ПТЗИ ФСТЭК России». Дата обращения 25.07.2023

4 Adversarial Machine Learning. A Taxonomy and Terminology of Attaks and Mitigations (Вредоносное машинное обучение. Таксономия и терминология атак, и способов снижения их отрицательных последствий). NIST AI 100-2e2023 ipd, 2023. nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2023.ipd.pdf

5 Artificial Intelligence Risk Management Framework. NIST AI 100-1, 2023. nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

6 Biggio B., Fumera G. and Roli F. Security evaluation of pattern classifiers under attack. IEEE transactions on knowledge and data engineering 26, 4. 2014. 984-996.

тенциалом) или внутренним нарушителем (со средним или высоким потенциалом) путем искажения («отравления») обучающих данных. Угроза обусловлена недостатками алгоритмов машинного обучения и осуществления процесса машинного обучения. Реализация угрозы возможна при наличии у нарушителя возможности воздействовать на процесс машинного обучения. Атаки с искажением («отравлением») по сути представляют собой целенаправленное злоумышленное изменение обучающих данных во время машинного обучения для компрометации всего процесса машинного обучения («отравление» - это буквальный перевод на русский язык англоязычного термина Poisoning Attack).

Анализ возможностей нарушителя на этапе обучения состоят в следующем. Нарушитель пытается напрямую повлиять на ММО или повредить ее, изменяя набор данных, используемый для обучения. Самая распространенная атака - это простой доступ к частичным или полным данным обучения.

На сегодня выделяются три применимые стратегии атаки для модификации ММО, основанные на возможностях нарушителя – это стратегии ввода данных, модификации данных и искажения логики ММО.

Стратегия ввода данных используется, когда нарушитель не имеет никакого доступа к обучающим данным, а также к алгоритму обучения, но имеет возможность добавить новые данные в обучающий набор. Он может исказить целевую ММО, вставив ложные выборки в обучающий набор данных. Это влечет за собой некорректность машинного обучения при разработке соответствующих ПС.

Стратегия модификации данных используется, когда нарушитель не имеет доступа к алгоритму обучения, но имеет полный доступ к обучающим данным. Нарушитель напрямую искажает обучающие данные (например, путем прямого изменения меток обучающих данных), изменяя их до того, как они будут использованы целевой ММО. Это также влечет за собой некорректность машинного обучения при разработке соответствующих ПС.

Стратегия искажения логики ММО используется, когда нарушитель имеет возможность вмешиваться в алгоритм обучения (например, путем манипулирования входными характеристиками в зависимости от своих возможностей). Это наиболее опасные атаки, поскольку очень трудно разработать стратегию упреждающего противодействия злоумышленным действиям нарушителя, способного законным образом изменить логику обучения (подобного рода нарушения легко могут быть замаскированы под неумышленную «ошибку»). Искажение логики целенаправленно влечет за собой некорректность машинного обучения при разработке соответствующих ПС, поскольку нарушителем контролируется и модифицируется сама целевая ММО.

Злоумышленные возможности нарушителя на этапе тестирования ПС состоят в следующем. Нарушитель пытается напрямую повлиять на ММО или повредить ее, изменяя набор данных. Атаки во время тестирования не влияют на целевую ММО, но приводят к неверным выходным результатам при использовании соответствующих ПС. Эффективность таких атак определяется главным образом объемом доступной нарушителю информации о целевой ММО.

Таким образом, реализация угроз злоумышленных действий по модификации ММО для СИИ рассчитана на «умного» нарушителя, понимающего свои возможности, представляющего и способного поставить достижимые задачи нарушения целостности системы. Вышеизложенные пояснения даны для понимания излагаемой далее формализации и системного анализа рассматриваемых угроз, мер противодействия этим угрозам и соответствующих рисков от реализации этих угроз.

3. Принятые предположения и допущения

Предполагается, что анализ рассматриваемых угроз может быть формализован с использованием понятия моделируемой системы. Получаемые результаты вероятностного моделирования используются в приложении к исходной СИИ, в интересах которой проводятся соответствующие исследования.

Под моделируемой системой понимается система, для которой решение задач системного анализа осуществляется с использованием ее формализованной модели и, при необходимости, формализованных моделей учитываемых сущностей в условиях их применения. В свою очередь под целостностью моделируемой системы понимается такое ее состояние, которое отвечает целевому назначению модели системы в течение задаваемого периода прогноза.

Примечание. В качестве модели системы могут выступать формализованные сущности, объединенные целевым назначением. Например, при проведении системного анализа в принимаемых допущениях, ограничениях и предположениях модель может формально описывать процесс, функциональные действия, множество активов или множество этих или иных сущностей в их целенаправленном применении в задаваемых условиях (по ГОСТ Р 59341).

С учетом неопределенностей расчет вероятностных показателей делается при условии или в предпо-

ложении реальной или гипотетической повторяемости возможных событий и их независимости. Для математической формализации приняты следующие предположения:

- в общем случае защищенность моделируемой системы от рассматриваемых угроз зависит от корректности машинного обучения при разработке и эксплуатации ПС;
- к началу периода прогноза целостность моделируемой системы полагается обеспеченной, в условиях неопределенностей возникновение и разрастание различных угроз целостности моделируемой системы описывается в терминах случайных событий;
- для различных вариантов развития угроз существуют технологии и меры для выявления признаков возникновения источников угроз и воспрепятствования реализации угрозам, а также следов реализации угроз.

Кроме того, делается предположение о наличии возможностей по определению предпосылок к реализации угроз, а также возможностей по приемлемому восстановлению нарушаемых условий функционирования моделируемой системы (с точки зрения противодействия угрозам). Обоснованное использование выбранных мер противодействия угрозам является предупреждающими контрмерами.

С учетом различных неопределенностей относительно возможных угроз принято допущение о пуассоновских потоках моментов возникновения событий на временной оси и об экспоненциальном распределении времени развития угроз. Предположение о пуассоновости обосновано тем, что в период прогноза общий поток моментов возникновения событий гипотетически представляет собой сумму большого числа составных разнородных потоков. Интенсивность каждого из слагаемых потоков мала по сравнению с интенсивностью суммарного потока – в такой ситуации действует предельная теорема Хинчина – Григолиониса, согласно которой суммарный поток будет близок к пуассоновскому. В свою очередь, экспоненциальное распределение обладает свойством отсутствия последействия. Это означает, что согласно предположению об экспоненциальности остаток времени до реализации угрозы всегда имеет то же распределение с тем же параметром, что и время с момента возникновения угрозы. Это предполагает более тяжелые условия функционирования моделируемой системы.

Принятые предположения и допущения позволяют предложить следующие вероятностные модели для анализа рассматриваемых угроз с использованием показателей рисков.

4. Модель для оценки риска невыявления некорректностей в машинном обучении при разработке ПС

Модель позволяет оценить возможность реализации рассматриваемых угроз при разработке ПС (в частности – при его тестировании) по показателям вероятности получения корректных результатов машинного обучения $P_{корр(1)}$ и риска невыявления некорректностей в машинном обучении. Модель адаптирует разработанные ранее авторские вероятностные подходы, которые доведены до уровня реализации в ГОСТ Р 59341, приложении В.3.7, а также учитывает иные научно-практические взгляды [5-14].

Определение: считается, что при разработке ПС машинное обучение (дообучение) проведено корректно в моделируемой системе, если в процессе контроля обученных ПС до истечения заданного срока его контроля все некорректности выявлены и новые алгоритмические ошибки не допущены. Некорректности при разработке ПС (в параметрах, исходных текстах программ, алгоритмах, обучающих фотографиях, метках и опорных векторах, действиях и др., способных привести к нарушениям нормального функционирования ПС при эксплуатации СИИ) – это в общем случае то, что искажает ожидаемые результаты последующего применения ПС после их машинного обучения в условиях рассматриваемых угроз по сравнению со случаем отсутствия каких-либо угроз. Некорректности появляются в результате реализации угроз, описанных выше в разделе 2, и характеризуют отсутствие корректности машинного обучения в моделируемой системе. Требуемая корректность машинного обучения при разработке ПС в идеале заключается в недопущении злоумышленной модификации адекватной ММО и использования небезопасных версий ПС, а также в исключении искажения обучающих данных. В общем случае под корректностью машинного обучения при разработке ПС для СИИ понимается свойство ПС, получаемых в результате машинного обучения, обеспечивать получение правильных согласованных результатов или эффектов обработки информации в соответствии с целевым назначением этой обработки в моделируемой системе. Корректность обеспечивается на основе применения адекватных способов машинного обучения и контроля результатов обучения, позволяющих выявить все имеющие место некорректности и не допустить алгоритмических ошибок при контроле

обученных ПС. Корректность машинного обучения после контроля информации по обучаемым ПС является следствием приемлемого соотношения между объемом контролируемой информации, частью важной для принятия решения информации, подлежащей учету, скоростью контроля информации, частотой ошибок контролера, длительностью его непрерывной работы и ограничениями на допустимое время контроля. В качестве контролера могут выступать человек – разработчик ПС, учитель, тестировщик или аналитик (в т.ч. лицо, принимающее решение), программно-технические инструментальные средства, ориентированные на выявление некорректностей в машинном обучении при разработке ПС, или их комбинация.

Для моделирования процесса контроля информации в моделируемой системе при разработке ПС приняты следующие обозначения:

V – объем информации по обучаемым (при тестировании – по обученным) ПС, подлежащий контролю (объем измеряется в безразмерных условных единицах – у.е., это могут, например, быть количество параметров, строк текста, алгоритмов, обучающих фотографий, меток и опорных векторов, действий, количество нарушений нормального функционирования ПС при тестировании и др.);

μ – часть важной для принятия решения информации, которая должна быть объективно использована при контроле информации в заданном объеме , измеряемая от 0 до 100% от анализируемого объема информации;

ν – скорость контроля (у.е. в единицу времени);

n – частота ошибок контроля 1-го рода (когда несущественная для принятия решения информация ошибочно воспринимается в качестве важной, влияющей на корректность машинного обучения);

$T_{нар}$ – среднее время наработки на алгоритмическую ошибку (когда объективно важная для принятия решения информация игнорируется, это – аналог ошибки контроля 2-го рода);

$T_{непр}$ – период непрерывной работы контролера;

$T_{зад}$ – задаваемое время на контроль информации.

Возможны 4 варианта соотношений между временем реального контроля всего контролируемого объема , задаваемым допустимым временем на контроль и непрерывным временем работы контролера .

Вариант 1. Задаваемое время на контроль информации не меньше, чем время реального контроля (т. е. $T_{реальн} \le T_{зад}$), а объем контролируемой информации относительно мал, что позволяет проверить его за один период непрерывной работы контролера ($T_{реальн} \le T_{непр}$).

Для экспоненциальной аппроксимации распределений интервалов между ошибками в контролируемой информации, времени до свершения ошибки 1-го рода и времени наработки контролера на ошибку, а также при условии независимости исходных характеристик вероятность $P_{после\ (1)}$ (V, μ, ν, n, $T_{нар}$, $T_{непр}$, $T_{зад}$) отсутствия некорректностей в машинном обучении после контроля для варианта 1 определяется выражением:

$$P_{после\ (1)} = \begin{cases} e^{-nV/\nu}\left[T_{нар}^{-1}\, e^{-\mu V} - \mu\nu e^{-V/(\nu\, T_{нар})}\right]/ \\ /\left(T_{нар}^{-1} - \mu\nu\right), \text{если } T_{нар}^{-1} \neq \mu\nu, \\ e^{-(n+\mu\nu)V/\nu}[1 - V\mu], \text{если } T_{нар}^{-1} = \mu\nu. \end{cases} \quad (1)$$

Вариант 2. Задаваемое время на контроль информации не меньше, чем время реального контроля (т. е. $T_{реальн} \le T_{зад}$), но объем контролируемой информации относительно большой ($T_{реальн} \le T_{непр}$). Это требует нескольких (N) периодов непрерывной работы контролера, в общем случае $N=V/(\nu\, T_{непр})$. Внутри каждого периода проверяют часть всего объема, равную в среднем $V_{части(2)}=V/N$, а допустимое время контроля информации для этой части принимается равным $T_{задчасти\ (2)} = T_{зад}/N$. Тем самым для каждой контролируемой части выполняются условия варианта 1. Вероятность $P_{после\ (2)} = P_{после(2)}$ (V, μ, ν, n, $T_{нар}$, $T_{непр}$, $T_{зад}$) отсутствия некорректностей в машинном обучении после контроля для варианта 2 определяется выражением:

$$P_{после\ (2)} = \\ = \{P_{после\ (1)}(V_{части\ (2)}, \mu, \nu, n, T_{нар}, T_{непр}, \\ T_{зад\ части\ (2)})\}^N. \quad (2)$$

Вариант 3. Задаваемое время на контроль информации меньше, чем время реального контроля ($T_{реальн} > T_{зад}$) при задаваемой средней скорости контроля ν, т. е. объективно может быть проконтролирована лишь часть от всего объема информации при контроле, эта часть равна $V_{части\ (3)} = \nu\, T_{зад}$. В свою очередь, сам объем контролируемой информации относительно мал и может быть проверен за один период непрерывной работы контролера, т. е. $T_{реальн} \le T_{непр}$ и для проверяемого объема $V_{части(3)}$ выполняются условия варианта 1. Вероятность $P_{после\ (3)} = P_{после(3)}$ (V, μ, ν, n, $T_{нар}$, $T_{непр}$, $T_{зад}$) отсутствия некорректностей в машинном обучении после его контроля для варианта 3 определяется выражением:

$$P_{после\ (3)} = [V_{части\ (3)}/V] \\ \cdot P_{после\ (1)}(V_{части\ (3)}, \mu, \nu, n, T_{нар}, T_{непр}, T_{зад}) + \\ +[(V - V_{части\ (3)})/V] \cdot P_{без\ контроля}, \quad (3)$$

где вероятность отсутствия некорректностей в непроверенной части информации, равной $V–V_{\text{части (3)}}$, составляет $P_{\text{без контроля}} = e^{-\mu(V-V_{\text{части (3)}})}$, а вероятность отсутствия некорректностей в объеме проверенной информации равна $P_{\text{после (1)}}$. ($V_{\text{части (3)}}$, μ, ν, n, $T_{\text{нар}}$, $T_{\text{непр}}$, $T_{\text{зад}}$).

Вариант 4. Задаваемое время на контроль информации меньше, чем время реального контроля ($T_{\text{реальн}} > T_{\text{зад}}$), а объем контролируемой информации относительно большой ($T_{\text{реальн}} > T_{\text{зад}}$). Аналогично варианту 3 реально может быть проконтролирована лишь часть от всего объема, равная $V_{\text{части (4)}} = \nu T_{\text{зад}}$. Относительно этой части возможны два подварианта:

- подвариант 4.1: $T_{\text{зад}} \le T_{\text{непр}}$, т. е. проверка будет завершена за один период непрерывной работы контролера;
- подвариант 4.2: $T_{\text{зад}} > T_{\text{непр}}$, т. е. потребуется несколько (N) периодов непрерывной работы контролера, $N = V_{\text{части (4)}} / (\nu T_{\text{непр}})$.

Для подварианта 4.1 вероятность отсутствия некорректностей в машинном обучении после контроля $P_{\text{после(4.1)}} = P_{\text{после(4.1)}}$ (V, μ, ν, n, $T_{\text{нар}}$, $T_{\text{непр}}$, $T_{\text{зад}}$) определяется выражением:

$$P_{\text{после (4.1)}} = [V_{\text{части (4)}}/V] \cdot \cdot P_{\text{после (1)}}\left(V_{\text{части (4)}}, \mu, \nu, n, T_{\text{нар}}, T_{\text{непр}}, T_{\text{зад}}\right) + +[V - V_{\text{части (4)}})/V] \cdot e^{-\mu\left(V-V_{\text{части (4)}}\right)}. \quad (4)$$

Для подварианта 4.2 внутри каждого периода проверяют новую часть, равную в среднем $V_{\text{части (4.2)}} = V_{\text{части (4)}}/N$, и допустимое время контроля для этой новой части принимают равным $T_{\text{зад части (4.2)}} = T_{\text{зад}}/N$.

Вероятность $P_{\text{после(4.2)}} = P_{\text{после(4.2)}}$ (V, μ, ν, n, $T_{\text{нар}}$, $T_{\text{непр}}$, $T_{\text{зад}}$) отсутствия некорректностей в машинном обучении после его контроля определяется выражением:

$$P_{\text{после (4.2)}} = [V_{\text{части (4)}}/V] \cdot \cdot \left\{P_{\text{после (1)}}\left(V_{\text{части (4.2)}}, \mu, \nu, n, T_{\text{нар}}, T_{\text{непр}}, T_{\text{зад части (4.2)}}\right)\right\}^N + +[V - V_{\text{части (4)}})/V] \cdot e^{-\mu\left(V-V_{\text{части (4)}}\right)}. \quad (5)$$

В итоге вероятность отсутствия некорректностей в машинном обучении после контроля $P_{\text{корр(1)}} = P_{\text{после}}$ определяется аналитическими выражениями для $P_{\text{после(1)}}$, $P_{\text{после(2)}}$, $P_{\text{после(3)}}$, $P_{\text{после(4.1)}}$, $P_{\text{после(4.2)}}$ в зависимости от варианта соотношений между исходными данными.

Для формирования исходных данных при моделировании могут использоваться статистические данные, включая данные для систем-аналогов, а также обоснованные гипотетические данные.

Для системного анализа результатов моделирования в оценках интегрального риска (см. раздел 6 статьи) рекомендуется задание допустимого уровня Рдоп корр(1) и условия α. Условие α касается не только обеспечения корректности машинного обучения при разработке ПС, но и возможного ущерба при реализации угроз. Условие α формулируется в виде ограничений: $P_{\text{корр(1)}} \ge P_{\text{доп корр(1)}}$ и возможный ущерб от нарушения не превышает допустимого (это - формулировка условия α). Учет результатов моделирования в оценках интегрального риска осуществляется с использованием индикаторного коэффициента $Z_{\text{корр(1)}}$ корректности машинного обучения при разработке ПС:

$$Z_{\text{корр(1)}} = \begin{cases} 1, \text{ если условие корректности машинного обучения при разработке ПС } \alpha \text{ выполнено,} \\ P_{\text{корр(1)}}, \text{ если условие } \alpha \\ \text{не выполнено или не задано.} \end{cases}$$

Сопоставление с возможным ущербом (или недополученным эффектом) позволяет рассматривать дополнение до единицы этого коэффициента (1 – Zкорр) в качестве вероятностного выражения риска невыявления некорректностей в машинном обучении при разработке ПС.

5. Модель для оценки риска невыявления некорректностей в машинном обучении при эксплуатации ПС

Модель позволяет оценить возможность реализации рассматриваемых угроз УБИ.222 или УБИ.221 при эксплуатации ПС по показателю вероятности получения корректных результатов машинного обучения $P_{\text{корр(2)}}$ и риска невыявления некорректностей в машинном обучении при эксплуатации ПС.

Определение: считается, что машинное обучение (дообучение) характеризуется корректностью при эксплуатации ПС в течение заданного периода прогноза, если в течение этого периода не были реализованы угрозы, связанные с использованием потенциально небезопасных версий ПС, при разработке которых могли быть использованы искаженные («отравленные») нарушителем обучающие данные или осуществлена подмена или модификация ММО. Некорректности при эксплуатации ПС – это в общем случае возникновение на временной оси негативных событий, вызванных допущенными и пропущенными ошибками при разработке ПС, уязвимостями в ПС, способствующих нарушению нормального функционирования СИИ, согласно ее назначению. Некорректности появляются в результате реализации угроз, описанных выше в разделе 2, и характеризуют отсутствие

корректности. Требуемая корректность машинного обучения при эксплуатации ПС достигается противодействием угрозам по факту выявления предпосылок или выявления непосредственного ущерба (недополученного эффекта) от реализации угроз при функционировании моделируемой системы. Корректность при эксплуатации ПС обеспечивается на основе анализа обращений пользователей на нарушения нормального функционирования СИИ с потенциально небезопасной версией ПС и/или на оперативное восстановление приемлемых условий ее функционирования (см. раздел 3). В качестве аналитика могут выступать оператор и пользователи системы, использующей СИИ, разработчик, осуществляющий сопровождение ПС, программно-аналитические инструментальные средства, ориентированные на выявление некорректностей в машинном обучении при эксплуатации ПС, или их комбинация.

Примечание. Нарушение нормального функционирования моделируемой системы должно быть определено формально. Возможно использование экспертных границ с применением универсальной вспомогательной модели показателя (УВМП) – см. раздел 8.

В моделях для оценки риска невыявления некорректностей в машинном обучении (дообучении) при эксплуатации ПС под моделируемой системой понимается множество функциональных действий модели СИИ, выполняемых с использованием потенциально небезопасных версий ПС, получаемых от разработчиков по результатам машинного обучения или дообучения.

Для моделируемой системы возможно либо отсутствие какого-либо контроля, либо периодический системный контроль хода выполнения функциональных действий. Предлагаемые вероятностные модели и методы адаптируют авторские вероятностные подходы, которые доведены до уровня реализации в ГОСТ Р 59341 (из-за изложения модели в этом стандарте, а также в [5,6, 9-14], она не приводится в полном объеме).

Моделируемая система представлена в виде «черного ящика». Специфика состоит в логическом переопределении исходных данных для моделирования. С формальной точки зрения результатом применения модели с учетом возможного ущерба (недополученного эффекта) является расчетный риск невыявления некорректностей в машинном обучении (дообучении) при эксплуатации ПС в моделируемой системе в течение заданного периода прогноза при реализации периодического системного контроля. Для расчета риска в моделируемой системе сложной структуры для каждого элемента используются исходные данные:

σ – частота возникновения источников угроз возникновения небезопасных версий ПС, при разработке которых были использованы искаженные («отравленные») нарушителем обучающие данные или была осуществлена подмена или модификация ММО;

β – среднее время развития угроз с момента их возникновения до нарушения нормального функционирования моделируемой системы;

$T_{меж}$ – среднее время между окончанием предыдущей и началом очередной диагностики целостности моделируемой системы;

$T_{диаг}$ – среднее время системной диагностики целостности моделируемой системы;

$T_{восст}$ – среднее время восстановления нарушаемой целостности моделируемой системы;

$T_{зад}$ – задаваемая длительность периода прогноза.

В итоге расчетная вероятность корректного машинного обучения ПС характеризуется вероятностью отсутствия нарушений целостности моделируемой системы в течение периода прогноза $T_{зад}$ и определяется теми же аналитическими выражениями (В.1) – (В.9), что и в моделях В.2.2, В.2.3, В.2.4 из ГОСТ Р 59341, в зависимости от варианта соотношений между исходными данными.

Сопоставление с возможным ущербом (или недополученным эффектом) позволяет рассматривать расчетную вероятность по формуле (В.1) как риск невыявления некорректностей в машинном обучении при эксплуатации ПС $P_{корр(2)}$ в моделируемой системе при реализации предпринимаемых технологических мер периодического системного контроля и восстановления целостности моделируемой системы. Вероятностное значение этого риска представляет собой дополнение до единицы вероятности корректного машинного обучения ПС в течение заданного периода прогноза.

В частном случае, когда период между диагностиками больше периода прогноза $T_{зад} < T_{меж}$, модель применима для прогноза риска при отсутствии какого-либо контроля.

Для системного анализа результатов моделирования в оценках интегрального риска (см. раздел 6) рекомендуется задание допустимого уровня $P_{доп\ корр(2)}$ и условия α. Условие α касается не только обеспечения корректности машинного обучения при эксплуатации ПС, но и возможного ущерба при реализации угроз. Условие α формулируется в виде ограничений: $P_{корр(2)} \geq P_{доп\ корр(2)}$ и возможный ущерб от нарушения не превышает допустимого (это - формулировка условия α). Учет результатов моделирования в оценках интегрального риска осуществляется с использованием

индикаторного коэффициента $Z_{корр(2)}$ ($T_{зад}$) корректности машинного обучения при эксплуатации ПС:

$$Z_{корр(2)}(Тзад) = \begin{cases} 1, \text{ если условие корректности машинного обучения при эксплуатации ПС } \alpha \text{ выполнено,} \\ P_{корр(2)}, \text{ если условие } \alpha \text{ не выполнено или не задано.} \end{cases}$$

Сопоставление с возможным ущербом (или недополученным эффектом) позволяет рассматривать дополнение до единицы этого коэффициента ($1 - Z_{корр(2)}(T_{зад})$) в качестве вероятностного выражения риска невыявления некорректностей в машинном обучении (дообучении) при эксплуатации ПС.

6. Метод оценки интегрального риска

Показатель интегрального риска нарушения корректности машинного обучения в моделируемой СИИ позволяет оценить способность нормального функционирования системы в условиях потенциальных угроз злоумышленной подмены и/или модификации ММО. Интегральный риск используется для сравнения весомости прогнозируемых частных рисков, выявления существенных угроз и поддержки принятия решений для задач системного анализа при разработке и эксплуатации моделируемой системы.

В качестве интегрального предлагается виртуальный показатель $\boldsymbol{R}_{интегр}(\boldsymbol{T}_{зад})$ риска нарушения корректности машинного обучения в условиях рассматриваемых угроз моделируемой СИИ, учитывающий в течение задаваемого периода прогноза $\boldsymbol{T}_{зад}$ риск невыявления некорректностей в машинном обучении (дообучении) при разработке ПС и риск невыявления некорректностей в машинном обучении (дообучении) при эксплуатации ПС. С учетом дополнительных условий α, а также в условиях независимости случайных событий (см. раздел 3) этот показатель может быть рассчитан с использованием моделей разделов 4 и 5:

$$R_{нтегр}(T_{зад}) = 1 - Z_{корр(1)} \cdot Z_{корр(2)}(T_{зад}).$$

Примечание. Для более общего случая модели угроз безопасности информации, учитывающей различные виды угроз, могут быть использованы другие вероятностные модели – см., например, ГОСТ Р 59341, ГОСТ Р 59346, ГОСТ Р 59349, ГОСТ Р 59989, ГОСТ Р 59991 и др.

7. Пример оценки риска при разработке ПС

Уже сегодня количество систем, использующих СИИ в различных сферах человеческой деятельности, измеряется многими тысячами, а с широким внедрением Интернета вещей и развитием «умных» систем в ближайшем будущем это количество возрастет на порядки. Тем не менее проблематика количественных оценок исследуемых рисков в России только начинает разворачиваться, критичных случаев злоумышленных модификаций ММО в СИИ не наблюдалось (мошенничество в финансовой сфере – это в общем случае комплекс более специфичных угроз, требующих специального исследования). Соответственно статистика для формирования исходных данных в интересах анализа угроз злоумышленной модификации ММО для СИИ на сегодня практически отсутствует. Поэтому в примере используются правдоподобные гипотетические исходные данные для ориентировочной оценки возможностей наличия некорректностей в машинном обучении при разработке ПС для СИИ.

Положим, по одному исследуемому объекту (например, связанному с распознаванием лиц или документов, строений или сооружений и их местоположений) объем контролируемой информации измеряется различными артефактами общим количеством 1010 у.е. (например, это могут быть параметры объектов, количество строк текста, алгоритмов, обучающих фотографий, меток и опорных векторов, действий, количество нарушений нормального функционирования ПС при тестировании и др.). Т.е. объем информации, подлежащий контролю, для определенности может быть оценен числом V =1010 у.е.

Примечание. Должно быть дано формальное содержательное наполнение у.е. контролируемого объема артефактов при машинном обучении.

В качестве контролера выступает человек – один или несколько разработчиков ПС, учитель, тестировщик или аналитик (в т.ч. лицо, принимающее решение). При этом контроль, как правило, осуществляется не только и не столько по результату, сколько в ходе работ, связанных с машинным обучением (например, в режиме разделения времени «обучение-контроль»). С точки зрения математического моделирования контролеры совместно со средствами, ориентированные на выявление некорректностей в машинном обучении при разработке ПС, представляют собой единое целое.

Часть важной для принятия решения информации, которая должна быть объективно использована при контроле информации в заданном объеме V, рассматривается на уровне до 100% от анализируемого объема в у.е., для определенности положим μ=50%, полагая, что при исследованиях возможны изменения до

100%. Скорость контроля для человека положим вполне реальные 20 у.е. в час, т.е. υ =20 у.е. в час. Период непрерывной работы контролера полагаем равным 1 часу, после чего следует восстановительный отдых, т.е. $T_{непр}$ =1 час. Предположим, что наработка контролера на ошибку 2-го рода (пропуск некорректности) составляет 1 год, что свойственно для специалистов квалификации выше средней, т.е. $T_{нар}$=365 суток. На практике при разработке ПС частота ошибок контроля 1-го рода на порядок меньше, нежели частота ошибок 2-го рода, поэтому соответственно положим n =0.00027 раз в сутки. Время на контроль информации задается таким образом, чтобы успеть завершить контроль всего заданного объема артефактов при установленной скорости контроля.

Тем самым все необходимые исходные данные для моделирования сформированы.

Результаты расчетов показывают, что вероятность получения корректных результатов машинного обучения $P_{корр(1)}$ = 0.994. Более того, достигается высокая степень устойчивости этих результатов (см. рис. 1-4) — вероятность получения корректных результатов машинного обучения не опускается ниже 0.988 (при ориентации на обоснование для системы-эталона по ГОСТ Р 59341, приложению Д допустимый уровень составляет не менее 0.95).

С привязкой к единой вероятностной шкале измерений в сравнении с допустимым уровнем это служит научно обоснованным доказательством несущественности рассмотренных типов угроз в рамках рассмотренного сценария.

Необходимо отметить, что эти положительные результаты получены в предположении, что частота ошибок контроля 1-го рода на порядок меньше, нежели частота ошибок 2-го рода. Это – для случая отсутствия целенаправленных действий по искажению («отравлению») обучающих данных (УБИ.221) или подмене или модификации ММО (УБИ.222).

Несколько изменим сценарий развития угроз, представив себе внедрение в состав разработчиков ПС и контролеров потенциального нарушителя (осуществляющего машинное обучение и контроль), злоумышленно реализующего угрозы УБИ.221 или УБИ.222. Сохраняя неизменными все предыдущие исходные данные для моделирования, проведем дополнительные исследования, изменив лишь частоту

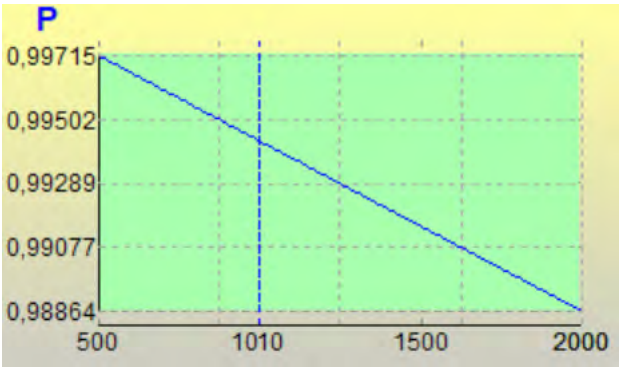

Рис.1. Зависимость вероятности получения корректных результатов машинного обучения от контролируемого объема артефактов (в у.е.)

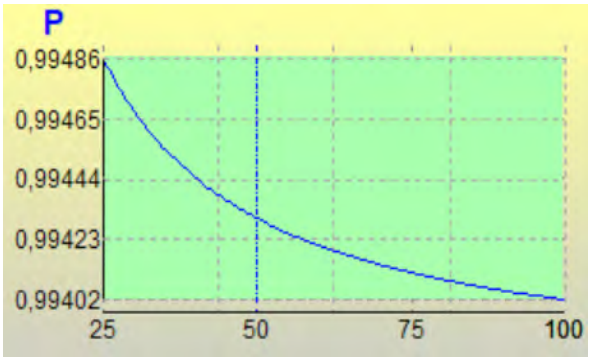

Рис.2. Зависимость вероятности получения корректных результатов машинного обучения от части важной для принятия решения информации (в %)

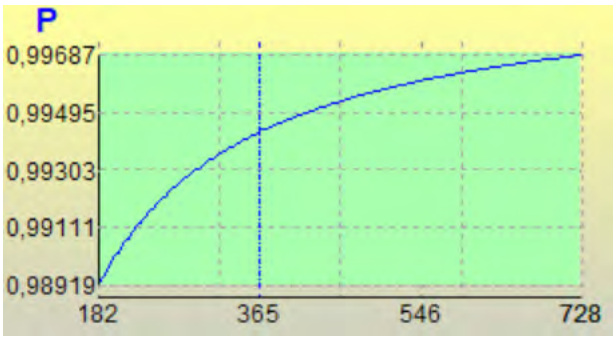

Рис.3. Зависимость вероятности получения корректных результатов машинного обучения от наработки на алгоритмическую ошибку (в сутках)

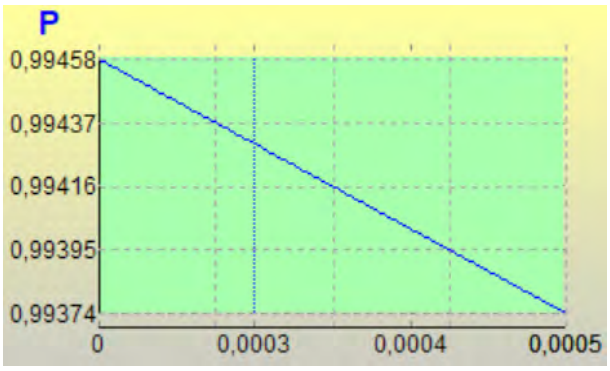

Рис.4. Зависимость вероятности получения корректных результатов машинного обучения от частоты ошибок контроля 1-го рода (раз в сутки)

ошибок контроля 1-го рода (когда несущественная для принятия решения информация ошибочно воспринимается в качестве важной), а именно: сделаем частоту ошибок контроля 1-го рода на порядок больше, нежели частота ошибок 2-го рода, т.е. положим n =0.027 раз в сутки.

Результаты расчетов показывают, что в точке расчета вероятность получения корректных результатов машинного обучения при разработке ПС $P_{корр(1)}$ = 0.939. Это меньше, нежели допустимый уровень 0.95 при ориентации на обоснование для системы-эталона по ГОСТ Р 59341, приложению Д (для вероятности получения корректных результатов обработки информации).

Примечание. При ориентации на прецедентный принцип допустимый уровень для $P_{корр(1)}$ по ГОСТ Р 59341, приложению Д соответствует уровню 0.90.

Более детальные оценки показали следующее. При прочих неизменных условиях контролируемый объем артефактов очень критичен с точки зрения получения корректных результатов машинного обучения – см. рис. 5. Так, при возрастании контролируемого объема до 2000 у.е. вероятность получения корректных результатов машинного обучения падает до 0.88. А допустимый уровень 0.95 будет преодолен, если контролируемый объем артефактов при прочих равных условиях не будет превышать 817 у.е. По этой причине актуальной для снижения риска невыявления некорректностей в машинном обучении при разработке ПС является следующая рекомендация: контролерам качества машинного обучения по возможности следует отбирать для проверки наиболее важные артефакты так, чтобы общее их количество в контролируемом объеме артефактов не превышало 817 у.е. Если этого достичь не удается, следует стараться применять рекомендации, излагаемые далее.

Часть важной для принятия решения информации, которая должна быть объективно использована при контроле информации в заданном объеме артефактов практически не критична – см. рис. 6. Это означает, что в условиях моделирования вся важная информация будет принята контролером во внимание. Скорость контроля и период непрерывной работы контролера практически не критичны. Вместе с тем сравнительно низкое абсолютное значение достигаемой вероятности получения корректных результатов машинного об-

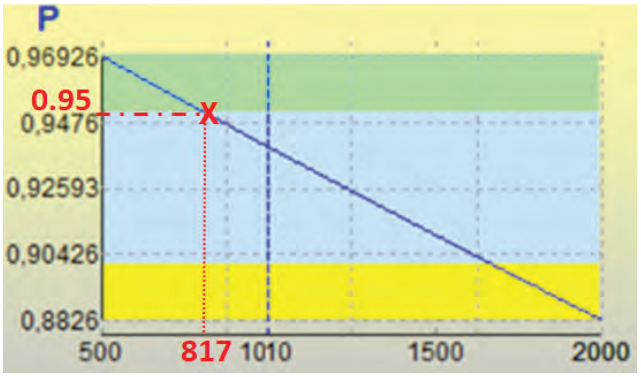

Рис.5. Зависимость вероятности получения корректных результатов машинного обучения от контролируемого объема артефактов (в у.е.)

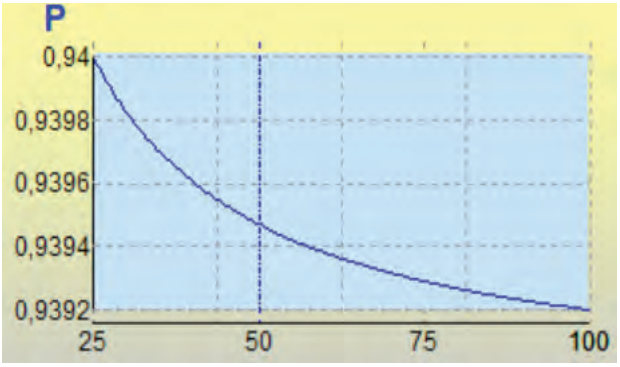

Рис.6. Зависимость вероятности получения корректных результатов машинного обучения от части важной для принятия решения информации (в %)

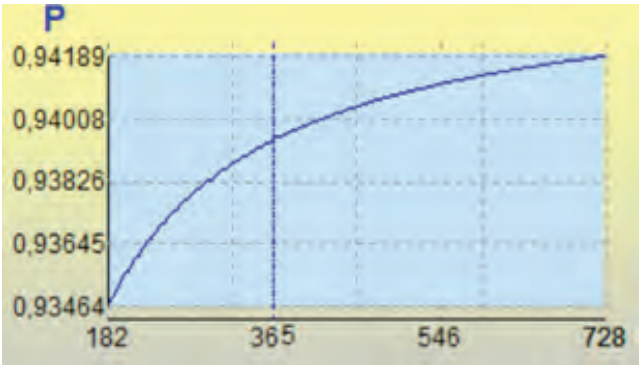

Рис.7. Зависимость вероятности получения корректных результатов машинного обучения от наработки на алгоритмическую ошибку (в сутках)

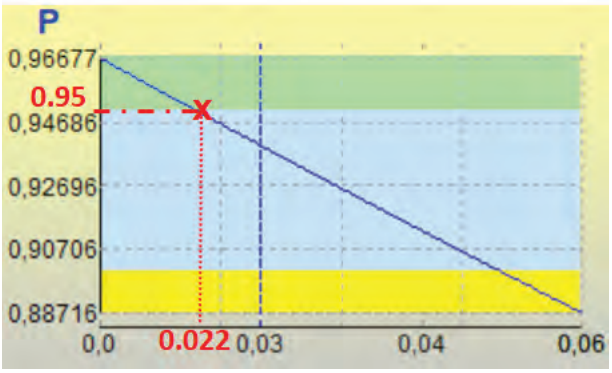

Рис.8. Зависимость вероятности получения корректных результатов машинного обучения от частоты ошибок контроля 1-го рода (раз в сутки)

учения (ниже 0.94) говорит о том, что снижения риска невыявления некорректностей в машинном обучении (дообучении) при разработке ПС следует искать в улучшении значений других параметров.

При прочих неизменных условиях в сравнении с ошибками 2-го рода частота ошибок контроля 1-го рода очень критична для получения корректных результатов машинного обучения – см. рис. 7-8. Так, при возрастании частоты ошибок контроля 1-го рода вдвое с 0.03 до 0.06 раз в сутки вероятность получения корректных результатов машинного обучения монотонно убывает с уровня 0.939 до 0.887. Это подчеркивает актуальность повышения квалификации контролеров машинного обучения. А допустимый уровень 0.95 будет преодолен, если частота ошибок контроля 1-го рода будет не выше 0.022 раз в сутки (что составляет приблизительно 8 раз в год).

Общая рекомендация: целесообразно отслеживать соотношение ошибок контроля 1-го и 2-го рода, не допуская превалирования ошибок 1-го рода (когда несущественная для принятия решения информация ошибочно воспринимается в качестве важной). Заметное превалирование ошибок 1-го рода является явным фактором возрастания риска невыявления некорректностей в машинном обучении при разработке ПС.

8. Предложения по формированию исходных данных для прогнозирования рисков

Выявление некорректностей в машинном обучении при эксплуатации ПС – это очень сложная практическая задача (в народном фольклоре она сродни ситуации, когда на вопрос учителя «Почему у Вас плохие результаты?» следует ответ ученика: «А Вы нас так учили»). В буквальном смысле как обучили ПС, такие прагматические эффекты и будут иметь место с точки зрения применения СИИ по назначению. Формально границы ожидаемых приемлемых эффектов применения СИИ должны быть определены. Например:

- недопустимое время простоя оборудования (использующего СИИ) на объекте с непрерывным производством, влекущее за собой сокращение прибыли или ущербы, должно составлять в среднем не более 0.5 часа за один останов оборудования и не более 4-х раз в месяц (в техническом задании на систему это требование бизнеса преобразуется чисто в техническое требование, к примеру: должна быть обеспечена приемлемая надежность выполнения функций системой в течение года – с вероятностью не ниже 0.995 при среднем времени восстановления после отказа не более 0.5 часа);
- приемлемый эффект применения навигаторов транспортного средства – не менее 99.9% адекватности в навигации на заданной территории;
- приемлемая удовлетворенность клиентов от использования биометрической системы платежей в метрополитене по сравнению с другими средствами платежей - не менее 88%;
- прирост числа пациентов, для которых с применением СИИ установлен верный диагноз на ранней стадии опасного заболевания должен составлять не менее 20% по сравнению с обычным диагностированием;
- число адекватно распознанных номеров транспортных средств нарушителей на автомобильных дорогах должно быть не менее 95%;
- приемлемый уровень экономии энергии в «умном» доме – не менее 25% по сравнению обычными домами, не оснащенными СИИ, и т.п.

Это – системный взгляд с одной стороны (со стороны лиц, ожидающих успешных результатов применения систем, использующих СИИ). При этом даже с использованием СИИ неизбежны случайные ошибки человека.

С другой стороны на практике просматриваются два основных варианта создания и эксплуатации ПС, в которых реализуются результаты машинного обучения:

- вариант 1 (редкий) – разработчики ПС, осуществляющие машинное обучение и дообучение, принадлежат одной и той же головной организации, которая разрабатывает и сопровождает всю систему, использующую СИИ. В этом случае предотвращение внедрения злоумышленников в состав разработчиков ПС, осуществляющих машинное обучение (дообучение), и всесторонний контроль – это прерогатива заказчика и разработчика системы. Угрозы подмены и/или модификации ММО слабоактуальны, риски пренебрежимо малы;
- вариант 2 (распространенный) – разработчики ПС, осуществляющие машинное обучение (дообучение), принадлежат сторонним организациям относительно разработчика системы, использующей СИИ. Взаимоотношения заказчик – разработчик для этого варианта подробно описаны во введении при обосновании актуальности настоящей работы. В этом случае угрозы становятся актуальными, риски могут оказаться недопустимо большими.

В случае варианта 2 становится остро востребованной предложенная модель для оценки риска невыявле-

ния некорректностей в машинном обучении при эксплуатации ПС (см. раздел 5). Однако здесь, если относительно определения средних времен между окончанием предыдущей и началом очередной диагностики целостности моделируемой системы ($T_{меж}$) и непосредственно самой системной диагностики целостности ($T_{диаг}$) трудностей не возникает, то с учетом отсутствия какой-либо статистики у аналитика встает правомерный вопрос – как приблизительно можно определить такие исходные данные, как частота возникновения источников угроз возникновения небезопасных версий ПС, при разработке которых были использованы искаженные («отравленные») нарушителем обучающие данные или была осуществлена подмена или модификация ММО (σ), среднее время развития угроз с момента их возникновения до нарушения нормального функционирования моделируемой системы (β), а также среднее время восстановления нарушаемой целостности моделируемой системы ($T_{восст}$).

Для ответа на этот вопрос предлагается использование универсальной вспомогательной модели показателя (УВМП) по ГОСТ Р 59349 «Системная инженерия. Защита информации в процессе системного анализа».

В любой момент времени у ответственных лиц, принимающих решение, имеет место формальное представление о том, какое состояние эксплуатируемой системы, использующей СИИ, «нормально» и «приемлемо», а какое «неприемлемо» и требует управляющей реакции для улучшения.

Т. е. на любой момент времени по каждому из критичных показателей (или по их совокупности) можно с однозначной уверенностью определить, что его (их) значения находятся в состоянии, которое может быть охарактеризовано как «Приемлемое» или «Приемлемое с отклонением» (когда за счет определенных организационных или обычных технических усилий по улучшению значения критичного показателя можно удерживать систему от перехода этого показателя в зону «Неприемлемого» состояния) или как «Неприемлемое» состояние (когда требуются кардинальные решения по восстановлению условий, которые в существующем виде уже не обеспечивают или в ближайшее время при бездействии не будут гарантировать требуемого уровня эффективности системы) – см. рис. 9. Переход критичного показателя в состояние «Неприемлемое» характеризует подозрение, что в ПС системы были реализованы потенциальные угрозы подмены и/или модификации ММО. Например, в качестве критичных показателей могут быть использованы показатели, перечисленные в начале этого раздела, а их допустимая граница – это вышеуказанные допустимые значения, ухудшение которых с точки зрения прагматического эффекта для системы характеризует зону состояния, именуемого как «Неприемлемое». Выбранные критичные показатели при существенном ухудшении их значений относительно установленных пределов до состояния «Неприемлемое» могут служить показателями возникновения некорректностей в машинном обучении при эксплуатации ПС. А граница «Приемлемое с отклонением» характеризует те некоторые уступки по сравнению с наилучшим достигнутым результатом для критичного показателя, которые могут быть допущены с учетом имеющих место неопределенностей.

При этом становятся определенными недостающие исходные данные σ, β, $T_{восст}$. Эти исходные данные формируются по следующему алгоритму, описанному ниже в привязке к регистрируемым значениям критичного показателя на рис. 9.

Частота возникновения источников угроз возникновения небезопасных версий ПС, при разработке которых были использованы искаженные («отравленные») нарушителем обучающие данные или была осуществлена подмена или модификация ММО, определяется выражением:

$$\sigma = 1/[(\tau_{возн.1} + \tau_{возн.2} + \tau_{возн.3} + \tau_{возн.4})/4].$$

Среднее время развития угроз с момента их возникновения до нарушения нормального функционирования моделируемой системы определяется выражением:

$$\beta = (\tau_{разв.1} + \tau_{разв.2} + \tau_{разв.3} + \tau_{разв.4} + \tau_{разв.5})/5.$$

Среднее время восстановления нарушаемой целостности моделируемой системы определяется выражением:

$$T_{восст} = (\tau_{восст.1} + \tau_{восст.2} + \tau_{восст.3})/3.$$

Здесь $\tau_{возн.i}$ – i-й интервал времени между возникновениями источника угроз; $T_{разв.j}$ – j-й интервал времени развития угроз с момента возникновения источника угроз до нарушения нормальных условий; $T_{восст.m}$ – m-й интервал времени восстановления нарушаемой целостности.

Значения σ, β, $T_{восст}$, получаемые в результате применения предложенного выше алгоритма к статистическим данным контроля на уровне УВМП, являются исходными данными для формального описания моделируемой системы с учетом возможности прогнозирования динамики разнородных событий. Роль в УВМП каждого из учитываемых критичных показателей сводится к определению σ, β, $T_{восст}$ для последующего моделирования.

Рис.9. Элементарные состояния контролируемого показателя УВМП во времени и временные характеристики для прогнозирования рисков

Применительно к конкретной СИИ остается установить границы для зон «Приемлемое», «Приемлемое с отклонением», «Неприемлемое». Тогда УВМП превращается в некий эталон, ухудшение показателя по которому до состояния «Неприемлемое» характеризует подозрение, что в ПС системы были реализованы потенциальные угрозы подмены и/или модификации ММО. Отличительная особенность изначальной версии этого эталона, который должен быть признан и утвержден, в том, что наблюдаемый приемлемый эффект от применения системы должен быть получен в полной уверенности в корректности соответствующих обученных ПС, т.е. в полной уверенности того, что если и были потенциальные угрозы, рассматриваемые в статье, то они не были реализованы в соответствующих ПС, которые применялись при формировании изначального эталона.

Сегодняшний период развития СИИ в России примечателен тем, что потенциальные угрозы могут быть реализованы главным образом из-за случайных ошибок, нежели из злоумышленных намерений нарушителя. Поэтому есть достаточно высокая уверенность в том, что заказчики систем, использующих СИИ, и организации-разработчики соответствующих ПС не только тщательно подходят к отбору кадров, но и многие математические вопросы по машинному обучению в условиях дефицита высококвалифицированных специалистов решаются сообща (создавая тем самым условия взаимоконтроля). В связи с этим предлагается в качестве эталона, ориентированного на УВМП, и начальных границ зон «Приемлемое», «Приемлемое с отклонением», «Неприемлемое» брать те значения, которые характеризуют достижение прагматического эффекта сразу, как только он появляется (по сравнению со случаем функционирования системы без использования СИИ).

С учетом широкомасштабных работ в области искусственного интеллекта эта ситуация может поменяться буквально через несколько лет. И тогда изначальная версия эталона по УВМП послужит отправной пограничной полосой для обоснованных подозрений о наличии или отсутствии реализации угроз подмены и/или модификации ММО. По мере эксплуатации системы и сбора соответствующей статистики этот изначальный эталон может быть усовершенствован.

9. Относительно примера оценки риска при эксплуатации ПС

В качестве примера оценки риска невыявления некорректностей в машинном обучении при эксплуатации ПС читателю рекомендуется работа «Подход к вероятностному прогнозированию защищенности репутации политических деятелей от «фейковых» угроз в публичном информационном пространстве» [15]. Достаточно представить себе, что речь идет об избирательной системе в расширенном ее понимании, где в систему входят средства массовой информации, использующие СИИ, которые целенаправленно реализуют рассмотренные в настоящей статье угрозы (в виде «фейковых» результатов машинного обучения) против репутации политических деятелей. В статье определены количественные границы относительно вероятностей сохранения и дискредитации изначально положительной репутации виртуального политического деятеля в условиях правового законодательства в России в период с конца 90-х по 2023гг. Выявлено, что в условиях отсутствия правовых норм по ограничению длительности рассмотрения исков в защиту репутации политического деятеля в России наблюдается недопустимо низкая степень защищенности изначально положительной репутации от таких «фейков», которые могут быть усилены потенциальными возможностями технологий нейролингвистического программирования и специальных политтехнологий психологического воздействия на электорат. Обоснованы востребованные способы защищенности репутации политических деятелей, включая комплексные меры мониторинга и выявления угроз, развития системы правосудия в защиты репутации политического деятеля с указанием количественных характеристик противодействия «фейковым» угрозам. Иными словами, если «фейки» рассматривать как результат реализации рассматриваемых угроз злоумышленной модификации ММО для СИИ, задействованных в избирательных кампаниях политических деятелей, то материалы статьи [15] могут послужить непосредственным примером оценки риска невыявления некорректностей в машинном обучении при эксплуатации ПС.

Заключение

1. Для систем, использующих СИИ, проведен анализ актуальных угроз подмены ММО (УБИ.222) и модификации ММО путем искажения («отравления») обучающих данных (УБИ.221). В условиях принятых предположений и допущений разработаны вероятностные модели для оценки частных рисков невыявления некорректностей в машинном обучении (дообучении) при разработке и эксплуатации ПС, а также метод оценки интегрального риска нарушения корректности машинного обучения в течение задаваемого периода прогноза.

2. Риск невыявления некорректностей в машинном обучении (дообучении) при разработке ПС предложено оценивать в зависимости следующих исходных данных: объема информации по обучаемым ПС, подлежащего контролю; части важной для принятия решения информации, которая должна быть объективно использована при контроле информации в заданном объеме; скорости контроля; частоты ошибок контроля 1-го рода; среднего времени наработки на алгоритмическую ошибку; периода непрерывной работы контролера; задаваемого времени на контроль информации.

3. Риск невыявления некорректностей в машинном обучении (дообучении) при эксплуатации ПС предложено оценивать в зависимости следующих исходных данных: частоты возникновения источников угроз возникновения небезопасных версий ПС, при разработке которых были использованы искаженные («отравленные») нарушителем обучающие данные или была осуществлена подмена модели машинного обучения; среднего времени развития угроз с момента их возникновения до нарушения нормального функционирования моделируемой системы; среднего времени между окончанием предыдущей и началом очередной диагностики целостности моделируемой системы; среднего времени системной диагностики целостности моделируемой системы; среднего времени восстановления нарушаемой целостности моделируемой системы; задаваемой длительности периода прогноза. В интересах формирования необходимых исходных данных для последующего моделирования предложено использовать универсальную вспомогательную модель показателя по ГОСТ Р 59349, адаптированную для анализа рассматриваемых угроз.

4. Интегральный риск предложено оценивать через виртуальный показатель риска нарушения корректности машинного обучения в условиях рассматриваемых угроз в течение задаваемого периода прогноза в зависимости от рисков невыявления некорректностей в машинном обучении (дообучении) при разработке и эксплуатации ПС, а через них – в зависимости от исходных данных, обеспечивающих расчет соответствующих рисков.

5. Предложенный методический аппарат, позволяет осуществлять вероятностную оценку корректности обучаемых ПС в системах, использующих СИИ, при их

разработке и эксплуатации в условиях потенциальных угроз злоумышленной подмены и/или модификации ММО. Работоспособность подхода проиллюстрирована на примерами.

Литература

1. Эртель В., Введение в искусственный интеллект.-М. «Эксмо», 2019. – 448с.
2. Лекун Ян, Как учится машина (революция в области нейронных сетей и глубокого обучения). – М. Альпина PRO, 2021. – 335с.
3. Арлазаров В. В., Мобильное распознавание и его применение к системе ввода идентификационных документов. – Диссертация на соискание ученой степени доктора технических наук. -М. ФИЦ ИУ РАН, 2023. – 358с.
4. Chakraborty A., Alam M., Dey V., Chattopadhayay A.U., Yay D.M., Adversarial attacks and defences: A survey //arXiv preprint arXiv:1810.00069. – 2018
5. Probabilistic modeling in system engineering. InTechOpen, 2018, 279p. – URL: http://www.intechopen.com/books/probabilistic-modeling-in-system-engineering
6. Климов С. М. Модели анализа и оценки угроз информационно-психологических воздействий с элементами искусственного интеллекта. / Сборник докладов и выступлений научно-деловой программы Международного военно-технического форума «Армия-2018». 2018. С. 273–277.
7. Манойло А. В., Петренко А. И., Фролов Д. Б. Государственная информационная политика в условиях информационно-психологической войны. 4-е изд., перераб. и доп. — Горячая линия-Телеком Москва, 2020. — 636 с.
8. Костогрызов А. И. Прогнозирование рисков по данным мониторинга для систем искусственного интеллекта / БИТ. Сборник трудов Десятой международной научно-технической конференции – М.: МГТУ им. Н.Э. Баумана, 2019, с. 220-229.
9. A. Kostogryzov and V. Korolev, Probabilistic Methods for Cognitive Solving of Some Problems in Artificial Intelligence Systems (Вероятностные методы для когнитивного решения некоторых задач в системах искусственного интеллекта). Probability, combinatorics and control / IntechOpen, 2020, pp. 3-34. — URL: https://www.intechopen.com/books/probability-combinatorics-and-control
10. Kostogryzov A., Nistratov A., Nistratov G. (2020) Analytical Risks Prediction. Rationale of System Preventive Measures for Solving Quality and Safety Problems. In: Sukhomlin V., Zubareva E. (eds) Modern Information Technology and IT Education. SITITO 2018. Communications in Computer and Information Science, vol 1201. Springer, pp.352-364. https://www.springer.com/gp/book/9783030468941
11. Kostogryzov A, Nistratov A., Probabilistic methods of risk predictions and their pragmatic applications in life cycle of complex systems. In “Safety and Reliability of Systems and Processes”, Gdynia Maritime University, 2020. pp. 153-174. DOI: 10.26408/srsp-2020
12. Нистратов А.А., Аналитическое прогнозирование интегрального риска нарушения приемлемого выполнения совокупности стандартных процессов в жизненном цикле систем высокой доступности. Часть 1. Математические модели и методы // Системы высокой доступности. 2021. Т.17 №3, с. 16–31, Часть 2. Программно-технологические решения. Примеры применения // Системы высокой доступности. 2022. Т.18 №2, с. 42–57.
13. Костогрызов А.И. О моделях и методах вероятностного анализа защиты информации в стандартизованных процессах системной инженерии //Вопросы кибербезопасности. 2022, №6(52), с.71-82.
14. Kostogryzov A., Makhutov N., Nistratov A., Reznikov G., Probabilistic predictive modeling for complex system risk assessments (Вероятностное упреждающее моделирование для оценок рисков в сложных системах). Time Series Analysis - New Insights. IntechOpen, 2023, pp. 73-105. http://mts.intechopen.com/articles/show/title/probabilistic-predictive-modelling-for-complex-system-risk-assessments
15. Костогрызов А. И. Подход к вероятностному прогнозированию защищенности репутации политических деятелей от «фейковых» угроз в публичном информационном пространстве // Вопросы кибербезопасности. 2023, №3. С. 114–133. DOI:1021681/2311-3456-2023-3-114-133

THREAT ANALYSIS OF MALICIOUS MODIFICATION OF THE MACHINE LEARNING MODEL FOR ARTIFICIAL INTELLIGENCE SYSTEMS

Kostogryzov A.I[7], Nistratov A.A.[8]

Objective: *to propose a methodological approach for probabilistic analysis of the correctness of the trained software tools (SW) for artificial intelligence systems (AIS) during their development and operation in conditions of potential threats of malicious modification of the machine learning model (MLM).*

7 Andrey I. Kostogryzov, Dr.Sc., Professor, Chief Researcher, Federal Research Center «Informatics and Control» of the Russian Academy of Sciences. Moscow, Russia. E-mail: Akostogr@gmail.com

8 Andrey A. Nistratov, PhD, Senior researcher, Federal Research Center «Informatics and Control» of the Russian Academy of Sciences. Moscow, Russia. E-mail: andrey.nistratov.job@yandex.ru

Research methods include *methods of probability theory, methods of system analysis. The approach is based on the adaptation of the author's probabilistic models developed earlier to assess the quality of the information used and risk management, which are fixed to the level of implementation in GOST R 59341-2021 "System engineering. Protection of information in system information management process".*

Result: *Under the conditions of accepted suppositions and assumptions, probabilistic models have been developed to assess the particular risks of non-detection of inaccuracies in machine learning during the development and operation of SW, as well as a method for assessing the integral risk of violation of the correctness of machine learning during specified period of prediction. Actual threat of MLM spoofing and the threat of MLM modification by poisoning the training data are analyzed. Proposals have been developed for the formation of input for risks prediction using the proposed models. The approach is illustrated by calculation examples with quantitative assessments, risk dependencies on the input and the rationale of recommendations.*

Scientific novelty: *For the conditions of potential threats of malicious MLM modification, models and methods for assessing the particular risks of non-detection of incorrectness in machine learning during AIS development and operation and the integral risk are proposed.*

Keywords: *probability, poisoning of training data, model, risk, system, threats.*

References

1. E`rtel` V., Vvedenie v iskusstvenny`j intellekt.-M. «E`ksmo», 2019. – 448s.
2. Lekun Yan, Kak uchitsya mashina (revolyuciya v oblasti nejronny`x setej i glubokogo obucheniya). – M. Al`pina PRO, 2021. – 335s.
3. Arlazarov V. V., Mobil`noe raspoznavanie i ego primenenie k sisteme vvoda identifikacionny`x dokumentov. – Dissertaciya na soiskanie uchenoj stepeni doktora texnicheskix nauk. -M. FICz IU RAN, 2023. – 358s.
4. Chakraborty A., Alam M., Dey V., Chattopadhayay A.U., Yay D.M., Adversarial attacks and defences: A survey //arXiv preprint arXiv:1810.00069. – 2018
5. Probabilistic modeling in system engineering. InTechOpen, 2018, 279p. – URL: http://www.intechopen.com/books/probabilistic-modeling-in-system-engineering
6. Klimov S. M. Modeli analiza i ocenki ugroz informacionno-psixologicheskix vozdejstvij s e`lementami iskusstvennogo intellekta. / Sbornik dokladov i vy`stuplenij nauchno-delovoj programmy` Mezhdunarodnogo voenno-texnicheskogo foruma «Armiya-2018». 2018. S. 273–277.
7. Manojlo A. V., Petrenko A. I., Frolov D. B. Gosudarstvennaya informacionnaya politika v usloviyax informacionno-psixologicheskoj vojny`. 4-e izd., pererab. i dop. — Goryachaya liniya-Telekom Moskva, 2020. — 636 s.
8. Kostogry`zov A. I. Prognozirovanie riskov po danny`m monitoringa dlya sistem iskusstvennogo intellekta / BIT. Sbornik trudov Desyatoj mezhdunarodnoj nauchno-texnicheskoj konferencii – M.: MGTU im. N.E`. Baumana, 2019, s. 220-229.
9. A. Kostogryzov and V. Korolev, Probabilistic Methods for Cognitive Solving of Some Problems in Artificial Intelligence Systems (Veroyatnostny`e metody` dlya kognitivnogo resheniya nekotory`x zadach v sistemax iskusstvennogo intellekta). Probability, combinatorics and control / IntechOpen, 2020, pp. 3-34. — URL: https://www.intechopen.com/books/probability-combinatorics-and-control
10. Kostogryzov A., Nistratov A., Nistratov G. (2020) Analytical Risks Prediction. Rationale of System Preventive Measures for Solving Quality and Safety Problems. In: Sukhomlin V., Zubareva E. (eds) Modern Information Technology and IT Education. SITITO 2018. Communications in Computer and Information Science, vol 1201. Springer, pp.352-364. https://www.springer.com/gp/book/9783030468941
11. Kostogryzov A, Nistratov A., Probabilistic methods of risk predictions and their pragmatic applications in life cycle of complex systems. In "Safety and Reliability of Systems and Processes", Gdynia Maritime University, 2020. pp. 153-174. DOI: 10.26408/srsp-2020
12. Nistratov A.A., Analiticheskoe prognozirovanie integral`nogo riska narusheniya priemlemogo vy`polneniya sovokupnosti standartny`x processov v zhiznennom cikle sistem vy`sokoj dostupnosti. Chast` 1. Matematicheskie modeli i metody` // Sistemy` vy`sokoj dostupnosti. 2021. T.17 №3, s. 16–31, Chast` 2. Programmno-texnologicheskie resheniya. Primery` primeneniya // Sistemy` vy`sokoj dostupnosti. 2022. T.18 №2, s. 42–57.
13. Kostogry`zov A.I. O modelyax i metodax veroyatnostnogo analiza zashhity` informacii v standartizovanny`x processax sistemnoj inzhenerii //Voprosy` kiberbezopasnosti. 2022, №6(52), s.71-82.
14. Kostogryzov A., Makhutov N., Nistratov A., Reznikov G., Probabilistic predictive modeling for complex system risk assessments (Veroyatnostnoe uprezhdayushhee modelirovanie dlya ocenok riskov v slozhny`x sistemax). Time Series Analysis - New Insights. IntechOpen, 2023, pp. 73-105. http://mts.intechopen.com/articles/show/title/probabilistic-predictive-modelling-for-complex-system-risk-assessments
15. Kostogry`zov A. I. Podxod k veroyatnostnomu prognozirovaniyu zashhishhennosti reputacii politicheskix deyatelej ot «fejkovy`x» ugroz v publichnom informacionnom prostranstve // Voprosy` kiberbezopasnosti. 2023, №3. S. 114–133. DOI:1021681/2311-3456-2023-3-114-133

МНОГОУРОВНЕВАЯ КОНЦЕПЦИЯ БЕЗОПАСНОСТИ СИСТЕМ УПРАВЛЕНИЯ БОЛЬШИМИ ДАННЫМИ

Полтавцева М. А.[1], Зегжда Д. П.[2], Калинин М.О.[3]

Цель исследования. *Технологии и системы управления большими данными являются основой огромного числа современных цифровых сервисов. С одной стороны, они построены на традиционных решениях, а с другой, включают новые подходы, такие, как полихранилища или аутсорсинг данных. Ключевая роль в технологическом стеке цифровой экономики и новизна определяют как привлекательность таких активов для злоумышленника, так и несовершенство методов защиты. Целью работы является анализ больших данных как объекта защиты и разработка многоуровневой концепции их безопасности на основе консистентного подхода.*

Метод исследования. *В работе используется многоуровневый подход, которому соответствует также архитектура ANSI/SPARC систем управления базами данных. Большие данные рассматриваются на трех уровнях от инфраструктуры до бизнес-логики, выделяются ключевые технологии, уязвимости и методы защиты. Также более детально в рамках ANSI/SPARC определяется уровневая архитектура систем управления большими данными на базе полихранилищ, проводится анализ их безопасности. Задается технологический базис безопасности систем управления большими данными как система распределенного динамического аудита, приводится пример такой системы на основе распределенного реестра.*

Результаты исследования. *В статье выделены три уровня рассмотрения больших данных: инфраструктурный, инженерии данных и бизнес-логики. Авторами сформулированы эволюционные изменения систем больших данных по сравнению с традиционными СУБД с точки зрения информационной безопасности. Дано понятие системы управления большими данными, определены ее собственные архитектурные уровни на основе архитектуры ANIS/SPARK, для каждого из которых выделены проблемы безопасности, причины их появления и направления развития средств защиты. Авторами выделено ключевое требование безопасности систем управления большими данными - консистентное представление на уровне общего монитора безопасности. Для его выполнения, в части сбора данных о системе, предложено использование технологий распределенного динамического аудита. Проведена апробация системы распределенного динамического аудита при управлении большими данными на базе технологии HashGraph.*

Научная новизна. *В работе впервые сформулирована многоуровневая концепция безопасности систем управления большими данными, в рамках которой выделены и систематизированы на различных уровнях ключевые уязвимости систем больших данных, отличные от других классов систем и традиционных СУБД. Впервые предложено применение технологии распределенного реестра для сбора данных о жизненном цикле информации в системе управления большими данными. Проведенные исследования позволяют более комплексно подойти к обеспечению безопасности больших данных и систем управления ими, конкретизировать и согласовать наборы методов и средств защиты, а также закладывают основу построения таких систем в защищенном исполнении.*

Ключевые слова: *информационная безопасность, безопасность больших данных, консистентный подход, архитектура безопасности, модель безопасности, безопасность полихранилищ, распределенный реестр.*

DOI: 10.21681/2311-3456-2023-5-25-36

Введение

Технологии «больших данных» в начале века во многом изменили ландшафт и архитектуру современных информационных систем. Цифровая экономика, цифровое производство, электронное правительство, искусственный интеллект – все современные цифровые системы в большей или меньшей степени осно-

1 Полтавцева Мария Анатольевна, доктор технических наук, доцент, профессор СПбПУ Петра Великого, г. Санкт-Петербург, Россия. E-mail: poltavtseva@ibks.spbstu.ru ORCHID 0000-0001-9659-1244

2 Зегжда Дмитрий Петрович, член–корреспондент РАН, доктор технических наук., профессор, профессор СПбПУ Петра Великого, г. Санкт-Петербург, Россия. E-mail: dmitry@ibks.spbstu.ru ORCHID 0000-0002-2048-6189

3 Калинин Максим Олегович доктор технических наук, профессор, профессор СПбПУ Петра Великого, г. Санкт-Петербург, Россия. E-mail: max@ibks.spbstu.ru ORCHID 0000-0002-9732-0099

Рис.1. Уровни представления больших данных

ваны на данных. Для многих из них подходы на основе данных (data-driven) являются ключевыми. В то же время, как росла зависимость цифровых сервисов от данных, росло и их влияние на жизнь каждого отдельного человека. И также росло число злоумышленников и атак в киберпространстве.

Безопасность современных цифровых сервисов напрямую зависит, в том числе, и от безопасности обрабатываемых в них данных, которые не только используются для предоставления сервиса или получения результата, услуги, но и для создания и функциональности самих цифровых решений. Например, для обучения искусственного интеллекта. Огромные массивы персональных данных в информационных системах, как государственных, так и частных, востребованы злоумышленниками. Результаты утечек информации используются как для простого получения выгоды, рекламы и перепродажи, так и во множестве мошеннических схем, проведении OSINT – исследований с различными целями.

Безопасность современных больших данных, таким образом, является одной из важных современных задач в области кибербезопасности. Несмотря на то, что исследования в этой области перешагнули десятилетней рубеж[4], проблематика сегодня остается прежней [1]. Целью данной работы является, в первую очередь, анализ больших данных и систем управления ими как объекта защиты, выявление основных технологических проблем обеспечения их защищенности на современном этапе а также разработка многоуровневой концепции безопасности на основе консистентного подхода в данной области.

4 Запечников С. В. и др. Проблемы обеспечения информационной безопасности больших данных //Безопасность информационных технологий. 2014. Т. 21.(3). С. 8-17

Большие данные как объект защиты

Несмотря на более чем десятилетнюю историю, устоявшегося и общепринятого понятия больших данных до сих пор не существует [2]. С одной стороны, проблема обработки больших объемов данных, сегодня ассоциированная с big data, была обозначена еще в середине прошлого века [2,3]. Хотя взрывной интерес к этой области и появление современной терминологии отмечается с начала двадцать первого века [4], технологические основы и вызовы, обусловленные проблемами разнородности, объема и скорости данных на уровне систем управления базами данных (СУБД) возникали и ранее, имеют не абсолютно новый, а исторический характер. С другой стороны, сегодня большие данные касаются не только технологических вызовов разработчикам СУБД и инженерам данных, но и формируют новые вызовы для организаций в целом, правовой системы, государственного управления.

Качественным изменением, по сравнению с проблемами, приведшими к появлению методов оптимизации запросов, параллельным и распределенным системам баз данных, ETL (extract, transfer, loading)-технологии является переход проблематики с уровня технических систем на организационный. Большие данные пытаются определять не только в технических терминах, но и на уровне бизнес-процессов и нормативных документов [6]. В отечественном стандарте определение также достаточно расплывчато[5]. В ито-

5 «Большие данные (big data): большие массивы данных, главным образом, по таким характеристикам данных, как объем, разнообразие, скорость обработки и/или вариативность, – которые требуют использования технологии масштабирования для эффективного хранения, обработки, управления и анализа.» ГОСТ Р ИСО/МЭК 20546–2019 Большие данные. Обзор и словарь

ге сегодня можно говорить о, по крайней мере, трех уровневой архитектуре больших данных, на каждом уровне которой существуют свои понятия, технологии, угрозы, уязвимости и методы защиты (рисунок 1).

На самом нижнем, инфраструктурном уровне сегодня сложился ряд технологий, которые также называют технологиями больших данных. Это облачные и туманные (fog) вычисления [6], центры обработки данных [7,8] и другие связанные технологии [9]. Безусловно, фактически этот уровень представляет собой общую инфраструктуру современных распределенных информационных систем и применим не только для больших данных. Однако, говоря о технологическом стеке и безопасности больших данных нельзя его проигнорировать, так как традиционные угрозы от DDoS атак до программных и аппаратных закладок актуальны для систем больших данных именно на этом уровне.

Следующий уровень относится к области инженерии данных (data engineering) и является продолжением технологий традиционных систем управления базами данных [10]. На этом уровне специалистами рассматриваются вопросы структуризации данных, построения специализированных СУБД, хранилищ, новые архитектуры обработки информации [1]. Здесь же используется и самое устоявшееся определение данного явления: под большими данными понимается информация, которая, в силу своих характеристик (скорости поступления, объема и/или разнообразия), не может быть обработана общераспространенными, “универсальными” средствами. А фактические значения характеристик данных могут различаться от задачи к задаче.

Ключевой новой технологией на этом уровне сегодня становятся полихранилища – системы, объединяющие в себе несколько СУБД на базе различных моделей данных с разными операциями и степенью структуризации информации [11,12]. Технологии безопасности традиционных СУБД сегодня достаточно развиты, хотя и требуют совершенствования, и, к сожалению, далеко не всегда поддерживаются серверами [1,13,14]. Но эти технологии не являются легко переносимыми на полихранилища, так как тесно связаны со структуризацией, грануляцией и операциями с данными внутри каждой конкретной системы управления базами данных.

Третий уровень рассмотрения больших данных во многом определяет уникальность этого явления по сравнению с проблемами роста объема и разнообразия ранее. Большие данные рассматриваются как ценный актив на уровне предприятия [15] и к ним применяются подходы управления, на их основе строятся data-driven (основанные на данных) системы принятия решений [16]. В рамках “цепочки поставок” существует несколько различных способов использования или работы с большими данными [17]. Ключевыми здесь являются как внешние источники данных, так и задачи обмена данными между организациями, передачи данных для анализа на аутсорсинг, совместное использование данных для обучения алгоритмов искусственного интеллекта. Сохранить конфиденциальность данных при этом становится достаточно сложной задачей [18], для решения которой используются криптографические технологии [19], технологии анонимизации [20], федеративное обучение искусственного интеллекта [21]. Тем не менее, все эти технологии сегодня не могут гарантировать безопасность больших данных, а только снижают вероятность их утечки. Попытки формализации комплексной безопасности на этом уровне [22] пока не получили широкого развития.

Систематизация описанных выше уровне с точки зрения кибербезопасности, от ключевых новых технологий больших данных, до эволюции угроз и принятых сегодня методов защиты вместе с ограничениями приведены в табл. 1.

В силу того, что в рамках больших данных можно выделить три технологических уровня, рационально говорить о комплексной безопасности в этой области как о согласованной системе безопасности, охватывающей все уровни представления от инфраструктурного до бизнес-логики. В то же время, так как приведенные уровни практически независимы друг от друга, что подчеркивается, в том числе, разницей в понятийном аппарате, реализация мер защиты на каждом из них также может осуществляться независимо.

В дальнейшей работе будет более детально рассмотрен уровень инженерии данных, как, с одной стороны, высоко технологичный (в отличии от уровня бизнес-логики), а с другой специфический для рассматриваемой области, в отличии от инфраструктурного. К тому же, приведенное выше определение больших данных из ГОСТ Р ИСО/МЭК 20546-2019 также относится в наибольшей степени к уровню инженерии.

Безопасность систем управления базами данных и полихранилищ в экосистеме больших данных

Первым шагом в сторону развития современных технологий и архитектур инженерии больших данных, после появления распределенных СУБД, стало появ-

Таблица 1

Систематизация характеристик уровней представления больших данных с точки зрения кибербезопасности

Параметр	Уровень бизнес-логики	Уровень инженерии данных	Уровень инфра-структуры
Ключевые изменения	Совместное использование данных. Передача данных на аутсорсинг.	Полихранилища (поли-базы данных или гетерогенные базы данных)	Распределенная инфраструктура обработки информации
Эволюциони-ровавшие угрозы	Утечки данных (внутренний нарушитель). Логический вывод над данными (inference attack)	Утечки данных (внутренний и внешний нарушитель, ошибки контроля доступа), искажение и удаление данных.	Не доверенная среда обработки
Технологии защиты	Анонимизация данных. Платформы «безопасной» аналитики. Федеративное обучение искусственного интеллекта.	Контроль доступа. Шифрование. Аудит и журналирование. Обнаружение вторжений и внутреннего нарушителя.	Безопасность облачных технологий (криптография, безопасные вычисления и др). Безопасное ПО.
Ограничения	Нет гарантированных способов защиты от атак логического вывода.	Многие технологии существуют только для отдельных типов СУБД и часто даже не внедрены в промышленные решения.	Сложные цепочки поставок, большие объемы анализируемого кода.

ление не реляционных (NoSQL) систем управления базами данных и специализация в области СУБД. В итоге для решения задач, выходящих за рамки возможностей промышленного сервера баз данных, требовалось составление комбинаций из разнородных инструментов. Такие системы в разных источниках называются полихранилищами [23], поли-базами данных [24], гетерогенными базами данных[25], гетерогенными системами баз данных[26], системами управления большими данными [27]. Помимо полихранилищ в экосистему инженерии больших данных входят также еще два класса программного обеспечения. Это, во-первых, инструменты потоковой обработки данных и, во-вторых, программы и библиотеки, отвечающие за балансировку нагрузки, преобразования данных и выполняющие другие вспомогательные задачи. Все эти инструменты в совокупности с полихранилищами можно назвать системами управления большими данными.

Рассмотрение полихранилищ с точки зрения теории управления базами данных также позволяет выделить в них уровни обработки информации. Согласно базовой архитектуре всех систем управления базами данных ANSI/SPARC, не потерявшей сегодня своей актуальности [28], выделяется три уровня:

— физический уровень, на котором осуществляется хранение и физические операЗции с данными на диске;
— логический уровень, на котором определяется внутреннее представление данных (модель данных), которое используется для манипулирования информацией в терминах не файловой системы, а семантически значимых фрагментов;
— концептуальный уровень, на котором формируются представления данных для пользователей и внешних программ.

Полихранилища имеют организацию, отличную от систем управления базами данных, так как, фактически, включают в себя несколько СУБД с различной логической (и тем более физической) организацией данных. Для больших данных уровни модели ANSI/SPARC предлагается интерпретировать следующим образом (рисунок 2).

В данной интерпретации на физический уровень выносятся традиционные СУБД, выступающие как «коробочные» решения по управлению поступающей в них информацией. Безусловно, каждый такой инструмент также может (и должен) при построении системы безопасности оцениваться на приведенных выше архитектурных уровнях. Но, с точки зрения полихранилища в целом, манипулирование данными внутри такого инструмента не интерпретируемо.

Логический уровень представляет собой комбинацию структур данных в рамках каждого инструмента. Используемые внутренними СУБД полихранилища

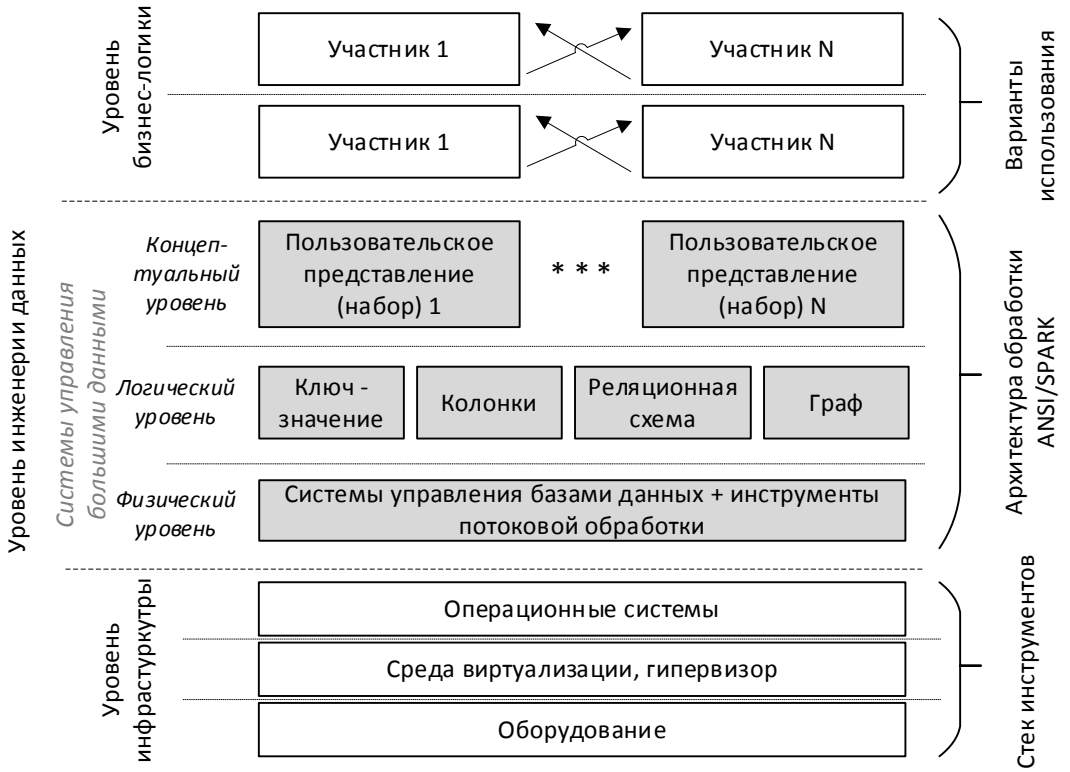

Рис.2. Уровни представления инженерии больших данных в концепции ANSI/SPARC

структуры данных должны быть оценены и согласованы для того, чтобы было возможно управление данными в рамках системы в целом. Такая задача возникла сразу при объединении нескольких систем с разной структуризацией и решается сегодня методами согласования моделей данных [29]. В то же время, на этом уровне, в отличии от СУБД еще рано говорить о едином представлении данных для всего хранилища, так как манипулирование информацией на логическом уровне в каждой модели осуществляется по-разному.

Концептуальным уровнем является сегодня представление больших данных полихранилища и всей системы управления данными для различных пользователей. Причем пользователями в этом случае определяются не только на уровне бизнес-логики, но и, например, ими являются подсистемы обеспечения безопасности верхнего уровня, реализующие политику безопасности на уровне хранилища в целом.

Систематизация объектов различных уровней систем управления большими данными, а также угроз и причин их возникновения приведена в табл. 2. В первую очередь рассматриваются полихранилища, так как именно эта ключевая технология отличает системы управления большими данными от всех других типов распределенных систем.

С появлением полихранилищ ландшафт угроз на уровне инженерии данных практически не изменился [13], как и основные методы защиты [14]. Основной проблемой является согласование реализации политики безопасности между различными инструментами. В чем-то это проблема сродни разработке и реализации единой политики безопасности для разнородных распределенных операционных систем (ОС) и информационных систем [30,31], однако разнородность структуризации данных разных СУБД на логическом уровне в составе одной системы не позволяет напрямую перенести известные практики.

Рассматривая концепцию безопасности систем управления большими данными как многоуровневую архитектуру безопасности, определим функции безопасности каждого уровня.

Физический уровень является технологической основной остальных операций с данными. Основными требованием к нему является безопасность и доверие к среде обработки данных, отсутствие программных закладок, несанкционированного доступа,

Таблица 2

Систематизация объектов различных уровней систем управления большими данными (полихранилищ), угроз безопасности и причин их возникновения

Уровни представления	Физический	Логический	Концептуальный
Объекты	Отдельные СУБД и инструменты обработки	В соответствии с моделью данных	В соответствии с моделью представления
Угрозы	Эксплуатация уязвимостей отдельных инструментов	Ошибки разграничения и контроля доступа, отсутствие аудита и оценки защищенности	Несогласованность политики безопасности верхнего уровня и реализованных ниже политик
Причины возникновения угроз	Аналогичны «классическим» СУБД	Разнородность структуризации и грануляции данных	Динамичность данных, сложный жизненный цикл данных

уязвимостей программных компонентов. Также на этом уровне для защиты от внутреннего нарушителя, которым может быть администратор системы, применима концепция нулевого доверия и безопасности операций. Аудит операций внутри инструментов обработки данных также относится к этому уровню.

Логический уровень определяет грануляцию данных и операции манипулирования с ними на более высоком уровне. Основное требование к безопасности логического уровня это согласованный контроль доступа между инструментами обработки данных и аудит операций с данными между инструментами обработки данных.

Концептуальный уровень определяет безопасность в отношении каждого набора данных, предоставляемых пользователям. С точки зрения защиты информации на этом уровне применимы технологии, использующиеся для технической защиты наборов данных с точки зрения бизнес-логики в системе больших данных в целом. Это защита данных при передаче их на аутсорсинг, в частности, методы анонимизации данных и защита от логического вывода (inferrence attack [32]).

На уровне систем управления большими данными уровни уже являются менее независимыми, чем в более высокоуровневом рассмотрении, и согласованность методов и средств безопасности между ними является ключевым требованием. В данной концепции логический уровень является связующим между представлениями данных для пользователей и физическим манипулированием ими, аналогично уровню инженерии данных в общей многоуровневой концепции больших данных и логическому уровню архитектуры ANSI/SPARC в СУБД. Сегодня основной проблемой, обуславливающей угрозы логического и концептуального уровней в таблице 2, является несогласованность данных в рамках инструментов физического уровня (и их логических моделей), и, как следствие:

- отсутствие единого представления данных в рамках системы в целом (в том числе, для реализации согласованного контроля доступа);
- отсутствие единой системы аудита данных не только внутри инструментов обработки, но и между узлами на которых они функционируют (для распределенных инструментов) а также между инструментами;
- отсутствие единой системы контроля реализации политики безопасности, включая защиту от внутреннего нарушителя.

Решение этих проблем является ключевым шагом для достижения безопасности в классе систем управления большими данными и в рамках безопасности больших данных в боле широкой интерпретации.

Технологический базис безопасности систем управления большими данными

Таким образом, ключевым требованием безопасности систем управления большими данными является консистентное представление на уровне общего монитора безопасности. Для этого необходимо решить целый ряд задач, таких, как построение общей концептуальной модели данных над всеми логическими моделями инструментов обработки данных (по крайней мере, для решения задач безопасности) и обеспечение согласованного аудита операций на физическом и логическом уровне. Решение этих задач позволит подойти к решению и проблемы в целом, включая контроль реализации политики безопасности, оценку защищенности и другие вопросы.

Построение обшей концептуальной модели данных является сложной научной задачей и заслуживает

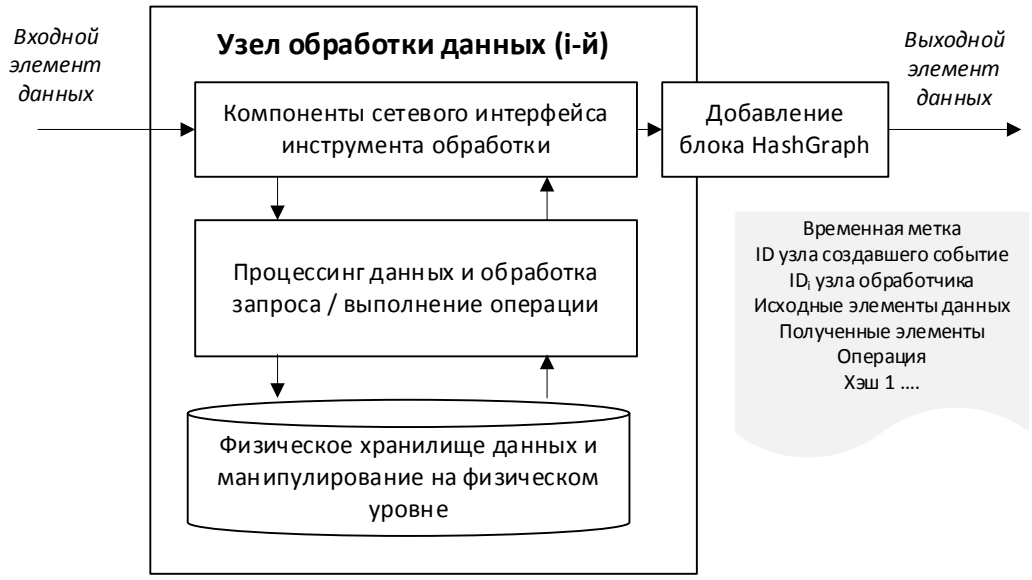

Рис.3. Введение распределенного реестра в систему управления большими данными

отдельного рассмотрения. Задача аудита представляется более простой, так как для большинства отдельных СУБД с точки зрения внутренних операций она уже решена [33], и даже для тех СУБД, в которых нет встроенных модулей аудита, такое решение несложно построить [14]. Основной проблемой остается аудит между инструментами данных и узлами обработки информации. Решение уже этой задачи сегодня может быть построено с использованием технологий распределенного реестра, в частности HashGraph [34]. В таком случае распределенный реестр используется для ведения информации об операциях с данными между узлами (рисунок 3). Под элементом данных понимается не отдельный пакет, а высокоуровневый семантически значимый фрагмент: банковская транзакция, отчет о продажах, запись файла лога и т.д.

В результате анализа цепочек блоков распределенного реестра, полученных в результате анализа прохождения элементов данных в системе, возможен полноценный аудит данных на протяжении цикла жизненного цикла. К тому же, защищенный от подделок со стороны внутреннего нарушителя – администратора системы. За счет высокого уровня абстрагирования от отдельных пакетов на концептуальный уровень семантически значимой информации при таком подходе можно получить достаточно хорошие показатели производительности. Это критически важно, так как именно баланс между производительностью и защищенностью – классическая дилемма систем управления данными (и базами данных).

Пример задержки при внедрении распределенного реестра для системы управления большими данными с достаточно маленькими отдельными элементами, анализирующей трафик для решения задач сетевой безопасности, приведен на рисунке 4. Для систем с семантически значимыми фрагментами большего размера задержка будет еще меньше.

При интеграции этой технологии с существующими журналами и системами аудита отдельных инструментов обработки данных и СУБД формируется комплексная система аудита, позволяющая проследить жизненный цикл каждого фрагмента данных. На основе графов жизненного цикла при наличии общей математической модели описания данных уже может быть согласована политика доступа, настроены параметры разграничения доступа и потом транслированы на уровень инструментов, а также – может проводиться мониторинг, анализ и оценка всей системы в целом.

Заключение

Использование многоуровневого подхода к безопасности сложных современных систем больших данных, как и их компонентов – систем управления большими данными, полихранилищ, позволяет системно взглянуть на задачу обеспечения их защищенности, выделить сходные технологии и методологии в смежных областях для каждого узкого круга задач, выявить проблемные области и искать пути решения проблем.

Сложность обеспечения безопасности больших данных во многом определяется широтой и комплексностью этого понятия. На каждом из трех выделенных в работе основных уровней представления: ин-

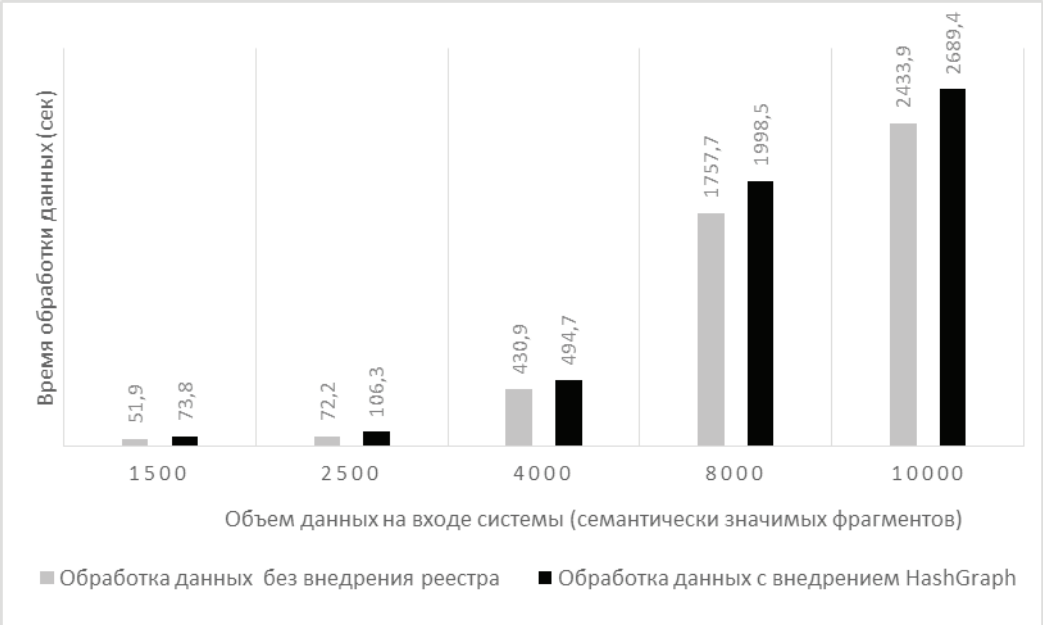

Рис.4. Время обработки данных в системе до введения распределенного реестра и после

фраструктурном, инженерии данных и бизнес-логики не только отличаются технологии, методы и средства работы с большими данными, но и даже само понятие. Каждому уровню присущи свои уязвимости, угрозы и развивающиеся сегодня направления защиты.

Уровень инженерии данных является ключевым технологическим уровнем, отличающим системы управления большими данными от других классов систем. Основными компонентами этого уровня являются разнородные системы управления базами данных, или полихранилища, дополняемые инструментами других классов. Экстраполируя на него принципы построения систем управления данными архитектуры ANSI/SPARC, можно говорить также о трех уровнях: физическом, логическом и концептуальном. Эти уровни коррелируют аналогичными с уровнями традиционных СУБД. Основным отличием является тот факт, что физический уровень включает в себя полностью реализации отдельных инструментов работы с данными.

Ключевые проблемы безопасности систем управления большими данными связаны как раз с разнородностью компонентов полихранилищ, а точнее, с отсутствием единого представления данных на общем логическом уровне, вместо гетерогенных структур данных различных СУБД. С другой стороны, существует проблема организации распределенного аудита на уровне системы управления большими данными в целом, а не, опять же, отдельных инструментов. Сочетание этих двух технологий позволит подойти к задаче разработки и практической реализации средств защиты, таких как компоненты контроля и разграничения доступа, анализа и мониторинга, аудита, оценки защищенности, форензики для систем управления большими данными. Использование технологии распределенного реестра, в частности – HashGraph, позволяет обеспечить эффективный аудит данных между инструментами системы и сделать шаг к решению всей комплексной проблемы.

Исследование выполнено за счет гранта Российского научного фонда № 23-11-20003, https://rscf.ru/project/23-11-20003/, грант Санкт-Петербургского научного фонда (Соглашение №23-11-20003 о предоставлении регионального гранта).

Литература

1. Naeem M. et al. Trends and future perspective challenges in big data //Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, 15–18 October 2019, Arad, Romania, Springer Singapore, 2022. pp. 309–325.
2. Корнев М. С. История понятия “большие данные” (Big Data): словари, научная и деловая периодика // Вестник РГГУ. Серия: Литературоведение. Языкознание. Культурология. 2018. №1 (34). С. 81-85
3. Otto B. The evolution of data spaces // Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Cham: Springer International Publishing. 2022. pp. 3-15.
4. Gupta D., Rani R. A study of big data evolution and research challenges //Journal of information science. 2019. Vol. 45. Is 3. pp. 322-340
5. Антипова К. Г. Способы определения больших данных: Российский и зарубежный опыт // Юридические исследования. 2021. № 9. С. 143–157. doi: 10.25136/2409-7136.2021.9.36591
6. Badidi E., Mahrez Z., Sabir E. Fog computing for smart cities’ big data management and analytics: A review // Future Internet. 2020. Vol. 12. No. 11. pp.1-28. doi: 10.3390/fi12110190
7. Wang J. et al. Big data service architecture: a survey // Journal of Internet Technology. 2020. Vol. 21. No 2. pp. 393-405.
8. Bhattarai B. P. et al. Big data analytics in smart grids: state of the art, challenges, opportunities, and future directions //IET Smart Grid. 2019. Vol. 2. No. 2. pp. 141-154. doi: 10.1049/iet-stg.2018.0261
9. Ndikumana A. et al. Joint communication, computation, caching, and control in big data multi-access edge computing //IEEE Transactions on Mobile Computing. 2019. Vol. 19. No 6. pp. 1359-1374. doi: 10.1109/TMC.2019.2908403.
10. Vogt M. et al. Polystore Systems and DBMSs: Love Marriage Marriage of Convenience? //Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB Workshops, Poly 2021 and DMAH 2021, Virtual Event, August 20, 2021, Revised Selected Papers 7. – Springer International Publishing, 2021. pp. 65-69.
11. Lu J., Holubová I., Cautis B. Multi-model databases and tightly integrated polystores: Current practices, comparisons, and open challenges //Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018. pp. 2301-2302.
12. Gobert M. Design, Manipulation and Evolution of Hybrid Polystores. [электронный ресурс] 2023. https://pure.unamur.be/ws/portalfiles/portal/74437131/2023_GobertM_these.pdf (дата доступа 01.08.2023)
13. Poltavtseva, M. A. Evolution of Data Management Systems and Their Security // Proceedings - 2019 International Conference on Engineering Technologies and Computer Science: Innovation and Application, EnT 2019, Moscow, 26–27 March 2019. – Moscow, 2019. pp. 25-29. doi: 10.1109/EnT.2019.00010. EDN AGZBGD.
14. Полтавцева, М. А. Безопасность баз данных / Санкт-Петербург: Федеральное государственное автономное образовательное учреждение высшего образования “Санкт-Петербургский политехнический университет Петра Великого”, 2023. 143 с. EDN RICQEN.
15. Титаренко, Д. В., Исмайлов Э. И. Безопасность больших данных // Проблемы информационной безопасности социально-экономических систем: VII Всероссийская с международным участием научно-практическая конференция, Гурзуф, 18–20 февраля 2021 года. – Симферополь: Крымский федеральный университет им. В. И. Вернадского, 2021. С. 121–122. EDN NFQQHT.
16. Скворцов Н. Константинов А., Кузнецов С. Ценность ваших данных / Москва: Альпина ПРО, 2022. – 750 с.
17. Ogbuke N. J. et al. Big data supply chain analytics: ethical, privacy and security challenges posed to business, industries and society // Production Planning & Control. 2022. vol. 33. №. 2-3. С. 123–137. doi: 10.1080/09537287.2020.1810764
18. Binjubeir M. et al. Comprehensive survey on big data privacy protection //IEEE Access. 2019. vol. 8. pp. 20067-20079. doi: 10.1109/ACCESS.2019.2962368
19. Madan S., Bhardwaj K., Gupta S. Critical analysis of big data privacy preservation techniques and challenges //International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Vol. 3. – Springer Singapore, 2022. – pp. 267-278. doi: 10.1007/978-981-16-3071-2_23
20. Mehta B. B., Rao U. P. Improved l-diversity: scalable anonymization approach for privacy preserving big data publishing //Journal of King Saud University-Computer and Information Sciences. 2022. vol. 34. Is. 4. pp. 1423-1430. doi: 10.1016/j.jksuci.2019.08.006
21. Dhiman G. et al. Federated learning approach to protect healthcare data over big data scenario //Sustainability. 2022. vol. 14. Is 5. pp. 1-14. doi: 10.3390/su14052500
22. Статьев В. Ю., Докучаев В. А., Маклачкова В. В. Информационная безопасность на пространстве” Больших данных” //T-Comm-Телекоммуникации и Транспорт. 2022. Т. 16. №. 4. С. 21-28.
23. Poudel M. et al. Development of a polystore data management system for an evolving big scientific data archive //Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB 2019 Workshops, Poly and DMAH, Los Angeles, CA, USA, August 30, 2019, Revised Selected Papers 5. – Springer International Publishing, 2019. pp. 167-182. doi: 10.1007/978-3-030-33752-0_12
24. Ecy M.T, Вальдуриес П. Принципы организации распределенных бах данных. – М.: ДМК Пресс. – 2021. 678с.
25. Fong J. S. P. et al. Heterogeneous Database Connectivity //Information Systems Reengineering, Integration and Normalization: Heterogeneous Database Connectivity. 2021. pp. 317-367. doi: 10.1007/978-3-030-79584-9_9
26. Abdennebi A. et al. Machine learning based load distribution and balancing in heterogeneous database management systems // Concurrency and Computation: Practice and Experience. 2022. vol. 34. Is 4. pp. 1-13. doi: 10.1002/cpe.6641
27. Kim T. et al. Similarity query support in big data management systems //Information Systems. 2020. vol. 88.pp. 1-61. doi: 10.1016/j.is.2019.101455
28. van Gils B. Data Storage and Operations //Data in Context: Models as Enablers for Managing and Using Data. – Cham: Springer Nature Switzerland, 2023. pp. 105-114.
29. Dziedzic A., Elmore A. J., Stonebraker M. Data transformation and migration in polystores //2016 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, 2016. pp. 1-6. doi: 10.1109/HPEC.2016.7761594
30. Зегжда, Д. П. Особенности обеспечения информационной безопасности вычислительных систем // Безопасность информационных технологий. 2021. Т. 28, № 1. С. 42–61. doi: 10.26583/bit.2021.1.04. EDN ETQPVN
31. Белим С. В., Белим С. Ю. Проблемы построения политики безопасности при объединении информационных систем //Математические структуры и моделирование. 2018. №. 3 (47). С. 126–131.

32. Полтавцев А. А., Хабаров А. Р., Селянкин А. О. Атаки логического вывода и защита информации в базах данных // Проблемы информационной безопасности. Компьютерные системы. 2019. № 4. pp. 20–25. EDN NTRSDO
33. Полтавцева М. А. и др. Модели форензики и расследование инцидентов в СУБД // Защита информации. Инсайд. 2021. № 3(99). С. 18–23. EDN OMKIRU.
34. Полтавцева М. А., Торгов В. А. Применение технологий распределенного реестра для аудита и расследования инцидентов в системах обработки больших данных // Проблемы информационной безопасности. Компьютерные системы. 2021. № 4. С. 144–156. doi: 10.48612/jisp/r3x6-ea4a-aaxn. EDN YRRVXG.

MULTI-LEVEL SECURITY CONCEPT FOR BIG DATA MANAGEMENT SYSTEMS

Poltavtseva M.A.[6], Zegzhda D.P.[7], Kalinin M. O.[8]

The purpose *of the study. Big data management technologies and systems are the basis for a huge number of modern digital services. On the one hand, they are built on traditional solutions, and on the other hand, they incorporate new approaches such as polystores or data outsourcing. The key role in the technology stack of the digital economy and novelty determine both the attractiveness of such assets for an attacker and the imperfection of protection methods. The aim of the paper is to analyze big data as an object of protection and to develop a multilevel concept of their security based on the consistency approach.*

Methods *of the study. The paper uses a layered approach, which also corresponds to the ANSI/SPARC architecture of database management systems. Big data is considered at three levels from infrastructure to business logic, key technologies, vulnerabilities and protection methods are highlighted. The ANSI/SPARC also defines in more detail the level architecture of big data management systems based on polystores and analyzes their security. The technological basis for the security of big data management systems as a system of distributed dynamic auditing is defined, an example of such a system based on a distributed registry is given.*

Results *of the study. The article identifies three levels of big data consideration: infrastructure, data engineering and business logic. The authors formulate the evolutionary changes of big data systems in comparison with traditional DBMS from the point of view of information security. The concept of big data management system is given, its own architectural levels based on ANIS/SPARK architecture are defined, for each of them security problems, reasons for their appearance and directions of protection development means are highlighted. The authors highlighted the key security requirement of big data management systems - consistent representation at the level of a global security monitor. For its implementation, in terms of collecting data about the system, the use of distributed dynamic ledger technologies is proposed. The system of distributed dynamic auditing for big data management based on HashGraph technology has been tested.*

Scientific novelty. *The paper is the first to formulate a multilevel security concept for big data management systems, within the framework of which the key vulnerabilities of big data systems, different from other classes of systems and traditional DBMSs, are identified and systematized at different levels. For the first time the application of distributed ledger technology for collecting data on the life cycle of information in the big data management system was proposed. The conducted research allows for a more comprehensive approach to ensuring the security of big data and big data management systems, specifies and coordinates the sets of protection methods and means, and lays the foundation for the construction of such systems in a secure design.*

6 Maria A. Poltavtseva, Dr.Sc., Associate Professor, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia. E-mail: poltavtseva@ibks.spbstu.ru

7 Dmitri P. Zegzhda, Corresponding member of RAS, Dr.Sc., Professor, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia. E-mail: dmitry@ibks.spbstu.ru

8 Maxim O. Kalinin, Dr.Sc., Professor, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia. E-mail: max@ibks.spbstu.ru

Keywords: *information security, big data security, consistency approach, security architecture, security model, polystore security, distributed registry.*

The study was supported by the grant of Russian Science Foundation No.23-11-20003, https://rscf.ru/project/23-11-20003/; grant of St.Petersburg Science Foundation (Agreement No.23-11-20003 on the regional grant).

References

1. Naeem M. et al. Trends and future perspective challenges in big data //Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, 15–18 October 2019, Arad, Romania, Springer Singapore, 2022. pp. 309-325.
2. Kornev M. S. Istorija ponjatija “bol’shie dannye” (Big Data): slovari, nauchnaja i delovaja periodika // Vestnik RGGU. Serija: Literaturovedenie. Jazykoznanie. Kul’turologija. 2018. №1 (34). pp. 81-85
3. Otto B. The evolution of data spaces // Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Cham: Springer International Publishing. 2022. pp. 3-15.
4. Gupta D., Rani R. A study of big data evolution and research challenges //Journal of information science. 2019. Vol. 45. Is 3. pp. 322-340
5. Antipova K.G. Sposoby opredelenija bol’shih dannyh: Rossijskij i zarubezhnyj opyt // Juridicheskie issledovanija. 2021. № 9. pp. 143 - 157. doi: 10.25136/2409-7136.2021.9.36591
6. Badidi E., Mahrez Z., Sabir E. Fog computing for smart cities’ big data management and analytics: A review // Future Internet. 2020. Vol. 12. No. 11. pp.1-28. doi: 10.3390/fi12110190.
7. Wang J. et al. Big data service architecture: a survey // Journal of Internet Technology. 2020. Vol. 21. No 2. pp. 393-405.
8. Bhattarai B. P. et al. Big data analytics in smart grids: state of the art, challenges, opportunities, and future directions //IET Smart Grid. 2019. Vol. 2. No. 2. pp. 141-154. doi: 10.1049/iet-stg.2018.0261.
9. Ndikumana A. et al. Joint communication, computation, caching, and control in big data multi-access edge computing //IEEE Transactions on Mobile Computing. 2019. Vol. 19. No 6. pp. 1359-1374. doi: 10.1109/TMC.2019.2908403
10. Vogt M. et al. Polystore Systems and DBMSs: Love Marriage or Marriage of Convenience? //Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB Workshops, Poly 2021 and DMAH 2021, Virtual Event, August 20, 2021, Revised Selected Papers 7. – Springer International Publishing, 2021. pp. 65-69.
11. Lu J., Holubová I., Cautis B. Multi-model databases and tightly integrated polystores: Current practices, comparisons, and open challenges //Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018. pp. 2301-2302.
12. Gobert M. Design, Manipulation and Evolution of Hybrid Polystores. 2023. https://pure.unamur.be/ws/portalfiles/portal/74437131/2023_GobertM_these.pdf (access date 01.08.2023)
13. Poltavtseva, M. A. Evolution of Data Management Systems and Their Security // Proceedings - 2019 International Conference on Engineering Technologies and Computer Science: Innovation and Application, EnT 2019, Moscow, 26–27 march 2019. – Moscow, 2019. pp. 25-29. doi: 10.1109/EnT.2019.00010. EDN AGZBGD.
14. Poltavceva, M. A. Bezopasnost’ baz dannyh / Sankt-Peterburg : Federal’noe gosudarstvennoe avtonomnoe obrazovatel’noe uchrezhdenie vysshego obrazovanija “Sankt-Peterburgskij politehnicheskij universitet Petra Velikogo”, 2023. 143 p. EDN RICQEN.
15. Titarenko, D. V., Ismajlov Je. I. Bezopasnost’ bol’shih dannyh // Problemy informacionnoj bezopasnosti social’no-jekonomicheskih sistem : VII Vserossijskaja s mezhdunarodnym uchastiem nauchno-prakticheskaja konferencija, Gurzuf, 18–20 fevralja 2021 goda. – Simferopol’: Krymskij federal’nyj universitet im. V.I. Vernadskogo, 2021. pp. 121-122. EDN NFQQHT.
16. Skvorcov N. Konstantinov A., Kuznecov S. Cennost’ vashih dannyh / Moskva: Al’pina PRO, 2022. – 750 s.
17. Ogbuke N. J. et al. Big data supply chain analytics: ethical, privacy and security challenges posed to business, industries and society // Production Planning & Control. 2022. vol. 33. №. 2-3. pp. 123-137. doi: 10.1080/09537287.2020.1810764
18. Binjubeir M. et al. Comprehensive survey on big data privacy protection //IEEE Access. 2019. vol. 8. pp. 20067-20079. doi: 10.1109/ACCESS.2019.2962368
19. Madan S., Bhardwaj K., Gupta S. Critical analysis of big data privacy preservation techniques and challenges //International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Vol. 3. – Springer Singapore, 2022. – pp. 267-278. doi: 10.1007/978-981-16-3071-2_23
20. Mehta B. B., Rao U. P. Improved l-diversity: scalable anonymization approach for privacy preserving big data publishing //Journal of King Saud University-Computer and Information Sciences. 2022. vol. 34. Is. 4. pp. 1423-1430. doi: 10.1016/j.jksuci.2019.08.006
21. Dhiman G. et al. Federated learning approach to protect healthcare data over big data scenario //Sustainability. 2022. vol. 14. Is 5. pp. 1-14. doi: 10.3390/su14052500
22. Stat’ev V. Ju., Dokuchaev V. A., Maklachkova V. V. Informacionnaja bezopasnost’ na prostranstve” Bol’shih dannyh” //T-Comm-Telekommunikacii i Transport. 2022. vol. 16. №. 4. pp. 21-28.
23. Poudel M. et al. Development of a polystore data management system for an evolving big scientific data archive //Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB 2019 Workshops, Poly and DMAH, Los Angeles, CA, USA, August 30, 2019, Revised Selected Papers 5. – Springer International Publishing, 2019. pp. 167-182. doi: 10.1007/978-3-030-33752-0_12
24. Esu M.T, Val’duries P. Principy organizacii raspredelennyh bah dannyh. – M.: DMK Press. – 2021. 678p.
25. Fong J. S. P. et al. Heterogeneous Database Connectivity //Information Systems Reengineering, Integration and Normalization: Heterogeneous Database Connectivity. 2021. pp. 317-367. doi: 10.1007/978-3-030-79584-9_9

26. Abdennebi A. et al. Machine learning based load distribution and balancing in heterogeneous database management systems // Concurrency and Computation: Practice and Experience. 2022. vol. 34. Is 4. pp. 1-13. doi: 10.1002/cpe.6641
27. Kim T. et al. Similarity query support in big data management systems //Information Systems. 2020. vol. 88.pp. 1-61. doi: 10.1016/j.is.2019.101455
28. van Gils B. Data Storage and Operations //Data in Context: Models as Enablers for Managing and Using Data. Cham : Springer Nature Switzerland, 2023. pp. 105-114.
29. Dziedzic A., Elmore A. J., Stonebraker M. Data transformation and migration in polystores //2016 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, 2016. pp. 1-6. doi: 10.1109/HPEC.2016.7761594
30. Zegzhda, D. P. Osobennosti obespecheniya informacionnoj bezopasnosti vy`chislitel`ny`x sistem // Bezopasnost` informacionny`x texnologij. 2021. Vol. 28, № 1. pp. 42-61. doi: 10.26583/bit.2021.1.04. EDN ETQPVN.
31. Belim S. V., Belim S. Ju. Problemy postroenija politiki bezopasnosti pri ob#edinenii informacionnyh sistem //Matematicheskie struktury i modelirovanie. 2018. №. 3 (47). pp. 126-131.
32. Poltavcev A. A., Habarov A. R., Seljankin A. O. Ataki logicheskogo vyvoda i zashhita informacii v bazah dannyh // Problemy informacionnoj bezopasnosti. Komp'juternye sistemy. 2019. № 4. pp. 20-25. EDN NTRSDO
33. Poltavceva M. A. i dr. Modeli forenziki i rassledovanie incidentov v SUBD // Zashhita informacii. Insajd. 2021. № 3(99). pp. 18-23. EDN OMKIRU.
34. Poltavceva M. A., Torgov V. A. Primenenie tehnologij raspredelennogo reestra dlja audita i rassledovanija incidentov v sistemah obrabotki bol'shih dannyh // Problemy informacionnoj bezopasnosti. Komp'juternye sistemy. 2021. № 4. pp. 144-156. doi: 10.48612/jisp/r3x6-ea4a-aaxn. EDN YRRVXG.

ПРОБЛЕМА МАСКИРОВАНИЯ И ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ В КИБЕРПРОСТРАНСТВЕ

Горбачев А.А.[1], Максимов Р.В.[2]

Цель исследования: *определение перспективных направлений научных исследований в области маскирования объектов киберпространства в контексте развития технологий машинного обучения.*

Используемые методы: *общая теория управления и моделирования, математическая статистика, общенаучные методы анализа и синтеза.*

Результат исследования: *определена научная проблема маскирования объектов киберпространства и применения технологий машинного обучения в условиях информационно-технических воздействий злоумышленников. Совершенствование методов маскирования на уровне сетевых узлов, локальных сегментов и информационных направлений с применением методов генеративного и состязательного машинного обучения позволит повысить защищенность объектов киберпространства за счет снижения эффективности сетевой разведки злоумышленников, основанной на методах и алгоритмах машинного обучения. Требуют глубокой теоретической и экспериментальной проработки вопросы: оценки формы, содержания, информативности, предварительной обработки и генерации «цифровых отпечатков» ложных и истинных информационных объектов, выбор типов и оптимальной архитектуры алгоритмов глубокого обучения, оценки качества маскирования как атак типа «уклонение» и «отравление» на алгоритмы машинного обучения потенциальных злоумышленников.*

Научная новизна: *заключается в рассмотрении концепции маскирования объектов киберпространства в условиях информационно-технического воздействия злоумышленников с позиции общей теории управления, моделирования и применения технологий машинного обучения.*

Ключевые слова: *моделирование, теория управления, проактивная парадигма защиты, сетевая разведка, машинное обучение.*

DOI:10.21681/4311-3456-2023-5-37-49

Введение

В последние десятилетия наблюдается процесс интеграции технических систем различного назначения (вычислительных и телекоммуникационных сетей, систем связи и автоматизированного управления) в единую систему, реализующую обмен информацией между гетерогенными элементами с целью решения широкого спектра задач. Это создает условия формирования единого информационного пространства или *киберпространства*, включающего в свою структуру такие понятия, как: Интернет, Интернет вещей, телекоммуникационные и вычислительные сети, процессоры, контроллеры [1]. Некоторые авторы киберпространство представляют в виде трех уровней: психо-когнитивного, программного обеспечения и приложений, а также уровня аппаратного обеспечения [2]. Киберпространство позволяет повысить доступность информационных ресурсов, оперативность информационного обмена и принятия решений, так как научно-технический прогресс приводит к росту информационных потребностей человечества и повышению требований к качеству услуг связи и автоматизированных систем управления.

В современных условиях имеет место *интеллектуализация* сфер деятельности человека, которая заключается в интенсивном развитии и внедрении систем искусственного интеллекта в связи с созданием эффективных архитектур вычислительных устройств, ростом их производительности и доступности, нако-

1 Горбачев Александр Александрович, кандидат технических наук, преподаватель Краснодарского высшего военного орденов Жукова и Октябрьской Революции Краснознаменного училища имени генерала армии С.М. Штеменко, г. Краснодар, Россия. E-mail: infosec23.00@ mail.ru

2 Максимов Роман Викторович, доктор технических наук, профессор, профессор Краснодарского высшего военного орденов Жукова и Октябрьской Революции Краснознаменного училища имени генерала армии С.М. Штеменко, г. Краснодар, Россия. E-mail: rvmaxim@yandex.ru

плением колоссальных объемов данных и развитием технологий их хранения.

Развитие технологий искусственного интеллекта и методов машинного обучения с одной стороны является предпосылкой, а с другой стороны – следствием изменения подходов к моделированию объектов окружающего мира вообще (рис. 1) [3, 4]. *Модель* как гомоморфное отображение существенных свойств объекта предназначена для *познания* и *управления*. Моделирование из общенаучного метода познания закономерностей окружающего мира превращается в инструмент управления для его эффективного преобразования. Рост вычислительной мощности аппаратной базы позволил развить алгоритмические и имитационные методы моделирования при исследовании объектов и процессов, имеющих принципиально стохастическую природу, так как зачастую отсутствует возможность или целесообразность проведения натурных экспериментов и построения аналитических зависимостей. С другой стороны, технологическое развитие создает потребность в оперативном и автоматизированном построении относительно точных моделей, использующих накопленные статистические наблюдения, которые имеют прикладное значение в системах поддержки принятия решений, системах управления технологическими процессами, робототехнике, системах распознавания образов, обработки естественного языка и других областях. В широком смысле технологии искусственного интеллекта предназначены для автоматизированного синтеза моделей процессов и объектов, то есть для универсальной нелинейной аппроксимации зависимостей между некоторыми входными наблюдениями и выходными количественными или категориальными переменными с целью управления этими объектами или процессами.

Рис. 1. Общие тенденции в теории и методологии математического моделирования

Отмеченные предпосылки привели к смещению методологии моделирования от создания познавательных моделей «механизмов» или «белых ящиков», имеющих детальную внутреннюю структуру, раскрывающую сущность явления к описательным моделям типа «черный ящик» или «статистическая фотография», име-

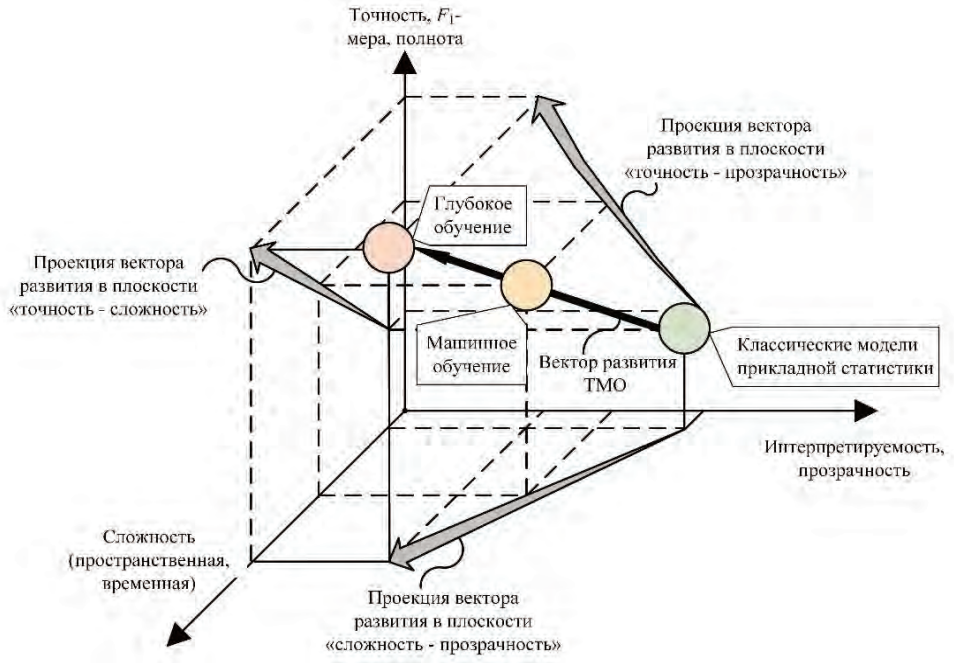

Рис. 2. Тенденции развития технологий машинного обучения и моделирования

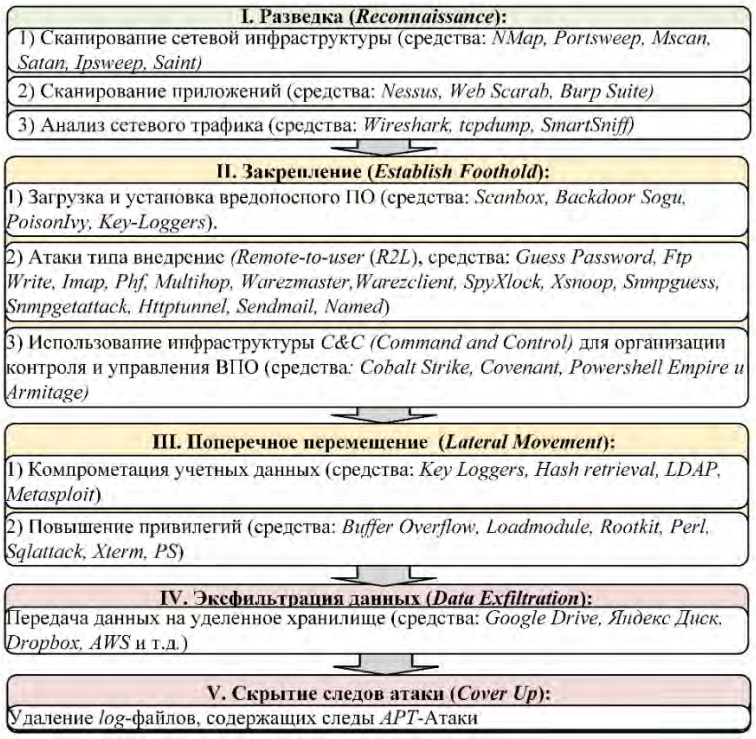

Рис. 3. Вариант реализации APT-атаки на объекты киберпространства

ющим высокую точность, обобщающую способность относительно обучающей выборки, но не раскрывающим сущности описываемых процессов (рис. 2).

Повышение качества статистической аппроксимации моделями глубокого машинного обучения осуществляется за счет их способности обрабатывать большие объемы неструктурированных данных, отсутствия теоретической интерпретации архитектуры модели, то есть ее инвариантности относительно природы объекта. Иными словами, качество аппроксимации наибольшим образом зависит не от архитектуры алгоритма (модели), а от качества обучающих и тестовых наборов данных. Так смещенность, недостаточная представительность обучающих и тестовых выборок могут привести к тому, что доверие к системе искусственного интеллекта не будет обеспечено[3]. В свою очередь наличие подобных уязвимостей процесса идентификации параметров (обучения) моделей инициирует исследование злоумышленниками методов атак на алгоритмы машинного обучения (отравление, уклонение, оракул), которые используются в системах безопасности (обнаружение вредоносного программного обеспечения, системы аутентификации) [5, 6].

3 ГОСТ Р 59276-2020. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения. Москва: Федеральное агентство по техническому регулированию и метрологии, 2020. 25 с.

Возможности злоумышленников и основные парадигмы защиты

Оборотной стороной указанных условий и факторов является использование технологий машинного обучения для реализации методов информационно-технического воздействия на элементы информационного пространства в форме компьютерных атак и морально-психологического воздействия на определенные группы лиц с целью достижения экономических, политических и военных целей. В статье рассматривается составляющая киберпространства, представленная информационно-телекоммуникационными и вычислительными сетями, а также информационно-технические воздействия в форме компьютерных атак. В мировой практике принято детализировать процесс реализации целевых компьютерных атак (*APT*-атак) на последовательность этапов (рис. 3) [7].

Ключевым этапом, на котором сфокусировано внимание в данной работе, является проведение злоумышленниками *разведки* (сетевой или компьютерной), обеспечивающей достижение целей информационно-технического воздействия на узлы (сетевые информационные объекты) вычислительных сетей за счет определения свойств программного и аппаратного обеспечения, то есть посредством *моделирования* объектов.

	Базовые парадигмы защиты		
	Пассивная	*Реактивная*	*Проактивная*
Характеристика	Физическое или логическое *дистанцирование* со злоумышленником. *Методы*: 1. Создание физически и логически обособленной инфраструктуры. 2.Создание и организационно-техническое обеспечение контролируемой зоны. 3. Логическое разграничение доступа к ресурсам (фильтрация).	Физическое или логическое *дистанцирование* со злоумышленником *при обнаружении* признаков компьютерных атак. *Методы*: 1. Обнаружение вредоносного программного обеспечения. 2. Обнаружение аномалий сетевого трафика, поведения пользователей, содержания файлов и т.д.	*Управление* структурно-функциональными характеристиками объектов вычислительной сети. *Методы*: 1. Защита с использованием подвижных целей (динамичность, многообразие, избыточность). 2. Маскирование информационного обмена и структуры (имитация, мимикрия); 3. Стеганография и шифрование.
Недостатки	1. Сложность обеспечения растущих потребностей и требований к качеству услуг связи и АСУ *между обособленными системами*; 2. Высокие капиталовложения на создание и поддержание инфраструктуры.	1. Реактивный (запаздывающий) характер по отношению к воздействию. 2. Вычислительная ресурсоемкость технических решений. 3. Низкая эффективность относительно угроз 0-дня.	1. Вычислительная ресурсоемкость технических решений. 2. Отрицательное влияние на качество услуг связи. 3. Отсутствие абсолютной защиты без использования пассивных или реактивных средств. 4. Стойкость к вскрытию, затраты на распределение ключей.

Рис. 4. Характеристика базовых парадигм защиты

Если рассматривать меры защиты объектов киберпространства по признаку *активности* по отношению к потенциальным воздействиям злоумышленника, то они реализуются с позиции трех *базовых парадигм защиты*: *пассивной*, *реактивной* и *проактивной* (рис. 4).

Для качественной оценки парадигм защиты рассмотрим процесс информационно-технического воздействия с точки зрения общей теории управления[4] как взаимодействие объекта воздействия (атаки), окружающей среды и системы управления информационно-технического воздействия злоумышленника или субъекта атаки (рис. 5). Окружающая среда (информационно-телекоммуникационная сеть, сопряженная с объектом) воздействует на объект посредством компоненты X (сетевой трафик, генерируемый средой). Объект в свою очередь осуществляет воздействие на среду посредством компоненты Y (сетевой трафик, генерируемый объектом). При этом объект может представлять собой как отдельный элемент или узел вычислительной сети, так и целую подсеть (сегмент, систему) узлов, на которые возложены взаимосвязанные цели функционирования Y_o и защиты Y_d, которые достигаются посредством выполнения алгоритмов функционирования и защиты объекта из множеств алгоритмов Ω_o и Ω_d соответственно.

Цель злоумышленника на этапе разведки состоит в построении *адекватной модели* F_a объекта атаки с использованием средств и алгоритмов разведки Ω_r. Иными словами, целью сетевой разведки является *идентификация* (структурная и параметрическая) модели объекта в широком смысле:

$$\left|Y' - F_a(X', U_r, U_a)\right| \to \min_{F_a \in Q_a, U_r \in \Omega_r, U_a \in \Omega_a}, \qquad (1)$$

где, Q_a – множество модельных операторов и их параметров, находящихся в распоряжении у субъекта атаки; F_a – искомый модельный оператор, связывающий свойства среды и объекта (X' и Y') с воздействиями средства разведки U_r и средствами информационно-технического воздействия U_a.

Построенная модель объекта в форме модельного оператора $F_a(X', U_r, U_a)$ используется в блоке управления для определения оптимального управляющего воздействия субъекта U_a^* из множества алгоритмов (способов) информационно-технического воздействия Ω_a на объект с целью достижения требуемого состояния системы с выходом Y_a.

Моделирование объекта в ходе разведки осуществляется на основании измерений X' и Y', производимых с использованием средств, выполняющих функцию датчиков D_1 и D_2. Датчик D_1 (канальная среда) предназначен для пассивного анализа среды (*пассивная разведка*: анализ сетевого трафика), взаимодействующей с объектом исследования (подконтрольный злоумышленнику элемент, выполняю-

4 Растригин Л.А., Марков В.А. Кибернетические модели познания. Вопросы методологии / Издательство «Зинатне». Рига. 1976 г. 264 с.

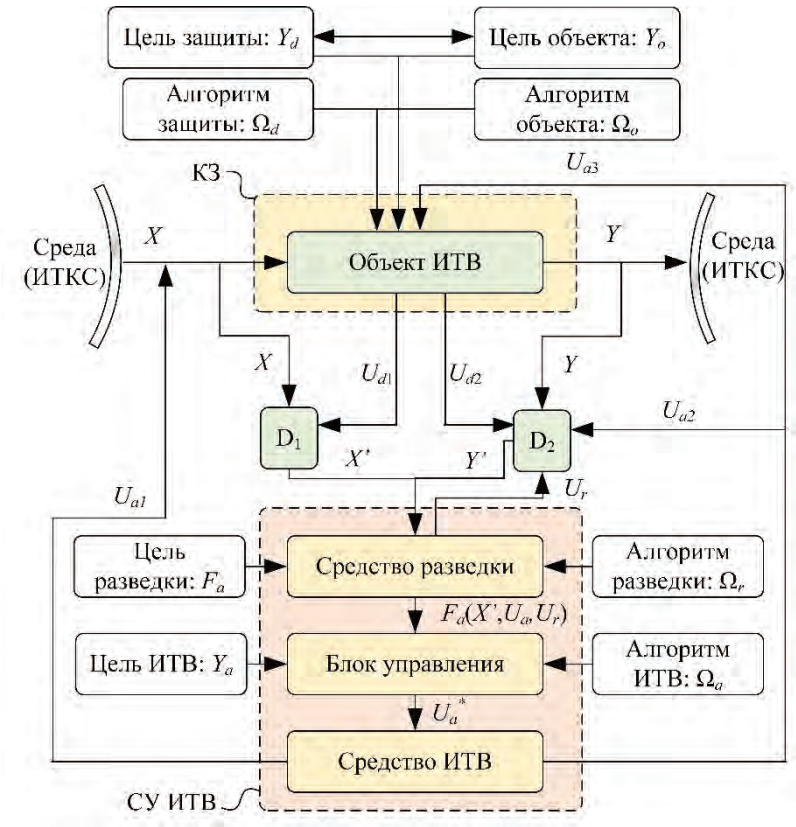

Рис. 5. Общая схема реализации процесса сетевой разведки, информационно-технического воздействия (ИТВ) и защиты в вычислительных сетях (КЗ – контролируемая зона, ИТКС – информационно-телекоммуникационная сеть)

щий функцию захвата сетевого трафика). Датчик D_2 (физический или логический интерфейс взаимодействия объекта с окружающей средой) позволяет осуществить измерение показателей свойств объекта при непосредственном взаимодействии средства разведки и целевого узла системы (*активная разведка*: сканирование).

На основании построенной модели, субъект воздействия обладает возможностью применения способов информационно-технического воздействия опосредованно U_{a1} (атаки типа «отказ в обслуживании», распределенные атаки), непосредственно взаимодействуя с объектом через сетевые интерфейсы U_{a2} (атаки типа: подбор пароля, эксплуатация уязвимостей), а также с использованием способов воздействия U_{a3} в пределах контролируемой зоны (внедрение вредоносного кода с использованием методов социальной инженерии, использование недекларированных возможностей программного и аппаратного обеспечения).

Изменение компонент X' и Y' позволяет системе защиты объекта противодействовать процессу информационного-технического воздействия.

Для построения адекватной модели объекта необходимо соблюдение общих принципов (постулатов) моделирования:

- *наблюдаемость*, которая заключается в возможности измерения интересующих показателей свойств (инвариантов) объекта исследования (вне контролируемой зоны наблюдаемость обеспечивается датчиками D_1 и D_2);
- *стабильность*, которая заключается в стационарности (статистической устойчивости) в узком смысле интересующих показателей свойств (инвариантов) объекта исследования во времени;
- *экстраполируемость*, которая заключается в возможности использования синтезированной модели объекта исследования для последующего управления в иных условиях (зависит от обобщающей способности модельного оператора F_a и стационарности инвариантов объекта исследования);
- *конечность*, которая заключается в конечномерности параметров, подлежащих оценке, в соответствии с требованиями аппаратного и программного обеспечения к сложности алго-

ритма моделирования (конечность определяется видом модельного оператора F_a, многообразием измеряемых характеристик объекта исследования через датчики D_1, D_2, временными и вычислительными ресурсами субъекта моделирования);

— *согласованность* объекта и субъекта, характеризующая наличие понятийного аппарата для качественной интерпретации результатов моделирования;

— *измеримость* указывает на существование системы мер, посредством которой производится измерение показателей свойств объекта исследования.

Соблюдение последних трех принципов обусловлено научно-техническим обеспечением злоумышленника, его вычислительными, временными ресурсами, ассортиментом моделей и методов, аппроксимирующих соответствующие отображения.

Меры защиты объекта, принятые в рамках базовых парадигм защиты, влияют на первые четыре принципа моделирования свойств объекта средством сетевой разведки. Для *парадигмы пассивной защиты* (физического или логического дистанцирования с потенциальным злоумышленником) свойственно редукция компонент X, Y, X' и Y' в связи с полной изоляцией объекта от внешней среды. Пассивная защита не позволяет средству разведки обеспечить принцип наблюдаемости при идентификации модели объекта, так как вне контролируемой зоны отсутствует возможность снятия показаний с датчиков D_1 и D_2. При этом, с одной стороны, не обеспечивается цель объекта по поддержанию заданного уровня показателей качества функционирования Y_0 (в смысле обеспечения обмена информацией с узлами распределенной информационной системы за пределами контролируемой зоны с требуемым качеством), с другой стороны, единственным способом реализации потенциала средств информационно-технического воздействия остается применение способов воздействия $U_{аз}$ в пределах контролируемой зоны, то есть поиск субъектом возможностей для реализации внутренних угроз.

Парадигма реактивной защиты, основанная на принципах реагирования на инциденты безопасности, не накладывает ограничений на компоненты X, Y, X' и Y' в связи с тем, что существует возможность информационного обмена со средой (удаленными узлами), соответственно все основные принципы моделирования (наблюдаемости, стабильности и экстраполируемости) соблюдаются. И лишь при обнаружении признаков атак (вторжений) осуществляется физическое (логическое) отключение защищаемого объекта от среды или блокирование действий субъекта. Принципиальной уязвимостью данного подхода является запаздывающий характер реакции на попытки воздействия злоумышленником, что связано как с принципами настройки средств защиты, так и их функционирования. Основная задача злоумышленника при реализации компьютерных атак сводится к таким воздействиям, которые классифицируются средствами защиты как легитимные (ошибка II рода). В случае эксфильтрации данных злоумышленником может пройти длительный промежуток времени (от нескольких часов до нескольких месяцев) прежде, чем будет обнаружен факт вторжения. Также существует принципиальная возможность блокирования легитимных взаимодействий с объектом (ошибка I рода). Техническая реализация реактивного подхода представлена широким классом средств защиты: системами обнаружения вторжений или атак (СОВ, СОА), средствами антивирусной защиты, фильтрации и анализа данных (трафика), обнаружения аномалий поведения. Несмотря на принципиальные недостатки, разработка и применение средств и способов защиты в рассмотренных подходах являются необходимыми и перспективными направлениями в области информационной безопасности.

Парадигма проактивной защиты предполагает постоянное влияние на компоненты X, Y, X' и Y' без необходимости физического и логического дистанцирования с системами потенциального злоумышленника. В пределах данной парадигмы различают основные подходы – это *защита с использованием подвижной цели*, *маскирование*, *стеганография* и *шифрование*.

Защита с использованием подвижной цели (*moving target defense*) это подход, который использует динамическое изменение структурно-функциональных характеристик защищаемых узлов вычислительной сети. Принципиально данный подход основан на *динамичности* (*shuffling* – изменение параметров объекта во времени), *многообразии* (*diversity* – генерация многообразия возможных значений параметров объекта) и *избыточности* (*redundancy* – синтез дополнительных объектов). Динамичность свойств объекта позволяет нарушить соблюдение принципа стабильности и экстраполируемости при определении модельного оператора объекта атаки F_a злоумышленником, а повышение многообразия и избыточности влияют на конечность измеряемых характеристик посредством изменения структуры и мощности мно-

Рис. 6. Характеристика способов реализации проактивной парадигмы защиты в рамках концепций защиты с использованием подвижной цели и маскирования

жества способов воздействия Ω_a. Данный подход предполагает принципиальную недостижимость абсолютной защищенности, то есть такого состояния объекта защиты, в котором реализация угрозы безопасности информации является невозможным событием [8]. Недостижимость абсолютной защиты связана с тем, что воздействия злоумышленников носят принципиально неопределенный, случайный характер и невозможно достоверно предсказать характеристики этих воздействий (за исключением вырожденного случая, в котором объект перестает выполнять свои функции по предназначению).

Маскирование (*маскировка*), аналогом которого в зарубежной литературе является *киберобман* (*Cyber Deception*) [9], это подход, направленный на создание ложного представления о свойствах объекта атаки методами *мимикрии* и *имитации*. Маскирование и защита с использованием подвижной цели используют общий фундаментальный принцип: динамический характер свойств объекта защиты. Особенностью маскирования является целевой характер изменения характеристик объекта, направленный на создание у злоумышленника *заданного ложного представления* об объекте атаки. В терминах общей теории управления оно направлено на идентификацию злоумышленником *заданного* ложного модельного оператора F'_a и соответственно синтез и применение неоптимального способа информационно-технического воздействия из *заданного* подмножества Ω'_a множества алгоритмов Ω_a.

Одной из проблем при реализации маскирования является неопределенность, связанная с выбором того свойства, характеристики или набора свойств объекта, по которому потенциальный злоумышленник классифицирует реальную или ложную цель, поэтому важным различием подходов является то, что маскирование основано на управлении *произвольным цифровым отпечатком* объекта. Этот отпечаток может содержать любое подмножество или его отображение из множества возможных параметров управления истинным или ложным объектом.

Технические реализации данного подхода предполагают управление параметрами аппаратного и программного обеспечения объектов вычислительных сетей (рис. 6).

Маскирование свойств объекта защиты в форме *мимикрии* позволяет настроить функционирование существующего объекта защиты (узла или системы узлов) в некотором смысле (с точки зрения некоторого функционала степени близости) похожего на какой-либо другой целевой объект (*редукция многообразия*). Например, *сервер-приманка* имеет близкие статические и динамические признаки функционирования с признаками сервера критической информационной инфраструктуры, что приводит к повышению неопределенности для злоумышленника относительно идентификации наиболее важной цели для атаки. Мимикрия может быть использована без создания ложных информационных объектов с целью снижения информативности демаскирующих признаков, свойственных критически важным узлам.

Второй формой маскирования является *имитация*, которая представляет собой создание ложных объектов с заданными характеристиками (*порождение многооб-*

разия). Предназначением ложных объектов является их использование в качестве мишеней для снижения возможностей противника по реализации атак на реальные объекты, а также для проведения *сетевой контрразведки*, то есть для получения информации относительно возможных алгоритмов (способов) Ω_a информационно-технического воздействия злоумышленника.

Стеганография и *шифрование* (криптографическое преобразование информации) являются обособленными методами, рассмотрение которых выходит за рамки данной работы, которые позволяют либо скрыть сам факт передачи информации, либо скрыть содержание передаваемой информации. При этом стойкость стеганографии определяется секретностью алгоритма (способа) скрытой передачи данных, а во втором случае секретностью, характеристиками ключа и криптоалгоритма.

Текущее состояние и перспективы исследований

С учетом типовой архитектуры клиент-серверных вычислительных сетей процесс маскирования структуры и процесса функционирования объектов защиты целесообразно рассмотреть на уровнях: *сетевых узлов*, *локального сегмента* (в пределах контролируемой зоны), *информационных направлений* (между обособленными локальными сегментами за пределами контролируемой зоны).

Анализ предметной области показал, что на уровне маскирования свойств *сетевых узлов* особое внимание уделялось вопросам: имитации ложными и мимикрии реальными сетевыми объектами свойств канальной среды низкого качества с целью исчерпания временных и вычислительных ресурсов средств сетевой разведки при взаимодействии с ними (управление потоком данных и параметрами фрагментации сообщений на сетевом, транспортном и прикладном уровнях стека протоколов *TCP/IP*), а также динамическому изменению структурно-функциональных характеристик узлов (управление *IP*, *MAC*-адресами, *TCP/UDP* портами, частотой их смены) [10-12].

На уровне маскирования свойств *локального сегмента* основными задачами явились: поиск оптимальных режимов многоадресного сетевого соединения и динамической адресации с учетом ресурсных ограничений (управление маской подсети; временем аренды *IP*-адресов, таблицами маршрутизации сети) [13-16].

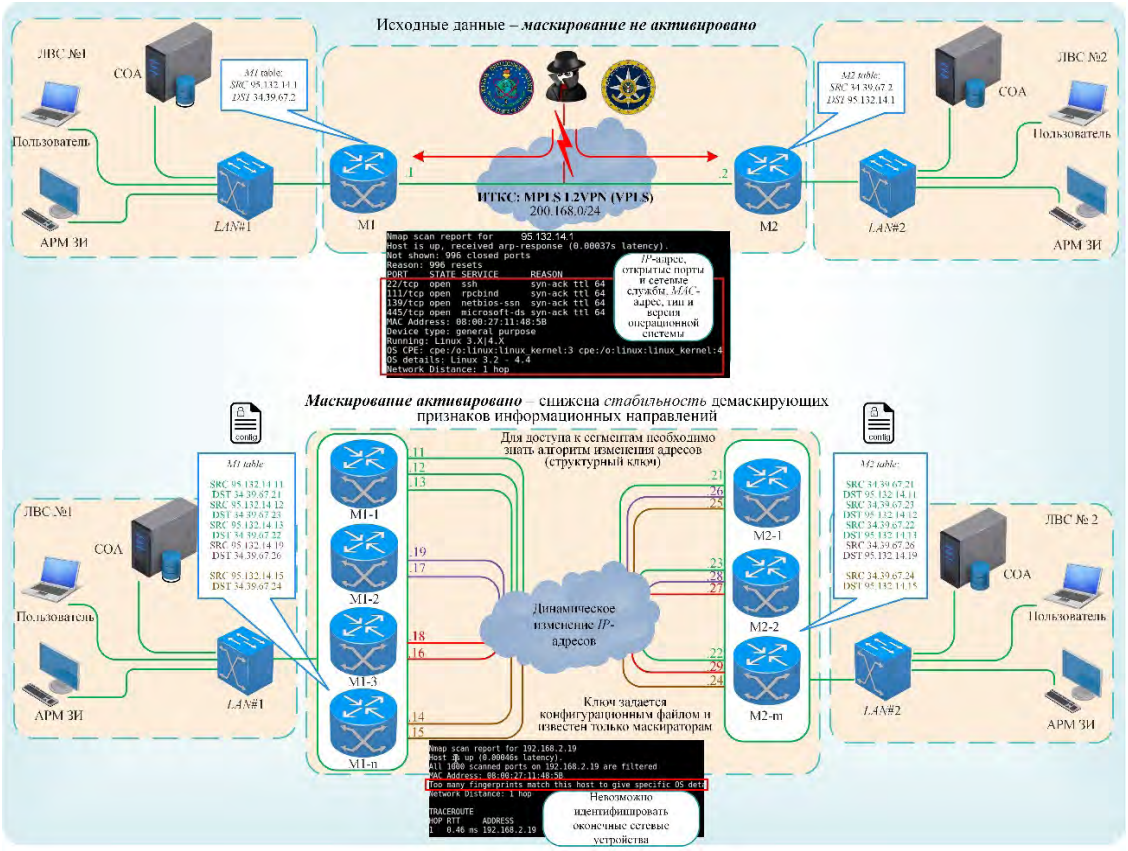

Рис. 7. Принципиальная схема маскирования информационных направлений

Исследования по маскированию свойств *информационных направлений* (рис. 7) были сосредоточены: на мимикрии реальной или имитации ложной иерархической структуры взаимодействующих локальных сегментов и узлов под заданную структуру (управление маршрутизацией между сегментами); моделировании динамических свойств для синтеза ложных (имитация), а также конфигурировании реальных сетевых информационных объектов (мимикрия) посредством генерации ложного (маскирующего) трафика) [17-19].

Научное обоснование вышеописанных разработок осуществлялось на основе: методов математической статистики (первичная обработка экспериментальных данных, проверка статистических гипотез, оценка неизвестных параметров статистических моделей), теории случайных процессов (марковские, полумарковские случайные процессы с дискретным пространством состояний для оценки вероятностно-временных характеристик неблагоприятных состояний информационного объекта, а также для использования полученных моделей в качестве уравнений связи или функционала качества в оптимизационных задачах), теории графов (алгоритмы построения минимальных остовных деревьев, синтез моделей информационных процессов в форме орграфов, решение транспортной задачи для маршрутизации трафика), методов прогнозирования и моделирования временных рядов (имитация ложного трафика с заданными динамическими характеристиками с использованием линейных регрессионных моделей, критериев самоподобия), теории алгоритмов (построение алгоритмов и оценка их свойств), теории оптимизации (использование точных и приближенных методов поиска экстремума функционала качества, заданного аналитически или алгоритмически), а также теории игр. Техническая реализация идей маскирования представлена в программно-аппаратных комплексах, генераторах ложного трафика, ложных сетевых информационных объектах («песочницы», *honeypot*, *honeynet*), специализированном программном обеспечении для реализации мимикрии в существующих системах.

Таким образом, в основу проведенных исследований положено выявление закономерностей с использованием классических математических моделей, имеющих теоретическую интерпретируемость результатов с учетом относительно грубых допущений (ограниченное последействие, нормальное распределение ошибки в значениях временных рядов), но основным недостатком традиционных методов с точки зрения синтеза моделей управления является необходимость структуризации данных, определения существенных свойств объектов. Структура и функционирование информационных объектов характеризуются значительными объемами данных, имеющими в общем случае нелинейные зависимости, оценка информативности которых крайне затруднительна.

Основной проблемой маскирования объектов вычислительных сетей является поиск и устранение *демаскирующих признаков,* по которым осуществляется классификация ложных узлов от истинных, узлов одной категории уязвимости от другой. Несмотря на глубокую теоретическую и прикладную проработку остаются практически не затронутыми методы, модели и алгоритмы машинного обучения при исследовании вопросов маскирования статических и динамических свойств (признаков) объектов, в частности методы *генеративного* и *вредоносного машинного обучения* (*состязательного*), подходящих под специфику научных задач.

Особенности алгоритмов машинного обучения, а именно зависимость их качества от свойств обучающих и тестовых наборов данных позволяет сделать предположение, что маскирование имеет значительное влияние на качество моделей, используемых в процессе компьютерной разведки. Маскирование инвариантов объектов вычислительных сетей по отношению к системам машинного обучения злоумышленника может быть интерпретировано как вредоносное машинное обучение, включающее в себя класс атак типа *уклонение* и *отравление данных* [20]. В терминах вредоносного машинного обучения уклонение использует *вредоносные образы*, которые позволяют избежать правильной классификации объекта алгоритмом машинного обучения. Сущность маскирования заключается в создании подобных вредоносных образов: изменении свойств или цифровых отпечатков сетевых информационных объектов, по которым осуществляется их классификация.

Отравление данных актуально при их использовании в процессе обучения классификаторов, то есть с целью изменения значений параметров моделей, оптимизируемых при обучении с маркированными данными. Так как неизвестно, когда и какие данные могут быть использованы злоумышленниками для обучения алгоритмов, то маскирование структуры и процессов функционирования информационных объектов приводит к уклонению в случае, если классификаторы злоумышленников уже обучены и к отравлению в противном случае.

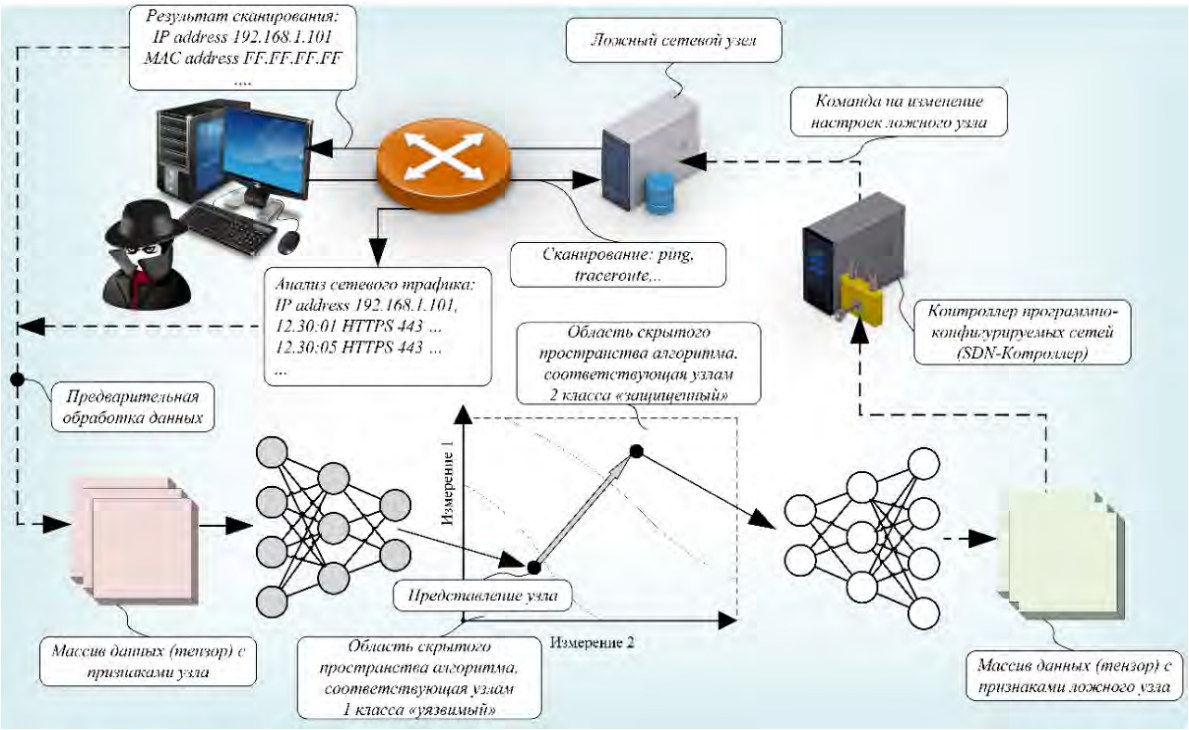

Рис. 8. Принципиальная схема использования алгоритмов машинного обучения для классификации и генерации узлов вычислительных сетей с заданными свойствами

Компьютерная разведка на уровне *сетевого узла* с использованием сканирования и анализа сетевого трафика позволяет получить множество статических (*IP*, *MAC*-адреса, *TCP/UPD*-порты, версии операционных систем, типы и версии сетевых протоколов) и динамических структурно-функциональных характеристик (распределение пакетов от одного источника по протоколам, объемы трафика за фиксированные интервалы времени), образующих *цифровой отпечаток узла* в необработанном виде. Методами глубокого обучения существует принципиальная возможность обработки и обобщения подобных неструктурированных массивов данных для получения сведений о распределении различных объектов вычислительной сети в метрике *скрытого* (*кодового*) *представления* алгоритма (рис. 8).

С точки зрения развития методов маскирования данное обстоятельство может быть использовано: для *генерации ложных узлов* с заданной степенью близости к реальным узлам вычислительной сети; для оценки *значимости* (информативности) первичных демаскирующих признаков посредством построения *карт значимости* признаков. Конкретная форма предварительной обработки (кодирования, масштабирования) и алгоритмов глубокого обучения зависят от специфики первичных данных о структурно-функциональных характеристиках информационных объектов. Сгенерированная конфигурация ложных или истинных сетевых узлов и сегментов сети может быть реализована технологией программно-конфигурируемых сетей.

На уровне *локальных сегментов* и *информационных направлений* использование технологий машинного обучения позволяет дополнительно к рассмотренным тензорам о свойствах узлов учесть сведения, относительно *топологии* и особенностях информационного обмена между узлами и подсетями (интенсивности и распределения трафика по узлам и сегментам, типы и версии сетевых протоколов, матрицы смежности пограничных маршрутизаторов).

Выводы

Научно-технический прогресс вызвал смещение принципов моделирования объектов окружающего мира от синтеза познавательных моделей типа «белый ящик» к автоматическому синтезу описательных моделей типа «черный ящик», предназначенных для целей управления. Моделирование или идентификация адекватной модели объекта вычислительной сети в ходе сетевой разведки является ключевым этапом, от которого зависит успех реализации информацион-

но-технических воздействий злоумышленников.

Проактивная защита как одна из базовых парадигм защиты позволяет обеспечить растущие информационные потребности с учетом качества услуг связи и автоматизации, а также защищенности объектов киберпространства за счет нарушения принципов моделирования их свойств. Маскирование и защита с использованием подвижной цели основаны на общих принципах управления структурно-функциональными характеристиками объектов.

Автоматизированный синтез цифровых отпечатков или инвариантов сетевых узлов, локальных сегментов и информационных направлений с применением технологий генеративного и состязательного машинного обучения, открывает новые возможности для развития средств и методов генерации ложных информационных объектов с заданными динамическими и статическими свойствами, соответствующими замыслу системы защиты.

Дальнейшие исследования в данной области будут направлены на разработку научно-методического аппарата и экспериментальную оценку эффективности маскирования объектов вычислительной сети с использованием различных архитектур алгоритмов глубокого обучения и технологий программно-конфигурируемых сетей.

Литература

1. Стародубцев Ю.И., Закалкин П.В., Иванов С.А. Техносферная война как основной способ разрешения конфликтов в условиях глобализации // Военная мысль. 2020. № 10. С. 16-21.
2. Daniel Ventre. Artificial Intelligence, Cybersecurity and Cyber Defense. Wiley. 2020. 237 p. ISBN 978-1-78630-467-4.
3. Сейновски, Т. Антология машинного обучения: «Издательство «Эксмо», 2022. 509 с.
4. Дауни, Аллен Б. Изучение сложных систем с помощью Python / пер. с анг. Д.А. Беликова. – М.: ДМК Пресс, 2019. 160 с.
5. Грибунин В.Г., Кондаков С.Е. К вопросу о защите информации в интеллектуализированных образцах вооружения // Вопросы кибербезопасности. 2021. № 5(45). С. 5–11. DOI: 10.21681/2311-3456-2021-5-5-11.
6. Aneesh Sreevallabh Chivukula, Xinghao Yang, Bo Liu, Wei Liu, Wanlei Zhou. Adversarial Machine Learning. Attack Surfaces, Defense Mechanisms, Learning Theories in Artificial Intelligence. Springer. 2023. 302 p. ISBN 978-3-030-99771-7.
7. Wang G., Ciptadi A., Ahmadzadeh A. Deployable Machine Learning for Security Defense. Communications in Computer and Information Science. 2020. Vol. 1271. 163 p.
8. Jin-Hee Cho, Dilli P. Sharma, Hooman Alavizadeh, Seunghyun Yoon, Noam Ben-Asher, Terrence J. Moore, Dong Seong Kim, Hyuk Lim, and Frederica F. Nelson. Toward Proactive, Adaptive Defense: A Survey on Moving Target Defense. arXiv:1909.08092v1 [cs.NI] 12 Sep 2019.
9. Kristin E. Heckman, Frank J. Stech. Roshan K. Thomas, Ben Schmoker, Alexander W. Tsow. Cyber Denial, Deception and Counter Deception. A Framework for Supporting Active Cyber Defense. Springer. 2023. 302 p. ISBN 978-3-030-99771-7.
10. Максимов Р.В., Орехов Д.Н., Соколовский С.П. Модель и алгоритм функционирования клиент-серверной информационной системы в условиях сетевой разведки // Системы управления, связи и безопасности. 2019. № 4. С. 50-99.
11. Лебедкина Т.В., Хорев Г.А. Модель функционирования и алгоритм конфигурирования адресации ложных сетевых информационных объектов в условиях сетевой разведки // Системы управления, связи и безопасности. 2023. № 2. С. 23-62.
12. Горбачев А.А. Модель и параметрическая оптимизация проактивной защиты сервиса электронной почты от сетевой разведки // Вопросы кибербезопасности. 2023. № 3(49). С. 69–81. DOI: 10.21681/2311-3456-2023-3-69-81.
13. Максимов Р.В., Соколовский С.П., Ворончихин И.С. Алгоритм и технические решения динамического конфигурирования клиент-серверных вычислительных сетей // Информатика и информатизация. 2020. № 5. С. 1018-1049.
14. Fraunholz D., Anton S.D., Lipps C., Reti D., Krohmer D., Pohl F., Tammen M., Schotten H.D. Demystifying Deception Technology: A Survey. pp. 1-25. arXiv:1804.06196v1 [cs.CR] 17 Apr 2018.
15. Ворончихин И.С., Иванов И.И., Максимов Р.В., Соколовский С.П. Маскирование структуры распределенных информационных систем в киберпространстве // Вопросы кибербезопасности. 2019. № 6(34). С. 92-99. DOI: 10.21681/2311-3456-2019-6-92-99.
16. Москвин А.А., Максимов Р.В., Горбачев А.А. Модель, оптимизация и оценка эффективности применения многоадресных сетевых соединений в условиях сетевой разведки // Вопросы кибербезопасности. 2023. № 3(55). С. 13-22.
17. Кучуров В.В., Максимов Р.В., Шерстобитов Р.С. Модель и методика маскирования адресации корреспондентов в киберпространстве // Вопросы кибербезопасности. 2020. № 6(40). С. 2-13. DOI: 10.21681/2311-3456-2020-6-2-13.
18. Соколовский С.П., Теленьга А.П. Методика формирования ложного сетевого трафика информационных систем для защиты от сетевой разведки // Вестник компьютерных и информационных технологий. 2022. № 2(212). С. 40-47.
19. Шерстобитов Р.С., Шарифуллин С.Р., Максимов Р.В. Маскирование интегрированных сетей связи ведомственного назначения // Системы управления, связи и безопасности. 2018. № 4. С. 136-175.
20. Уорр Кэти. Надежность нейронных сетей: укрепляем устойчивость ИИ к обману / СПб.: Питер, 2021. — 272 с.: ил. ISBN 978-5-4461-1676-8.

THE PROBLEM OF MASKING AND APPLYING OF MACHINE LEARNING TECHNOLOGIES IN CYBERSPACE

Gorbachev A.A.[5], Maximov R.V.[6]

The purpose of the study: *is to identify promising areas of scientific research in the field of masking cyberspace objects in the context of machine learning technologies.*

Methods used: *methods of general control theory and modeling, mathematical statistics, general scientific methods of analysis and synthesis.*

The result of the study: *the scientific problem of masking cyberspace objects and applying of machine learning technologies in the conditions of information and technical influences of intruders is determined. Improving masking methods at the level of network nodes, local segments and information directions using generative and adversarial machine learning methods will increase the security of cyberspace objects by reducing the effectiveness of network intelligence of intruders based on machine learning methods and algorithms. The following issues require deep theoretical and experimental study: evaluation of the form, content, informativity, preprocessing and generation of «digital fingerprints» of fake and true information objects, selection of types and optimal architecture of deep learning algorithms, evaluation of the quality of masking methods as «evasion» and «poisoning» attacks on machine learning algorithms of potential attackers.*

Scientific novelty: *consists in considering the concept of masking cyberspace objects in the conditions of information and technical impact of intruders from the standpoint of the general theory of control, modeling and application of machine learning technologies.*

Keywords: *modeling, control theory, proactive protection paradigm, network intelligence, machine learning.*

References

1. Starodubcev YU.I., Zakalkin P.V., Ivanov S.A. Tekhnosfernaya vojna kak osnovnoj sposob razresheniya konfliktov v usloviyah globalizacii // Voennaya mysl'. 2020. № 10. S. 16-21.
2. Daniel Ventre. Artificial Intelligence, Cybersecurity and Cyber Defense. Wiley. 2020. 237 p. ISBN 978-1-78630-467-4.
3. Sejnovski, T. Antologiya mashinnogo obucheniya: «Izdatel'stvo «Eksmo», 2022. 509 s.
4. Dauni, Allen B. Izuchenie slozhnyh sistem s pomoshch'yu Python / per. s ang. D.A. Belikova. – M.: DMK Press, 2019. 160 s.
5. Gribunin V.G., Kondakov S.E. K voprosu o zashchite informacii v intellektualizirovannyh obrazcah vooruzheniya // Voprosy kiberbezopasnosti. 2021. № 5(45). S. 5–11. DOI: 10.21681/2311-3456-2021-5-5-11.
6. Aneesh Sreevallabh Chivukula, Xinghao Yang, Bo Liu, Wei Liu, Wanlei Zhou. Adversarial Machine Learning. Attack Surfaces, Defense Mechanisms, Learning Theories in Artificial Intelligence. Springer. 2023. 302 p. ISBN 978-3-030-99771-7.
7. Wang G., Ciptadi A., Ahmadzadeh A. Deployable Machine Learning for Security Defense. Communications in Computer and Information Science. 2020. Vol. 1271. 163 p.
8. Jin-Hee Cho, Dilli P. Sharma, Hooman Alavizadeh, Seunghyun Yoon, Noam Ben-Asher, Terrence J. Moore, Dong Seong Kim, Hyuk Lim, and Frederica F. Nelson. Toward Proactive, Adaptive Defense: A Survey on Moving Target Defense. arXiv:1909.08092v1 [cs.NI] 12 Sep 2019.
9. Kristin E. Heckman, Frank J. Stech. Roshan K. Thomas, Ben Schmoker, Alexander W. Tsow. Cyber Denial, Deception and Counter Deception. A Framework for Supporting Active Cyber Defense. Springer. 2023. 302 p. ISBN 978-3-030-99771-7.
10. Maksimov R.V., Orekhov D.N., Sokolovskij S.P. Model' i algoritm funkcionirovaniya klient-servernoj informacionnoj sistemy v usloviyah setevoj razvedki // Sistemy upravleniya, svyazi i bezopasnosti. 2019. № 4. S. 50-99.
11. Lebedkina T.V., Horev G.A. Model' funkcionirovaniya i algoritm konfigurirovaniya adresacii lozhnyh setevyh informacionnyh ob"ektov v usloviyah setevoj razvedki // Sistemy upravleniya, svyazi i bezopasnosti. 2023. № 2. S. 23-62.
12. Gorbachev A.A. Model' i parametricheskaya optimizaciya proaktivnoj zashchity servisa elektronnoj pochty ot setevoj razvedki // Voprosy kiberbezopasnosti. 2023. № 3(49). S. 69–81. DOI: 10.21681/2311-3456-2023-3-69-81.
13. Maksimov R.V., Sokolovskij S.P., Voronchihin I.S. Algoritm i tekhnicheskie resheniya dinamicheskogo konfigurirovaniya klient-servernyh vychislitel'nyh setej // Informatika i informatizaciya. 2020. № 5. S. 1018-1049.

5 Alexander A. Gorbachev, Ph.D., Assistant Professor, Krasnodar Higher Military Orders of Zhukov and the October Revolution Red Banner School named after General of the Army S.M. Shtemenko, Krasnodar, Russia. E-mail: infosec23.00@mail.ru

6 Roman V. Maximov, Dr.Sc., Professor, Professor, Krasnodar Higher Military Orders of Zhukov and the October Revolution Red Banner School named after General of the Army S.M. Shtemenko, Krasnodar, Russia. E-mail: rvmaxim@yandex.ru

14. Fraunholz D., Anton S.D., Lipps C., Reti D., Krohmer D., Pohl F., Tammen M., Schotten H.D. Demystifying Deception Technology: A Survey. pp. 1-25. arXiv:1804.06196v1 [cs.CR] 17 Apr 2018.
15. Voronchihin I.S., Ivanov I.I., Maksimov R.V., Sokolovskij S.P. Maskirovanie struktury raspredelennyh informacionnyh sistem v kiberprostranstve // Voprosy kiberbezopasnosti. 2019. № 6(34). S. 92-99. DOI: 10.21681/2311-3456-2019-6-92-99.
16. Moskvin A.A., Maksimov R.V., Gorbachev A.A. Model', optimizaciya i ocenka effektivnosti primeneniya mnogoadresnyh setevyh soedinenij v usloviyah setevoj razvedki // Voprosy kiberbezopasnosti. 2023. № 3(55). S. 13-22.
17. Kuchurov V.V., Maksimov R.V., SHerstobitov R.S. Model' i metodika maskirovaniya adresacii korrespondentov v kiberprostranstve // Voprosy kiberbezopasnosti. 2020. № 6(40). S. 2-13. DOI: 10.21681/2311-3456-2020-6-2-13.
18. Sokolovskij S.P., Telen'ga A.P. Metodika formirovaniya lozhnogo setevogo trafika informacionnyh sistem dlya zashchity ot setevoj razvedki // Vestnik komp'yuternyh i informacionnyh tekhnologij. 2022. № 2(212). S. 40-47.
19. SHerstobitov R.S., SHarifullin S.R., Maksimov R.V. Maskirovanie integrirovannyh setej svyazi vedomstvennogo naznacheniya // Sistemy upravleniya, svyazi i bezopasnosti. 2018. № 4. S. 136-175.
20. Uorr Keti. Nadezhnost' nejronnyh setej: ukreplyaem ustojchivost' II k obmanu / SPb.: Piter, 2021. — 272 s.: il. ISBN 978-5-4461-1676-8.

МАСКИРОВАНИЕ МЕТАСТРУКТУР ИНФОРМАЦИОННЫХ СИСТЕМ В КИБЕРПРОСТРАНСТВЕ

Теленьга А.П.[1]

Цель исследования: *повышение защищенности информационных систем в киберпространстве от компьютерной разведки.*

Метод исследования: *методы математической статистики, нелинейной динамики, многокритериальной оптимизации.*

Результат исследования: *рассмотрены современные подходы к выделению уровней киберпространства, введено понятие метаструктуры информационной системы как протоколов и механизмов, которые обеспечивают интерфейс на различных уровнях киберпространства между базовыми компонентами системы, приложениями и сервисами, обслуживающими их, а также данными и информацией, сформулирована научная проблема маскирования метаструктур информационных систем в киберпространстве, заключающаяся в управления демаскирующими признаками метаструктур информационных систем: интенсивностью трафика между топологически локализованными сетевыми информационными объектами распределенной информационной системы, сетевыми протоколами взаимодействия и иерархическими уровнями (рангами) элементов информационной системы, постоянным перемещением между множественными конфигурациями информационной системы, на примере сетевого трафика поставлены непараметрическая и параметрическая задачи идентификации моделей метаструктур информационных систем.*

Научная новизна: *предложенная концепция отличается от известных выделением метаструктур информационных систем на различных уровнях киберпространства, постановкой задач маскировки, отравления, мимикрии и имитации информационной системы, управлением демаскирующими признаками метаструктур путем идентификации моделей информационных систем.*

Ключевые слова: *компьютерная разведка, компьютерная атака, сетевой трафик, идентификация модели, показатель Хёрста, динамическая трансформация временной шкалы, расстояние Кульбака-Лейблера.*

DOI:10.21681/4311-3456-2023-5-50-59

Введение

Рядом авторов, как зарубежных, так и отечественных [1-4], вводится понятие киберпространства как искусственного неоднородного технологического пространства со множеством разноуровневых органов оперативного и технологического управления, процесс создания и эксплуатации которого не предопределяется требованиями одной системы управления, а функционирует в интересах множества разнородных, в том числе антагонистических, систем управления, при этом его свойства зависят как от характеристик собственных элементов, так и от объема и свойств реализуемых процессов в интересах внутренних и внешних потребителей.

В соответствии с данным определением, информационные системы (далее – ИС), функционирующие в киберпространстве, представляют собой совокупность территориально распределенных сегментов, объединенных каналами связи различной протяженности с использованием коммуникационных технологий (оборудования) через сети связи общего пользования (ССОП) с целью предоставления пользователям информационных систем информационных ресурсов (программ и сервисов).

Формирование структуры ИС не происходит одномоментно. Она изменяется в ходе повседневной деятельности, внештатных ситуаций или преднамеренного информационно-технического воздействия. Образно говоря, ИС «мерцает» во времени, поэтому традиционное представление в виде двух- или трехмерных конструкций (граф, матрица связности) даёт возможность наблюдать лишь «срезы» ИС, причём в различные моменты времени наблюдаются их различные проявления.

1 Теленьга Александр Павлович, кандидат педагогических наук, докторант Краснодарского высшего военного орденов Жукова и Октябрьской Революции Краснознаменного училища имени генерала армии С. М. Штеменко, г. Краснодар, Россия. E-mail: alexander.telenga@yandex.ru, ORCID: 0000-0001-6193-0656

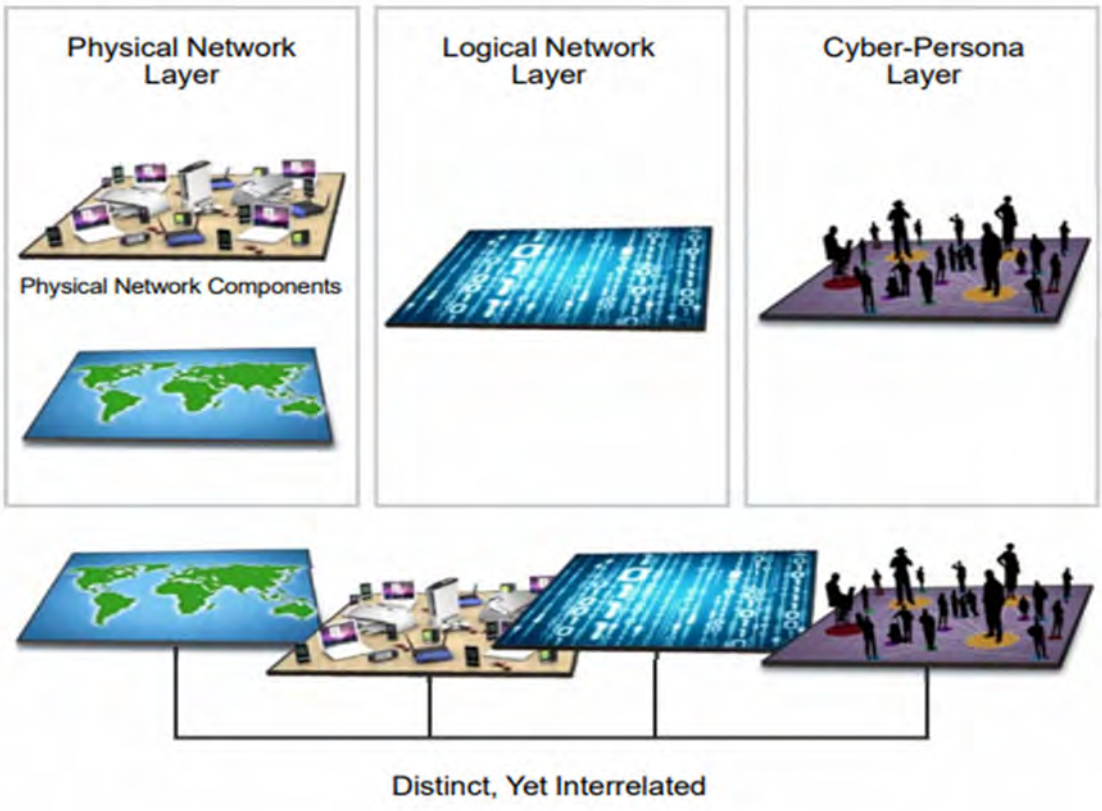

Рис. 1. Три взаимосвязанных уровня киберпространства согласно концепции Киберкомандования США

Согласно концепции Киберкомандования США [5] (рис. 1), в киберпространстве выделяются следующие уровни: кибер-идентификации, логической сети и физической сети.

Исходя из этого, логическая модель информационной системы в терминах структур позволяет выделить следующие уровни (рис. 2):

1. **Инфраструктура** – базовые компоненты системы: вычислительные мощности, сеть и хранилище данных.
2. **Апплиструктура** – приложения информационной системы и сервисы, обслуживающие их.
3. **Инфоструктура** – данные и информация. Содержимое баз данных, файловых хранилищ и т.д..
4. **Метаструктура** – протоколы и механизмы, которые обеспечивают интерфейс между инфраструктурой, структурой приложений и структурой данных в информационной системе, закон группы, которую образуют разнородные структуры.

Подобно «вторичным структурам» в геологии, возникающим в горной породе под влиянием позднейших процессов, например, механического, термального или химического воздействия, можно говорить о формировании в киберпространстве метаструктур ИС, которые могут обнаруживаться как информационные следы на соответствующем уровне киберпространства. Так, примером статистического следа на уровне логической сети является сетевой трафик ИС, семантического следа на уровне кибер-идентификации – служебная информация операционных систем, приложений, а структурный след на уровне физической сети проявляется в перколяционных процессах кластеров сетей передачи данных.

Предпосылки к идентификации моделей метаструктур информационных систем

В настоящее время киберпространство продолжает развиваться и усложняться. В связи с этим необходимо глубокое научное понимание закономерностей и тенденций этого развития. Требует в том числе внимания проблема информационно-технологического противоборства, методов ведения оборонительных, наступательных и разведывательных операций в киберпространстве, проблемы обеспечения надежного сдерживания в этой области.

Конфликты в киберпространстве характеризуются тем, что все его участники имеют развитые системы мониторинга и наблюдения состояния антагониста,

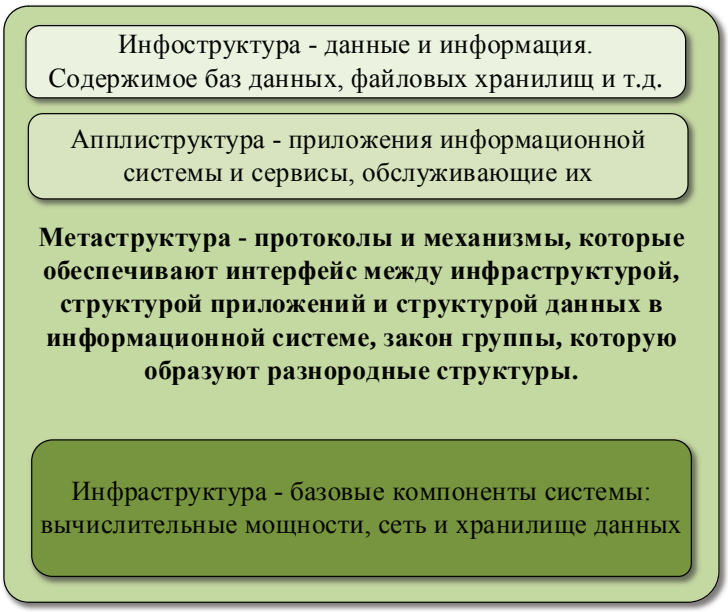

Рис. 2. Модель ИС в терминах структур

Рис. 3. Техники компьютерной разведки

системы информационного воздействия, а также собственные защищаемые информационно-управляющие системы.

Статичность, однородность и детерминированность, определяющие постоянство состава, структуры и алгоритмов функционирования ИС, обуславливают наличие у злоумышленника ряда преимуществ в использовании временного и вычислительного ресурса для ведения компьютерной разведки (КР), обеспечивающих ему:

— возможность с небольшими ресурсными затратами проводить крупномасштабную атаку после успешного проведения мелкомасштабной атаки;
— высокую достоверность результатов КР в течение длительного времени, что позволяет осуществлять планирование, выбор времени и технологического процесса ИС для начала компьютерных атак (КА);
— возможность бескомпроматного применения средств КР и реализации КА, в любое удобное для этого время, за счет заблаговременного (планового) формирования и применения их оптимального набора;
— возможность неоднократного обнаружения и анализа уязвимостей аппаратного и программного обеспечения, с последующим их тестированием на проникновение для конкретной цели и т.д.

Согласно методологии описания этапов компьютерной атаки *Cyber KillChain*, разработанной компанией *Lockheed Matrin* [6], любая атака начинается с разведки. База знаний тактик и техник злоумышленников *MITRE ATT&CK* выделяет 10 групп техник разведки (рис. 3).

Разнообразие этих техник и их комбинаций на различных уровнях киберпространства, а также возможность анализа полученной информации методами

глубокого анализа данных, существенно повышает вероятность обнаружения вскрытия структуры ИС через выявленные метаструктуры, а значит, и осуществления вредоносного воздействия.

В настоящее время в состав мер защиты информации в государственных информационных системах включены:

— сокрытие архитектуры и конфигурации ИС;
— создание (эмуляция) ложных ИС или их компонентов, предназначенных для обнаружения, регистрации и анализа действий нарушителей в процессе реализации угроз безопасности информации;
— воспроизведение ложных и (или) скрытие истинных отдельных информационных технологий и (или) структурно-функциональных характеристик ИС или ее сегментов, обеспечивающее навязывание нарушителю ложного представления об истинных информационных технологиях и (или) структурно-функциональных характеристиках ИС.

Это вызвано тем, что достаточно большое количество КА выполняется средствами КР с целью получения информации о составе, структуре и алгоритмах функционирования, местоположении и принадлежности ИС, а также данных, хранимых, обрабатываемых и передаваемых в таких ИС. Наряду с угрозами безопасности информации, связанными с диалоговым взаимодействием нарушителя и ИС (в частности – автоматизированными средствами сетевого сканирования), на реконструкцию структурно-функциональных характеристик ИС нацелена угроза определения топологии ИС, бескомпроматно реализуемая анализом сетевого трафика. Результат – вскрытие топологии распределенной в киберпространстве ИС, определение важности ее узлов – может быть использован нарушителем для реализации спланированных *APT*-атак (от англ. *advanced persistent threat* – «развитая устойчивая угроза», целевая кибератака).

Задача реализации перечисленных выше мер защиты информации может быть решена маскированием метаструктур ИС [7] – совокупностью ложных (маскирующих) техник, выполняемых сетевыми информационными объектами (СИО) с целью управления демаскирующими признаками (ДМП) метаструктур ИС: интенсивностью трафика между топологически локализованными СИО распределенной ИС, сетевыми протоколами взаимодействия и иерархическими уровнями (рангами) элементов ИС, постоянным перемещением между множественными конфигурациями ИС, что увеличивает неопределенность данных у КР и лишает её возможности ретроспективного анализа данных разведки.

Обстановка в разных условиях функционирования ИС требует решения следующих задач маскирования [8-10]:

1. Необходимо дополнить метаструктуры ИС до метаструктур ССОП, иными словами, необходима *маскировка* метаструктур ИС под метаструктуры ССОП, поскольку разнообразие уникальных цифровых отпечатков устройств ИС – набора параметров, позволяющих однозначно идентифицировать устройство пользователя –несоизмеримо меньше, чем устройств ССОП.
2. Необходимо *«отравить» (насытить)* метаструктуру ИС ложными данными для снижения эффективности средств КР. В этом случае применение злоумышленником, например, методов глубокого анализа данных для ведения КР будет существенно затруднено.
3. Необходимо *имитировать* метаструктуру ИС для обеспечения успешного киберманеврирования. Сущность киберманевра заключается в искусственном расширении поверхности атаки за счёт создания ложных целей.
4. Необходима *мимикрия* метаструктур ССОП под метаструктуры ИС с целью введения в заблуждение средств КР и отвлечения внимания.

При этом перечисленные выше задачи могут решаться как по отдельности, так и совместно, образуя комплекс средств маскирования.

Очевидно, что для успешного решения поставленных задач необходимо вскрыть закономерности изменения метаструктур ИС во времени, проведя реконструкцию динамических моделей (такой термин принят в нелинейной динамике), или идентификацию систем (в терминах математической статистики), т.е. определение структуры и параметров (параметрическая идентификация) или наилучшей аппроксимации характеристик (непараметрическая идентификация) по полученному экспериментальному набору данных (записанных входных и выходных сигналов) [11-13]. Математическая модель метаструктур ИС в этом случае задаётся в виде уравнений, описывающих связь одной или нескольких случайных переменных с другими переменными (случайными и детерминистическими).

Задача непараметрической идентификации модели метаструктур информационных систем

Особенностью методов непараметрической идентификации (получение описания одной или нескольких случайных переменных модели) является то, что в них либо не учитывается закон распределения полученных

данных, либо учитывается с неявно определенными параметрами. Другими словами, в методах непараметрического идентификации статистическая модель (и ее структура) не имеет фиксированного числа параметров.

К таковым относятся:

— построение графика функции плотности распределения вероятности (например, гистограммой);
— ядерная оценка (сглаживание) плотности распределения (под ядром понимается некоторая весовая функция);
— непараметрическая регрессия (на базе ядер, сплайнов, вейвлетов и др.);
— ряд методов классификации и кластеризации (например, kNN, SVM);
— проверка статистических гипотез.

Построение графика функции плотности распределения вероятности гистограммой проводится следующим образом. Для набора из N случайных переменных $\{X_1, X_2, ..., X_N\}$ количество c_i попаданий X_j в i-ый подинтервал $[a_{i-1}, a_i]$ исходного интервала $[a_0, a_n]$ для $i = 1, 2, ..., n$ (теоретически, $[-\infty, +\infty]$) определяется как

$$c_i = \sum_{j=1}^{N} \{1 : X_j \in [a_{i-1}, a_i]\} \quad (1)$$

Тогда кусочнопостоянная функция $h(x)$, называемая нормализованной гистограммой, оценивается следующим образом:

$$h(x) = \frac{c_i}{N\Delta a_i} = \frac{c_i}{N(a_i - a_{i-1})}. \quad (2)$$

Нормализованная гистограмма есть графическая интерпретация функции плотности распределения вероятности.

Кроме нормализованной гистограммы используется интегральная гистограмма. Для набора из N случайных переменных $\{X_1, X_2, ..., X_N\}$ количество c_i попаданий X_j в i-ый подинтервал $[a_0, a_i]$ исходного интервала $[a_0, a_n]$ для $i = 1, 2, ..., n$ (теоретически, $[-\infty, +\infty]$) определяется как:

$$c_i = \sum_{j=1}^{N} \{1 : X_j \in [a_0, a_i]\} \quad (3)$$

Нормализованная интегральная гистограмма $h(x)$, являющаяся функцией распределения вероятности, оценивается следующим образом:

$$h^{инт}(x) = \frac{c_i}{N\Delta a_i} = \frac{c_i}{N(a_i - a_0)}. \quad (4)$$

Другим видом непараметрической идентификации является ядерная оценка плотности или ядерное сглаживание (kernel density estimation). Для набора из N случайных переменных $\{X_1, X_2, ..., X_N\}$ форма (огибающая) функции плотности распределения вероятности определяется как

$$f_{KDE}(x) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - X_j) = \frac{1}{Nb}\sum_{j=1}^{N} K_h\left(\frac{x - X_j}{b}\right) \quad (5)$$

где $K(\cdot)$ – ядро, неотрицательная функция, интегрируемая в 1, $b > 0$ – параметр сглаживания.

Проверка статистических гипотез направлена на численное принятие решения о том, удовлетворяет ли статистическая выборка заданной статистической гипотезе. С практической точки зрения, как правило, речь идёт о следующем:

— описывается ли случайная переменная из имеющейся выборки заданным законом распределения;
— принадлежат ли две случайные величины из имеющейся выборки к одному закону распределения.

Задача параметрической идентификации модели метаструктур информационных систем

В случае параметрической идентификации под моделью метаструктуры ИС будем понимать отображение структуры S и параметров X в свойства Y:

$$Y = F(S, X_S). \quad (6)$$

Пусть $\mathbf{Y}^{треб}$ – вектор требуемых свойств метаструктуры ИС. Тогда задача идентификации модели заключается в определении множества структур Ω_S и параметров $\Omega_X : \mathbf{Y}^{треб} = F(S, X_S), S \in \Omega_S, X_S \in \Omega_X(S)$.

Задача сводится к экстремальной задаче

$$\left|\mathbf{Y}^{треб} - F(S, X_S)\right| \to \min_{S, X_S}. \quad (7)$$

Все возможные результаты решения экстремальной задачи образуют множества Ω_S и Ω_X.

На стадии оптимизации производится синтез оптимальной конструкции из допустимого множества структур и параметров. Для этого прежде всего необходимо задать критерии оптимальности. Эти критерии могут быть трех типов:

а) типа неравенств:

$H(S, X_S) \geq 0$ или $h_i(S, X_S) \geq 0, i = 1, ..., m$, где $H = (h_1, h_2, ..., h_m)$;

б) типа равенств:

$G(S, X_S) = 0$ или $g_i(S, X_S) = 0, i = 1, ..., p$, где $G = (g_1, g_2, ..., g_p)$;

в) экстремального типа:

$Q(S, X_S) \to extr$ или $q_i(S, X_S) \to 0, i = 1, ..., k$, где $Q = (q_1, q_2, ..., q_k)$, т. е. экстремальная задача имеет многокритериальный характер.

Вид критериев H, G и Q определяется, исходя из технологических, эксплуатационных и других соображений. В общем виде задача оптимальной идентификации моделей метаструктур информационных формулируется в виде

$$Q(S, X_S) \to \underset{S, X_S \in \Psi}{extr}, \quad (8)$$

где

$$\Psi = \begin{cases} H(S, X_S) \geq 0, \\ G(S, X_S) = 0, \\ S \in \Omega_S, \\ X_S \in \Omega_X(S) \end{cases}. \quad (9)$$

Постановка задачи идентификации модели сетевого трафика

Рассмотрим в качестве метаструктуры, модель которой необходимо идентифицировать, сетевой трафик ИС.

Существует несколько подходов к описанию сетевого трафика ИС: в виде потоков и в виде последовательности пакетов («сырой» трафик).

Потоки содержат заголовочную информацию о сетевых соединениях между двумя конечными устройствами, такими как серверы или рабочие станции. Каждый поток представляет собой совокупность переданных сетевых пакетов, которые имеют некоторые общие свойства. Как правило, все передаваемые сетевые пакеты с одинаковыми IP-адресом источника, портом источника, *IP*-адресом назначения, порт назначения и транспортным протоколом в пределах временного окна объединяются в один поток [14, 15].

«Сырой» трафик, как правило, представляет собой последовательность пакетов, каждый из которых содержит время отправки пакета, *IP*-адрес источника, порт источника, IP-адрес назначения, порт назначения, протокол, размер пакета, установленные флаги и поле данных, в которое записывается полезная нагрузка [16, 17].

Известно, что сетевой трафик обладает свойством самоподобия его статистических свойств в *IP*-сетях не только в текущий момент времени, но и ретроспективно. Это означает, что присутствует повторяемость статистических характеристик естественных временных рядов с изменением масштаба. Процессы, обладающие свойствами самоподобия, характеризуются наличием последействия за счет факторов, вызывающих сложные зависимости: при относительно низкой средней скорости поступления пакетов сообщений возможны большие всплески интенсивности [18]. Статистические характеристики такого процесса – ДМП конкретной информационной системы.

Общепринятым показателем самоподобия процесса является показатель Хёрста H, в зависимости от значений которого делают следующие выводы об исследуемых процессах: при $0 \leq H \leq 0,5$ случайный процесс не обладает самоподобием; при $H > 0,5$ – процесс обладает длительной памятью и является самоподобным.

Таким образом, разность между показателем Хёрста эталонного и модельного трафика может выступать в качестве метрики близости модели и реального сетевого трафика:

$$q_1(S, X_S) = \left| H_{ориг} - H_{модель} \right| \to \min. \quad (10)$$

Ещё одним критерием выступает динамическая трансформация временной шкалы *Dynamic time warping* (*DTW*) – это метод анализа временных рядов, который позволяет сравнивать и выявлять сходства между двумя временными рядами, имеющими различную скорость изменения [19]. Он основан на алгоритме динамического программирования, который вычисляет оптимальное выравнивание между двумя временными рядами, учитывая возможные различия в скорости изменения.

$$q_2(S, X_S) = \left| DTW(Y^{треб}, F(S, X_s)) \right| \to \min. \quad (11)$$

Формальная постановка непараметрической идентификации модели времени задержек между пакетами сетевого трафика формулируется следующим образом: необходимо обеспечить минимальность разности между показателями самоподобия исходного и модельного временных рядов (10), а также минимальность динамической трансформации временной шкалы (11) в соответствующей метрике при заданном допустимом множестве структур S и параметров X_S:

$$\Psi = \begin{cases} H_{модель} \geq 0,5, \\ q_1(S, X_S) \leq 10^{-3}, \\ q_2(S, X_S) \leq 10, \\ S \in \{S_1, S_2, S_3, S_4, S_5\}, \\ X_s \in \{X_{S_1}, X_{S_2}, X_{S_3}, X_{S_4}, X_{S_5}\} \\ S_1 = f(X_{S_1}), S_2 = f(X_{S_2}), S_3 = f(X_{S_3}), \\ S_4 = f(X_{S_4}), S_5 = f(X_{S_5}) \\ X_{S_1} = \{N_{KN}, b_{KN}\}, X_{S_2} = \{N_{KE}, b_{KE}\}, X_{S_3} = \{A, B\}, \\ X_{S_4} = \{\mu_{LN}, \sigma_{LN}\}, X_{S_5} = \{\mu_{HN}, \sigma_{HN}\}, \\ b_{KN} > 0, b_{KE} > 0, A > 0, B > 0, \sigma_{LN} \geq 0, \ \sigma_{HN} \geq 0 \end{cases}. \quad (12)$$

S – тип модельного оператора.

X_S – параметры модельного оператора.

Прямое решение задачи векторной оптимизации представляет собой множество параметров, оптимальных по Парето. Исходя из равнозначности критериев, характера достижимого критериального пространства и поставленной цели исследования, выбор одного оптимального набора типа и параметров модельного оператора целесообразно осуществлять с использованием метода идеальной точки. За «идеальную точку» принимаются экстремальные (идеальные) значения целевых функций $q_i(S, X_S)$.

Тогда скалярная целевая функция $R(S, X_S)$ имеет физический смысл евклидовой метрики (расстояния) между «идеальной точкой» и точкой фронта Парето, а выбор оптимального набора значений исходных целевых функций и факторов аргументов соответствует минимальному значению указанного расстояния, тогда скалярная целевая функция имеет вид:

$$\begin{cases} R = \sqrt{(q_1(S, X_S) - 0)^2 + (q_2(S, X_S) - 0)^2} \\ R \to \min\limits_{S, X_S \in \Psi} \end{cases}, \quad (13)$$

где (0,0) – координаты идеальной точки в критериальном пространстве $q_1(S, X_S) \times q_2(S, X_S)$.

Исходя из характера исследуемых процессов, в качестве модельных операторов могут использоваться следующие.

— Гауссово ядро:

$$S_1 = f(x \mid N, b) = \frac{1}{\sqrt{2\pi} Nb} \sum_{j=1}^{N} e^{-\frac{1}{2}\left(\frac{x - X_j}{b}\right)^2} \quad (14)$$

— Епанечниково ядро:

$$S_2 = f(x \mid N, b) = \frac{3}{4Nb} \sum_{j=1}^{N} \left(1 - \left(\frac{x - X_j}{b}\right)^2\right), |x| < 1 \quad (15)$$

— Функции плотностей распределений Вейбула, логнормального и полунормального соответственно:

$$S_3 = f(x \mid a, b) = \begin{cases} \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-(x/a)^b}, x \geq 0 \\ 0, x < 0 \end{cases} \quad (16)$$

$$S_4 = f(x \mid \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left\{\frac{-(\log x - \mu)^2}{2\sigma^2}\right\}, x > 0 \quad (17)$$

$$S_5 = f(x \mid \mu, \sigma) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \geq \mu \quad (18)$$

Динамическая трансформация временной шкалы DTW использует метрику расстояния между временными рядами. Поскольку непараметрическая идентификация подразумевает вероятностный характер моделей, в качестве метрики выбрана симметричная дивергенция Кульбака-Лейблера [20].

$$d_{mn}(\mathbf{X}, \mathbf{Y}) = \sum_{k=1}^{K} (x_{k,m} - y_{k,n})(\log x_{k,m} - \log y_{k,n}) \quad (19)$$

Формальная постановка параметрической идентификации модели времени задержек между пакетами сетевого трафика формулируется следующим образом: необходимо обеспечить минимальность разности между показателями самоподобия исходного и модельного временных рядов (10), а также минимальность динамической трансформации временной шкалы (11) в метрике евклидового расстояния при множестве допустимых параметров

$$\Psi = \begin{cases} H_{модель} \geq 0,5, \\ q_1(S, X_S) \leq 10^{-3}, \\ q_2(S, X_S) \leq 30, \\ S \in \{S_1, S_2, S_3\}, \\ X_s \in \{X_{S_1}, X_{S_2}, X_{S_3}\}, \\ S_1 = f(X_{S_1}), S_2 = f(X_{S_2}), S_3 = f(X_{S_3}), \\ X_{S_1} = \{A, B\}, X_{S_2} = \{\sigma, \rho, \beta\}, X_{S_3} = \{a, b, \tau\}, \\ 0 < A \leq 50, 0 < B \leq 50, 0 < \sigma \leq 20, 0 < \beta \leq 20, \\ 0 < \rho \leq 17, 0 < a \leq 20, 0 < b \leq 20, 0 < \tau \leq 17 \end{cases} \quad (20)$$

При этом в качестве метрики расстояния между рядами для DTW может быть выбрано евклидово расстояние:

$$d_{mn}(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{k=1}^{K} (x_{k,m} - y_{k,n})(x_{k,m} - y_{k,n})} \quad (21)$$

В качестве модельных операторов используются динамические системы, известные самоподобным поведением [21].

— Уравнение нелинейного осциллятора Ван дер Поля

$$S_1 = f(x \mid A, B) = \frac{d^2x}{d^2t} - A(1 - B\frac{dx}{dt}) + x = 0. \quad (22)$$

— Уравнение Лоренца

$$S_2 = f(x, y, z \mid \sigma, \rho, \beta) = \begin{cases} \frac{dx}{dt} = \sigma(y - x) \\ \frac{dy}{dt} = x(\rho - z) - y \\ \frac{dz}{dt} = xy - \beta z \end{cases} \quad (23)$$

— Уравнение генератора Мэки-Гласса

$$S_3 = f(x \mid a,b,\tau) = \frac{dx(t)}{dt} - \frac{ax(t-\tau)}{1+x(t-\tau)^{10}} + bx(t) = 0 \quad (24)$$

Получившаяся задача векторной оптимизации также может быть сведена к скалярной методом идеальной точки (13).

Выводы

Анализ современных подходов к выделению уровней киберпространства позволяет утверждать, что функционирование антагонистических систем происходит на уровне кибер-идентификации, логической сети и физической сети. Предложенное понятие метаструктуры информационной системы, имманентное её информационным следам, может быть определено как протоколы и механизмы, которые обеспечивают интерфейс на различных уровнях киберпространства между базовыми компонентами системы, приложениями и сервисами, обслуживающими их, а также данными и информацией.

Статичность, однородность и детерминированность, определяющие постоянство состава, структуры и алгоритмов функционирования информационных систем обуславливают наличие у злоумышленника ряда преимуществ в использовании временного и вычислительного ресурса для ведения компьютерной разведки, в связи с чем сформулирована научная проблема маскирования метаструктур информационных систем в киберпространстве, заключающаяся в управления демаскирующими признаками метаструктур информационных систем: интенсивностью трафика между топологически локализованными сетевыми информационными объектами распределенной информационной системы, сетевыми протоколами взаимодействия и иерархическими уровнями (рангами) элементов информационной системы, постоянным перемещением между множественными конфигурациями информационной системы.

Решение указанных задач невозможно без вскрытия закономерностей изменения метаструктур информационных систем во времени, для чего необходимо провести идентификацию их моделей или реконструкцию динамических систем, описывающих процесс функционирования информационных систем. На примере сетевого трафика поставлены непараметрическая и параметрическая задачи идентификации моделей метаструктур информационных систем, определены критерии оптимальности и сформулирована экстремальная задача выбора оптимальных параметров модели.

Таким образом, в условиях ведения злоумышленником компьютерной разведки и реализации компьютерных атак выделение метаструктур ИС, идентификация их моделей и маскирование методами введения в заблуждение и повышения неопределенности позволит обеспечить гибкость, адаптивность и повысит эффективность системы защиты.

Научный консультант: Максимов Роман Викторович, доктор технических наук, профессор, профессор Краснодарского высшего военного орденов Жукова и Октябрьской Революции Краснознаменное училища имени генерала армии С. М. Штеменко, г. Краснодар, Россия. E-mail: rvmaxim@yandex.ru

Литература

1. Стародубцев, Ю. И. Техносферная война как основной способ разрешения конфликтов в условиях глобализации / Ю. И. Стародубцев, П. В. Закалкин, С. А. Иванов. // Военная Мысль. — 2020. — № 10. — С. 16-21.
2. Стародубцев, Ю. И. Концептуальные направления решения проблемы обеспечения устойчивости Единой сети электросвязи Российской Федерации / Ю. И. Стародубцев, С. А. Иванов, П. В. Закалкин. // Военная Мысль. — 2021. — № 4. — С. 39-49.
3. Zdzikot, T. Cyberspace and Cybersecurity / T. Zdzikot. // Cybersecurity in Poland. — Cham: Springer, 2022. — С. 9-21. DOI 10.1007/978-3-030-78551-2_2
4. Theoretical basis and technical methods of cyberspace geography / Gao Chundong, Guo Qiquan, Jiang Dong [и др.]. // Journal of Geographical Sciences. — 2019. — № 29. — С. 1949–1964.
5. Joint Chiefs of Staff. Cyberspace operations. Joint Chiefs of Staff (US); 19 2022 Dec 19. Joint Publication No.: JP 3-12. // Official Website of the Joint Chiefs of Staff: [сайт]. — URL: https://www.jcs.mil/Doctrine/Joint-Doctrine-Pubs/3-0-Operations-Series/ (дата обращения: 26.08.2023).
6. Lockheed Martin's Cyber-Kill Chain. // Leading Aerospace and Defense | Lockheed Martin: [сайт]. — URL: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html (дата обращения: 26.08.2023).
7. Теленьга А.П. Анализ возможностей обнаружения метаструктур информационных систем специального назначения. / А. П. Теленьга // Кибербезопасность: угрозы, тенденции, технологии защиты: материалы II Межведомственной науч.-практич. конф., 19-20 мая 2022 г. / Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С.М.Штеменко. – Краснодар: КВВУ, 2022. С.70-76
8. Zheng, Y. Dynamic defenses in cyber security: Techniques, methods and challenges / Y. Zheng, Z. Li, X. Xu // Digital Communications and Networks. — 2022. — № 8. — С. 422–435.

9. Shaping Attacker Behavior: Evaluation of an Enhanced Cyber Maneuver Framework / J. A. McKneely, T. K. Sell, K. A. Straub [и др.] // HCII 2022: HCI for Cybersecurity, Privacy and Trust. — Cham: Springer, 2022. — С. 358–379. – DOI: 10.1007/978-3-031-05563-8_23
10. Lilli, E. How Can We Know What We Think We Know about Cyber Operations? / E. Lilli // Journal of Global Security Studies. — 2023. — № 8(2). — С. 1–18. – DOI 10.1093/jogss/ogad011
11. Dynamic-chaos information technologies for data transmission, storage, and protection / Yu. V. Gulyaev, R. V. Belyaev, G. M. Vorontsov [et al.] // Радиоэлектроника. Наносистемы. Информационные технологии. – 2018. – Vol. 10, No. 2. – P. 279-312. – DOI 10.17725/rensit.2018.10.279.
12. Четвертакова, Ю. С. Построение моделей стохастических нелинейных динамических систем на основе двухэтапной процедуры параметрической идентификации / Ю. С. Четвертакова, О. С. Черникова // Наука. Технологии. Инновации: Сборник научных трудов. В 9-ти частях, Новосибирск, 30 ноября – 04 2020 года / Под редакцией А.В. Гадюкиной. Том Часть 2. – Новосибирск: Новосибирский государственный технический университет, 2020. – С. 94-98.
13. Карганов, В. В. К вопросу о необходимости моделирования информационных систем организации, функционирующих в условиях угроз безопасности / В. В. Карганов // Национальная безопасность России: актуальные аспекты : Сборник статей Всероссийской научно-практической конференции, Санкт-Петербург, 30 июля 2019 года. – Санкт-Петербург: Частное научно-образовательное учреждение дополнительного профессионального образования Гуманитарный национальный исследовательский институт «НАЦ-РАЗВИТИЕ», 2019. – С. 20-30.
14. Будко, Н. П. Общие принципы функционирования и требования к построению структур перспективных систем мониторинга распределенных информационно-телекоммуникационных сетей / Н. П. Будко // Техника средств связи. – 2021. – № 2(154). – С. 38-59.
15. Голованов, А. А. Сравнительный анализ систем обнаружений аномалий с использованием потока сетевого трафика и протокола NETFLOW / А. А. Голованов, О. И. Мельникова // Международный научно-исследовательский журнал. – 2023. – № 6(132). – DOI 10.23670/IRJ.2023.132.18.
16. Полтавцева, М. А. Формирование структур данных в задачах активного мониторинга безопасности // Проблемы информационной безопасности. Компьютерные системы. – 2021. – № 1. – С. 9-19.
17. Alothman, B., Raw Network Traffic Data Preprocessing and Preparation for Automatic Analysis // 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). – 2019. – P. 1-5, DOI: 10.1109/CyberSecPODS.2019.8885333.
18. Черниговский, А. В. Оценка степени самоподобия сетевого трафика / А. В. Черниговский, М. В. Кривов // Информационные технологии. Проблемы и решения. – 2019. – № 1(6). – С. 115-120.
19. Chen, L. A deep multi-task representation learning method for time series classification and retrieval / L. Chen, D. Chen, F. Yang, J. Sun // Information Sciences. – 2021. – Vol. 555. – P. 17-32. – DOI 10.1016/j.ins.2020.12.062.
20. Калинин, М. Ю. Энтропийные оценки решающих статистик алгоритма классификации случайных процессов / М. Ю. Калинин, О. Н. Чопоров // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8, № 4(31). – DOI 10.26102/2310-6018/2020.31.4.034.
21. Соколовский, С. П. Методика формирования ложного сетевого трафика информационных систем для защиты от сетевой разведки / С. П. Соколовский, А. П. Теленьга // Вестник компьютерных и информационных технологий. – 2022. – Т. 19, № 2(212). – С. 40-47. – DOI: 10.14489/vkit.2022.02.pp.040-047.

MASKING METASRTRUCTURES OF INFORMATION SYSTEMS IN CYBERSPACE

***Telenga A.P.*[2]**

Research objective: *to improve the security of information systems in cyberspace against computer reconnaissance.*

Research method: *methods of mathematical statistics, nonlinear dynamics, multicriteria optimization.*

Research results: *modern approaches to the allocation of cyberspace levels are considered, the concept of information system metastructure is introduced as protocols and mechanisms that provide an interface at different levels of cyberspace between the basic components of the system, applications and services that serve them, as well as data and information, the scientific problem of masking metastructures of information systems in cyberspace is formulated, which consists in the management of demasking features of metastructures of information systems in cyberspace.*

Scientific novelty: *the proposed concept differs from the known ones by singling out metastructures of information systems at different levels of cyberspace, setting tasks of masking, poisoning, mimicry and imitation of information systems, management of demasking features of metastructures by identifying models of information systems.*

Keywords: *computer reconnaissance, computer attack, network traffic, model identification, Hurst index, dynamic time warping, Kulbak-Leibler distance.*

2 Telenga Alexander Pavlovich, candidate of pedagogical sciences, Doctoral student, Krasnodar Higher Military Orders of Zhukov and the October Revolution Red Banner School named after General of the Army S.M. Shtemenko, Krasnodar, E-mail: alexander.telenga@yandex.ru, ORCID: 0000-0001-6193-0656

References

1. Starodubcev, Yu. I. Texnosfernaya vojna kak osnovnoj sposob razresheniya konfliktov v usloviyax globalizacii / Yu. I. Starodubcev, P. V. Zakalkin, S. A. Ivanov. // Voennaya My`sl`. — 2020. — № 10. — S. 16-21.
2. Starodubcev, Yu. I. Konceptual`ny`e napravleniya resheniya problemy` obespecheniya ustojchivosti Edinoj seti e`lektrosvyazi Rossijskoj Federacii / Yu. I. Starodubcev, S. A. Ivanov, P. V. Zakalkin. // Voennaya My`sl`. — 2021. — № 4. — S. 39-49.
3. Zdzikot, T. Cyberspace and Cybersecurity / T. Zdzikot. // Cybersecurity in Poland. — Cham: Springer, 2022. — S. 9-21. DOI 10.1007/978-3-030-78551-2_2
4. Theoretical basis and technical methods of cyberspace geography / Gao Chundong, Guo Qiquan, Jiang Dong [i dr.]. // Journal of Geographical Sciences. — 2019. — № 29. — S. 1949–1964.
5. Joint Chiefs of Staff. Cyberspace operations. Joint Chiefs of Staff (US); 19 2022 Dec 19. Joint Publication No.: JP 3-12. // Official Website of the Joint Chiefs of Staff: [sajt]. — URL: https://www.jcs.mil/Doctrine/Joint-Doctrine-Pubs/3-0-Operations-Series/ (data obrashheniya: 26.08.2023).
6. Lockheed Martin's Cyber-Kill Chain. // Leading Aerospace and Defense | Lockheed Martin: [sajt]. — URL: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html (data obrashheniya: 26.08.2023).
7. Telen`ga A.P. Analiz vozmozhnostej obnaruzheniya metastruktur informacionny`x sistem special`nogo naznacheniya. / A. P. Telen`ga // Kiberbezopasnost`: ugrozy`, tendencii, texnologii zashhity`: materialy` II Mezhvedomstvennoj nauch.-praktich. konf., 19-20 maya 2022 g. / Krasnodarskoe vy`sshee voennoe ordenov Zhukova i Oktyabr`skoj Revolyucii Krasnoznamennoe uchilishhe imeni generala armii S.M.Shtemenko. – Krasnodar: KVVU, 2022. C.70-76
8. Zheng, Y. Dynamic defenses in cyber security: Techniques, methods and challenges / Y. Zheng, Z. Li, X. Xu // Digital Communications and Networks. — 2022. — № 8. — S. 422–435.
9. Shaping Attacker Behavior: Evaluation of an Enhanced Cyber Maneuver Framework / J. A. McKneely, T. K. Sell, K. A. Straub [i dr.] // HCII 2022: HCI for Cybersecurity, Privacy and Trust. — Cham: Springer, 2022. — S. 358–379. – DOI: 10.1007/978-3-031-05563-8_23
10. Lilli, E. How Can We Know What We Think We Know about Cyber Operations? / E. Lilli // Journal of Global Security Studies. — 2023. — № 8(2). — S. 1–18. – DOI 10.1093/jogss/ogad011
11. Dynamic-chaos information technologies for data transmission, storage, and protection / Yu. V. Gulyaev, R. V. Belyaev, G. M. Vorontsov [et al.] // Radioe`lektronika. Nanosistemy`. Informacionny`e texnologii. – 2018. – Vol. 10, No. 2. – P. 279-312. – DOI 10.17725/rensit.2018.10.279.
12. Chetvertakova, Yu. S. Postroenie modelej stoxasticheskix nelinejny`x dinamicheskix sistem na osnove dvuxe`tapnoj procedury` parametricheskoj identifikacii / Yu. S. Chetvertakova, O. S. Chernikova // Nauka. Texnologii. Innovacii: Sbornik nauchny`x trudov. V 9-ti chastyax, Novosibirsk, 30 noyabrya – 04 2020 goda / Pod redakciej A.V. Gadyukinoj. Tom Chast` 2. – Novosibirsk: Novosibirskij gosudarstvenny`j texnicheskij universitet, 2020. – S. 94-98.
13. Karganov, V. V. K voprosu o neobxodimosti modelirovaniya informacionny`x sistem organizacii, funkcioniruyushhix v usloviyax ugroz bezopasnosti / V. V. Karganov // Nacional`naya bezopasnost` Rossii: aktual`ny`e aspekty` : Sbornik statej Vserossijskoj nauchno-prakticheskoj konferencii, Sankt-Peterburg, 30 iyulya 2019 goda. – Sankt-Peterburg: Chastnoe nauchno-obrazovatel`noe uchrezhdenie dopolnitel`nogo professional`nogo obrazovaniya Gumanitarny`j nacional`ny`j issledovatel`skij institut «NACzRAZVITIE», 2019. – S. 20-30.
14. Budko, N. P. Obshhie principy` funkcionirovaniya i trebovaniya k postroeniyu struktur perspektivny`x sistem monitoringa raspredelenny`x informacionno-telekommunikacionny`x setej / N. P. Budko // Texnika sredstv svyazi. – 2021. – № 2(154). – S. 38-59.
15. Golovanov, A. A. Sravnitel`ny`j analiz sistem obnaruzhenij anomalij c ispol`zovaniem potoka setevogo trafika i protokola NETFLOW / A. A. Golovanov, O. I. Mel`nikova // Mezhdunarodny`j nauchno-issledovatel`skij zhurnal. – 2023. – № 6(132). – DOI 10.23670/IRJ.2023.132.18.
16. Poltavceva, M. A. Formirovanie struktur danny`x v zadachax aktivnogo monitoringa bezopasnosti // Problemy` informacionnoj bezopasnosti. Komp`yuterny`e sistemy`. – 2021. – № 1. – S. 9-19.
17. Alothman, B., Raw Network Traffic Data Preprocessing and Preparation for Automatic Analysis // 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). – 2019. – P. 1-5, DOI: 10.1109/CyberSecPODS.2019.8885333.
18. Chernigovskij, A. V. Ocenka stepeni samopodobiya setevogo trafika / A. V. Chernigovskij, M. V. Krivov // Informacionny`e texnologii. Problemy` i resheniya. – 2019. – № 1(6). – S. 115-120.
19. Chen, L. A deep multi-task representation learning method for time series classification and retrieval / L. Chen, D. Chen, F. Yang, J. Sun // Information Sciences. – 2021. – Vol. 555. – P. 17-32. – DOI 10.1016/j.ins.2020.12.062.
20. Kalinin, M. Yu. E`ntropijny`e ocenki reshayushhix statistik algoritma klassifikacii sluchajny`x processov / M. Yu. Kalinin, O. N. Choporov // Modelirovanie, optimizaciya i informacionny`e texnologii. – 2020. – T. 8, № 4(31). – DOI: 10.26102/2310-6018/2020.31.4.034.
21. Sokolovskij, S. P. Metodika formirovaniya lozhnogo setevogo trafika informacionny`x sistem dlya zashhity` ot setevoj razvedki / S. P. Sokolovskij, A. P. Telen`ga // Vestnik komp`yuterny`x i informacionny`x texnologij. – 2022. – T. 19, № 2(212). – S. 40-47. – DOI: 10.14489/vkit.2022.02.pp.040-047.

УМНАЯ БОТ-СЕТЬ ИЛИ МОДЕЛЬ ИНТЕЛЛЕКТУАЛЬНОГО ДЕСТРУКТОРА

Рыженко А.А.[1]

Целью работы *является разработка модели интеллектуального деструктора бот-сети, содержащей автономные и полуавтономные ресурсы.*

Метод исследования: *методы мультимножества, концептуальное моделирование, алгоритмизация процессов.*

Результат исследования: *разработана модель формирования правил перехода состояний интеллектуальных деструкторов единой сети как автономного элемента и как части единой сети одновременно. Особенностью модели является адаптивность к внешним возмущениям за счет использования агентной модели методологии системы систем и семантики связей между ними с использованием единого неразрушимого ядра базы правил и множественного выбора древовидной иерархии поля решений баз ассоциаций. Правила продукционного типа представлены в упрощенной алгебраической форме по аналогии с современными алгоритмами построения цифровой подписи (организация зоны доверия с открытыми ключами). Полученная постановка решает такую проблему, как появление естественным образом отшельников и изгой в виде однозадачных автонов, что являлось одной из ключевых проблем полиморфных деструкторов.*

Научная новизна *заключается в разработке нового элемента концептуального моделирования деструкторов моделей – атрибутивного процесса, позволяющего адаптивно изменять правила перехода состояний.*

Ключевые слова: *деструктор, моделирование, интеллектуальный агент, фасет, иерархия, правила перехода, автоном, поле решений, полиморфик.*

DOI:10.21681/2311-3456-5-60-68

Введение

Многолетний анализ сводных отчетов и примеров утечек информации и атак на информационные ресурсы всемирной сети Интернет из новостных лент (дайджестов) показывает, что организуемые и используемые для деструктивного воздействия бот-сети вполне организованны и управляемы многоуровневой иерархической распределенной сетью с одним аналитическим центром (ядро системы) [1]. Общая модель управляемого деструктора аналогична модели конструктора информационного обмена, используемого поисковыми системами, но есть ряд особенностей, о чем и будет рассмотрено в статье. Раннее в некоторых публикациях поднималась данная тематика, но целевой модели так и не было обнаружено [2].

Стоит сразу отметить, что данная технология не нова, развитие продолжается десятилетиями, о чем свидетельствует неофициальная статистика одного из направлений социальной инженерии [3]. Отслеживаемые заказные атаки второго и третьего уровня (организованные средними группами и уровнем экспертов взломщиков) начиная с 90-х годов прошлого века показывают, насколько организованная сеть атакующих способна достаточно долго обрабатывать заказанного клиента любого уровня защищенности для достижения итоговой цели (обрушение, кража, подлог информации и т. п.). Данная особенность сформировала новое направление взлома – однозадачные ловушки деструкторов. Первое десятилетие 2000-х годов показало, что благодаря данным разработкам организованность атакующих начала плавно переходить от групповых атак с привлечением взломщиков к атакам с использованием автономных и полуавтономных сетей ловушек или бот-сетей. Данная тенденция породила новое направление развития интеллектуальных ботов деструкторов, способных не только автономно выполнять простые задачи на уровне вируса, но и также быть частью общей сети в нужный момент времени. Здесь уместно вспомнить алгоритм одного

1 Рыженко Алексей Алексеевич, кандидат технических наук, доцент, доцент департамента информационной безопасности, Финансовый университет при Правительстве РФ, г. Москва, Россия. E-mail: AARyzhenko@fa.ru

из самых эффективных деструкторов – алгоритм вируса полиморфик [4]. Многие публикации отметили, что данный алгоритм на примерах показывает, как в автономной сети в полуавтономном режиме может происходить эволюция программного кода, что дает основу для интеллектуальных систем деструкторов [5].

Официальная статистика выявленных попыток расставить боты-ловушки в сети Интернет со встроенным однозадачным алгоритмом для последующего деструктивного действия показали, насколько данная идея способна жить, развиваться и показывать результаты [6]. Примеры дайджестов, описывающих заказные атаки, основанные на исторических событиях разных государств (выборы государственных деятелей, реорганизация ключевых игроков глобального рынка, обрушение валюты государства или другого аналогичного рынка и т.д.) показывают эффективность интеллектуальных агентов деструкторов. Также часто используются более простые атаки с использованием бот-сетей для банковских систем или других финансовых структур [7]. Отмечено, что все чаще используют программируемые бот-сети для массового информационного вброса в электронные средства информации [8]. Развитие мобильных систем связи и коммуникаций, а также систем с открытым кодом и *app*-приложений открыло второе дыхание для систем долгосрочного управления сетью деструкторов [9] и т.д.

Систематизация полученной информации позволила заложить и описать модель бот-сети интеллектуального деструктора. Рассмотрим разработанную модель более подробно.

1. Обобщенное описание разрабатываемой модели бот-сети деструктора

Классическая теория управления включает множество компонентов взаимосвязанных между собой различных систем поддержки управления [10]. В основе связи элементов модели закладываются различные формы правил и ограничений. При переходе системы от простого уровня к комплексному устанавливается дополнительное условие: для дальнейшего развития системы помимо *конструктора* (порождающего новые объекты, процессы и потоки данных) должен использоваться *деструктор* ресурсов (уничтожающий временные структуры, коллизии, неиспользуемые архивы данных и т.д.). Для стабильного функционирования развивающейся системы ключевые задачи деструктора должны выполняться в полном объеме, иначе система начинает съедать себя изнутри, что приводит к эффекту самоуничтожения [11].

С другой стороны, информационные системы не всегда выполняют это требование, что связанно с ограничениями ресурсов в основном технической составляющей, т.е. деструктор также будет требовать ресурсы, а взять из негде. Например, самые популярные операционные системы внедрили первые деструкторы только в 2007 году при переходе на кроссплатформенные технологии (например, интеллектуальная корзина). Данное событие напрямую связано с развитием технической составляющей. Аналогичная история и со всемирной сетью. Переход на оптоволоконную сеть и массовое развитие спутниковой (и сотовой) связи способствовало развитию деструктивной сети в целом. Как следствие, современные информационные потоки также способствовали развитию не только конструктивной составляющей, но и деструктивной. Самым популярным примером являются организованные массовые *DDoS*-атаки [12].

Стоить отметить развивающуюся тенденцию существующих систем защиты информации – необходимость моделирования не столько самой системы защиты, сколько модели деструктора, что в дальнейшем позволяет комплексным системам защиты информации динамично принимать своевременные решения при атаках различного типа. Ранее были представлены некоторые элементы модели в публикациях [13]. Дальнейшее развитие системы позволило синтезировать полученную информации и представить в виде единой модели. Разработанная модель базируется на трех теоретических составляющих:

- бикубическая матрица «*атакуемые ресурсы – методы атаки*» – используется матричный подход организации данных, где одна из осей каждой матрицы смежная, а вторая – содержит ключевые параметры для построения узловых точек поля решений;
- модифицированные правила перехода состояний системы – позволяют формировать простые правила базы ассоциаций, позволяющие переходить между узловыми точками по определенным правилам;
- динамическое дерево последовательности атак на ресурсы – временно формируемый иерархические структуры, позволяющие устанавливать связи между узловыми точками, а также уровни иерархии принятия решений.

Данные методы используются при построении моделей угроз для объектов информационной инфраструктуры многих организация в том или ином виде [14].

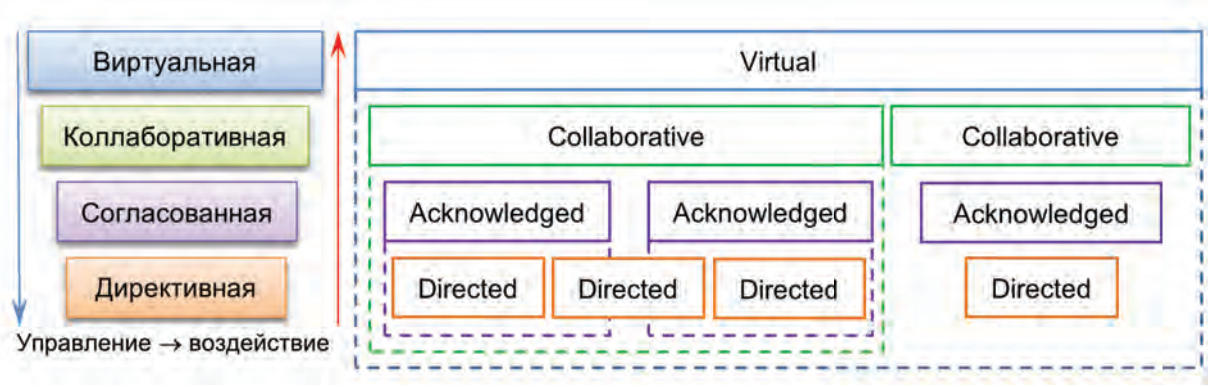

Рис. 1. Варианты иерархии «системы систем»

Приведенные выше составляющие имеют разный формат данных и разные формы построения ассоциативных правил. Анализ возможных форм взаимодействия выявил, что предлагаемый разными моделями онтологий системы взаимодействия не позволяют описывать процессные модели данных, что вызывает ряд вопросов в отношении универсальности предлагаемых разными авторами подходов [15]. Как следствие, для взаимосвязи неформатных данных выбрана для использования методология «*system of systems*» (*SoS*). Описание типов связей или процессов между элементами поля решений основано на следующей классификации *SoS*: виртуальная (*Virtual*), коллаборативная или мягкая (*Collaborative*), согласованная или смешанная (*Acknowledged*) и директивная или жесткая (*Directed*) [16]. Жесткие и смешанные системы позволяют собой условно управлять, что (опять же условно) может декларировать как более низкий (промежуточный) подкласс (рис. 1). Ранее в публикациях более подробно были рассмотрены данные подклассы[2].

Дальнейший анализ информационных источников не позволил выявить формального математического обоснования принципов организации *SoS*. Используемые на практике модели указывают, что необходимо применять дополнительные логические сценарии, позволяющие *одновременно работать и как часть целого и как само целое*, *пытаться «выжить» не управляя фактически ничем и никем*. Данный принцип (из зарубежных источников) получил название – *Russian babushka doll* или *Матрёшка*[3].

Подводя промежуточные итоги, можно сделать следующий вывод: существует множество технологий по заражению информационных ресурсов деструкторами (первый этап построения бот-сети), существуют методологии по организации управления бот-сетью для выполнения одной целевой задачи, также существуют методологии долгосрочного управления динамичной бот-сетью для решения пролонгируемых задач. Но единой теоретической модели, позволяющей не только слепо управлять бот-сетью, но и иметь встроенный инструмент анализа текущего состояния на практике не существует. В результате все чаще появляются такие автономные деструкторы как *изгои* и *отшельники*. Поведение данных категорий непредсказуемое, что доставляет ряд проблем как защищающимся, так и для самих атакующих злоумышленников и разработчиков в одном лице. Дальнейшее изложение материала будет как вариант решения данной задачи.

2. Теоретическая модель

Для формирования итоговой модели интеллектуального агента бот-сети произведена выборка из существующих теоретических подходов. В результате элемент структуры модели агента используется для:

- *матричное представление данных* – кортеж данных системы распределения ресурсов: динамические адреса источников деструкторов + задача автономного деструктора + алгоритм целевых воздействий деструктора. Располагаются в ячейках матрицы с активными границами;
- *иерархичное представление данных* – структурные древовидные алгоритмы последовательностей воздействий (атак) на информационные ресурсы. Привязано к матрице и фасету данных. В бикубической матрице является связующим звеном между кортежем данных и кортежем процессов;
- *сетевая структура данных* – условные алгоритмы возможных переходов состояний при целевых атаках на информационный ресурс. Позволяют анализировать не только общее состояние

2 Systems of Systems (SoS). URL: https://www.sebokwiki.org/wiki/Systems_of_Systems_(SoS) (дата обращения: 15.08.2023)

3 Molly Sharbach, Math enables custom arrangements of liquid “nesting dolls”. URL: https://engineering.princeton.edu/news/2020/11/30/math-enables-custom-arrangements-liquid-nesting-dolls (дата обращения: 15.08.2023)

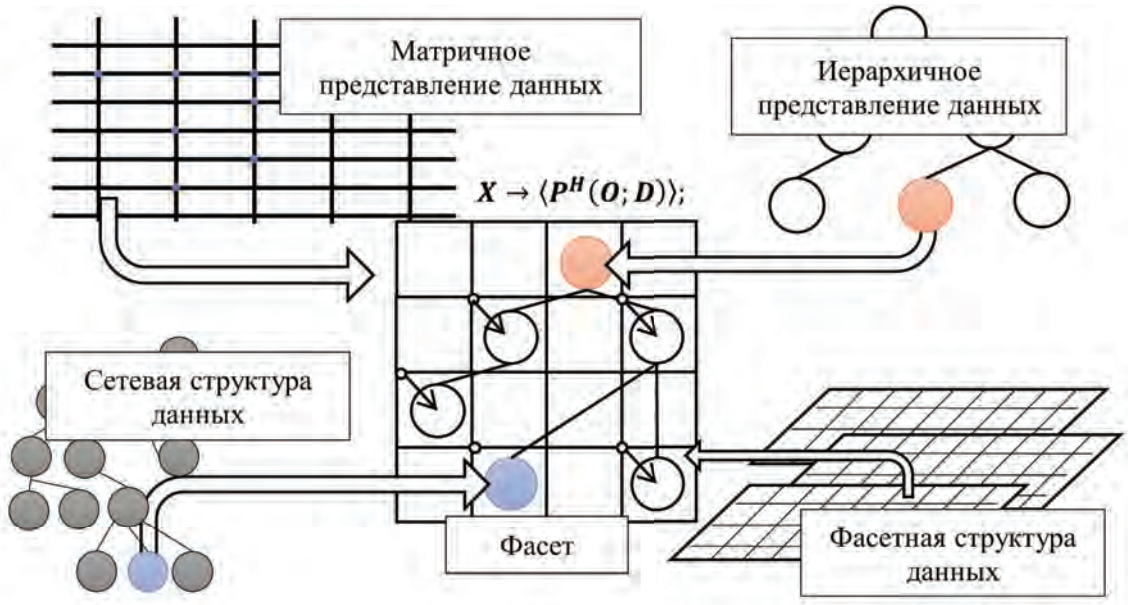

Рис. 2. Схематичное представление симбиоза используемых методологий

системы бот-сети, но и детализировать для каждого задействованного деструктора;

— *фасетная структура данных* – кортеж процессов системы распределения ресурсов: динамические адреса атакуемых ресурсов + исполняемые задачи деструкторов (например, вовремя *DoS*-атаки) + алгоритм изменения последующих ветвей иерархии данных в зависимости от текущего состояния. Используется формальная модель продукционного правила с нефиксированной правой частью.

Общее представление системы синтеза используемых в данной модели методологий представлено на рис. 2. Основной кортеж перехода состояний (1):

$$X \rightarrow P^H\left(O;D\right); \quad (1)$$

где:

X – *experience*, описание состояния системы бот-сети.

P^H – *hierarchical process*, иерархическая процессная составляющая, основанная на фасетной структуре данных.

O – *objects*, объектная составляющая, основанная на матричной структуре данных.

D – [*brain*] *data*, алгоритмическая составляющая – модель основного элемента поля решений интеллектуального агента.

Рассмотрим каждый компонент отдельно:

1. Матричное представление данных (бикубическая матрица «*атакуемые ресурсы – методы атаки*»). Для формирования одной границы смежности используется минимальное дуальное количество матриц, т. е. количество матриц ограничивается нижней границей равной двум. Одна матрица отвечает за динамическое распределение адресов источников деструкторов «*адрес источника – деструктор*», вторая – за адресную системы атакуемой цели «*деструктор – адрес цели*» (рис. 3).

Основная матрица перераспределения адресов *автономов* деструкторов располагается в ядре модели в базе правил (БП). Матрицы атакуемых ресурсов располагаются во временно создаваемых базах ассоциаций ($БА_i$). В чем принципиальной отличие комбинации «*база правил – базы комбинаций*» от «*базы знаний*» было подробно рассмотрено в публикациях на сторонних примерах [15]. Как было упомянуто ранее, взаимосвязь между ячейками матрицы происходит с использованием *алгоритмов последовательности атак*, подключается второй компонент – иерархическое дерево.

2. Динамическое дерево последовательности атак на ресурсы (рис. 4). Основные постулаты:

— *не существует заранее предопределенных унифицированных иерархий деревьев*. Достаточным условием является наличие правил переходов и правил разрешения простых коллизий. Эффект *метаморфа* позволяет конструировать иерархию деревьев на основе сетевой структуры за счет продукционных правил перехода состояний;

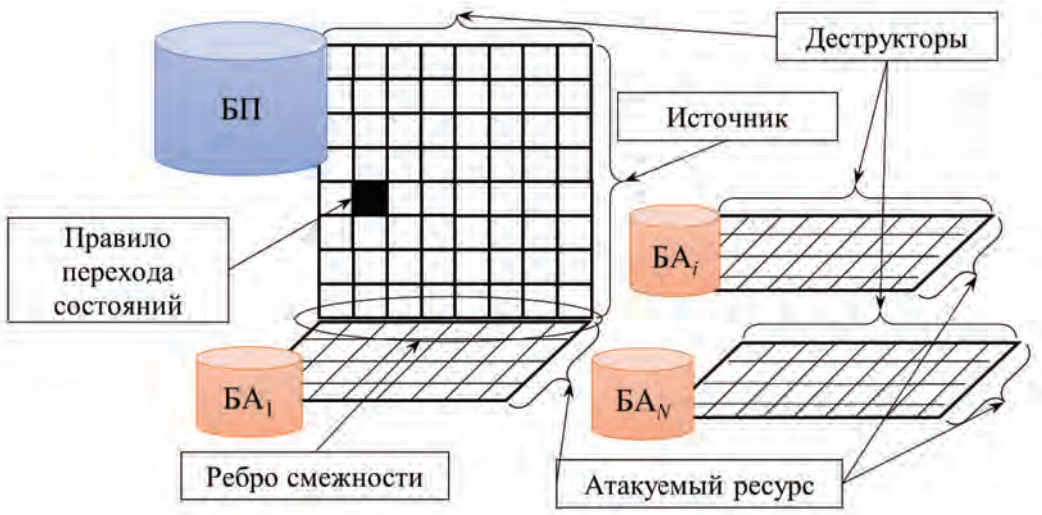

Рис. 3. Схематическое представление бикубической матрицы

— *не существует ограничений на количество уровней иерархии, минимальное количество ограничено базовым условием одноуровневого процесса*, т.е. минимальное количество уровней равно двум, что соответствует простой атаке деструктора;
— *первоначальное количество источников адресов деструкторов ограничивается одним условием – наличием минимум одной альтернативы, либо использованием отшельника в качестве источника*;
— *адрес каждого деструктора изгой заносится в отдельную матрицу без привязок к процессам*. Данная матрица также хранится в базе правил;
— *целевые ресурсы под постоянном мониторингом состояния осуществляются архитектором, анализ состояния не проводится*. Количество атакуемых ресурсов не может превышать количество источников с учетом временных ресурсов;
— *вариативные ресурсы используются для временного перехода в состояние источник до момента активной атаки, затем деструктор самоуничтожается (временный ресурс)* и т. д.

Количество первоначальных правил ровно 13, что соответствует количеству основных аксиом используемой в модели алгебры мультимножеств[4]. На рис. 4 не отражено, но параллельно с прямым деревом иерархии формируется обратное целевое дерево. Точки пересечения двух деревьев в узлах позволяют автономно конструировать набор исходных данных для правил перехода состояний.

3. Модифицированные правила перехода состояний системы. С одной стороны, в данной модели используется классическая форма продукционного правила, где элементами являются: описание класса ситуаций, условие, при котором продукция активизируется, ядро продукции и постусловие продукционного правила. Но, как было упомянуто ранее, в отличии от функционала конструктора, используется деструктивное свойство полиморфика. Ядро продукции не приводит к единому решению, а производит выборку вариантов решений в выделенном диапазоне. Более подробно данную особенность процессных моделей можно изучить в публикации [1]. Например, у узла дерева имеется максимально 5 источников, при этом атакуемых ресурсов всего 4. Количество атак на один ресурс не ограничено. Описать правило перехода ядра продукции для атаки из двух источников на один ресурс кортежа (1) можно следующим образом (2):

$$1+1\underset{5}{\rightarrow}[1;1..4] \qquad (2)$$

Вывод по теоретической части: представлен синтез теоретических подходов при формировании поля решений действий интеллектуального агента. Особенностью является использование модификаций правил продукционного типа в алгебраической форме в базе правил ядра разработанной модели, а также синтез ядра и множеств матриц атакуемых ресурсов баз ассоциаций. Алгебраическая форма представления ядра продукционного правила позволяет интеллектуальной части деструктора более оперативно принимать решения за счет искусственного устранения

4 Мультимножества. – режим доступа: https://life-prog.ru/1_47229_multimnozhestva.html (дата обращения: 15.08.2023)

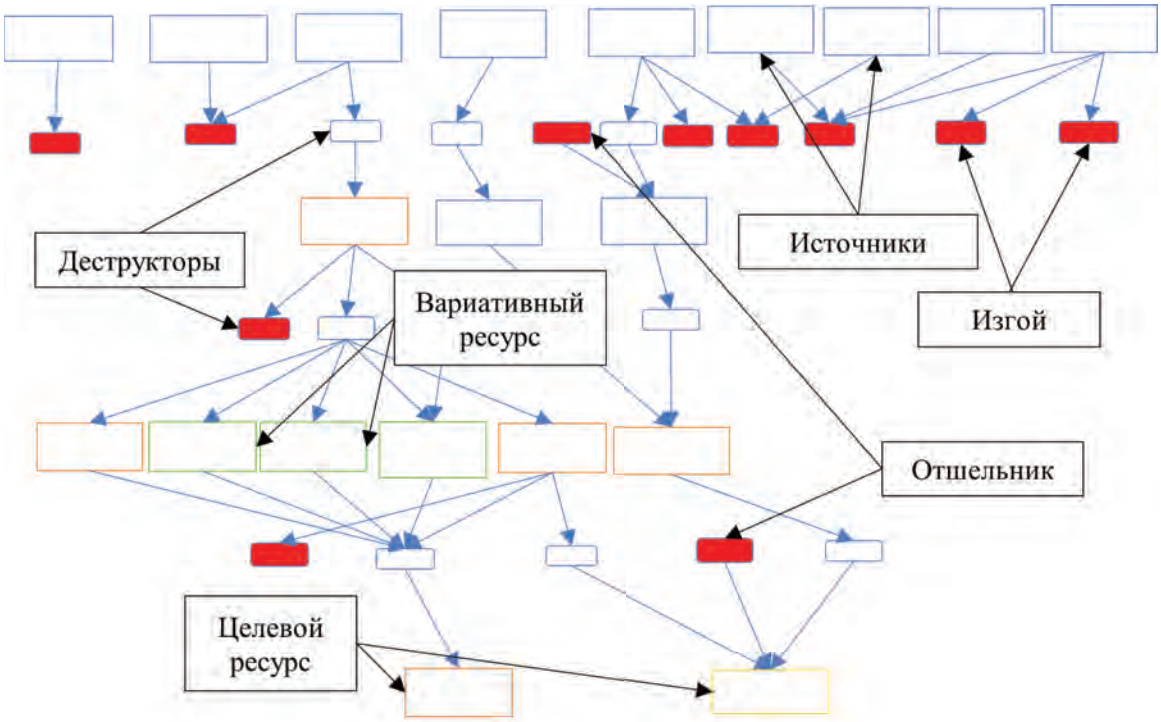

Рис. 4. Пример построения динамического дерева последовательности атак

источников парадоксов и коллизий. Вариативность альтернативных решений позволяет отрабатывать возможные сценарии, что приводит к однозначному решению.

3. Практические наработки

На рис. 5 представлен пример формирования поля решений деструктора с использованием шаблонов базы правил, а также сценариев баз ассоциаций. Для визуализации используем аффинную систему координат, где:

- ось абсцисс – ребро смежности между матрицами источников деструкторов и целевых атакуемых ресурсов;
- ось ординат – ребро смежности между матрицей целей и иерархией алгоритмов атак;
- ось аппликат – ребро смежности между матрицей источников и иерархией алгоритмов атак;

Для индикации состояний узловых точек используется классическая модель светофор: зеленый – успех, желтый – требуется действие, красный – неуспех. Организация мониторинга интеллектуальным деструкторов осуществляется через отклики автономов ботов следующим образом:

- *автоном деструктор с откликом* – после успешной атаки отправляет сигнал хозяину. Например, троян отправляет полученную информацию и меняет состояние индикатора или червь вносит изменения, отправляет запрос трояну и через посредника меняет состояние индикатора;
- *автоном деструктор без отклика* – однозадачный процесс, не отправляет хозяину отклик независимо от текущего состояния. Как правило данные автономы используются для массовых атак на информационные ресурсы и нет необходимости оперативного наблюдения;
- *полуавтоном деструктор с откликом* – деструктор наблюдатель. Данная категория не несет прямых деструктивных действий, но способствует атакам. Например, сканер портов, монитор активных потоков, монитор реестров, монитор файловой активности и т. д.;
- *полуавтоном деструктор без отклика* – деструктор посредник. Фактически выполняет роль поддержки управления для взломщиков. Проводит аудит системы на наличие поисковых механизмов обратного слежения потоков данных через сканеры портов, а также антивирусной активности в разных проявлениях. Работает совместно с предыдущими ботами в роли координатора и контролера. Готовых программных приложений для данной категории не существует. Каждая бот-сеть готовит данную категорию уникально.

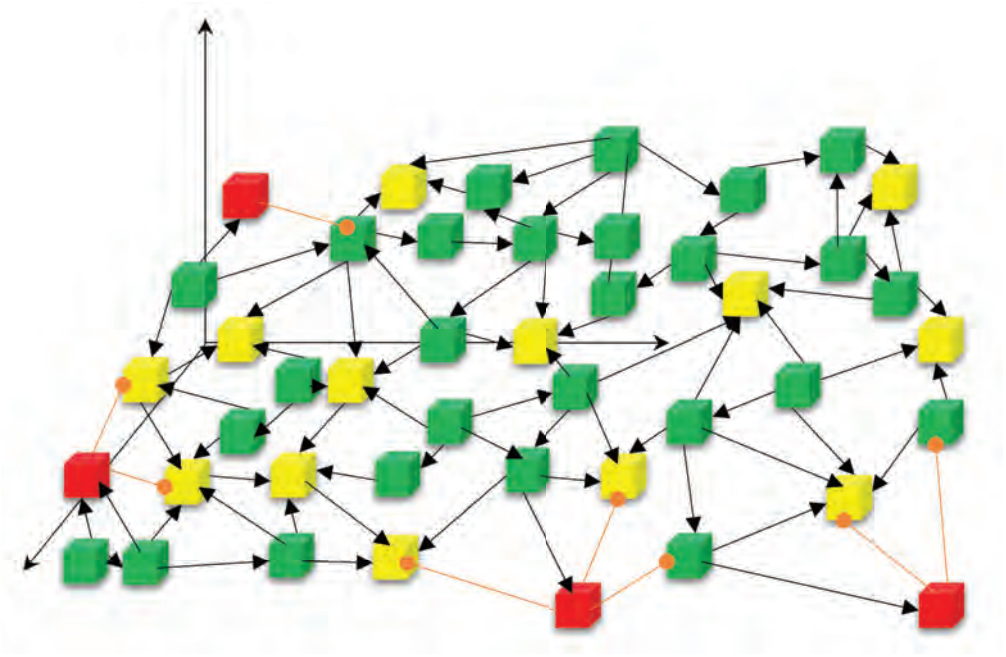

Рис. 5. Пример формирования поля решений деструктора

Вывод перед заключением: особенностью разработанной модели бот-сети интеллектуального деструктора является свойство адаптивности, основанное на полиморфной модели вирусных атак и метаморфной модели построения сценариев последовательных атак. Разработанный механизм позволяет визуализировать деятельность бот-сети в оперативном режиме.

Заключение

Данная статья является очередным звеном цикла статей по организации контура безопасности систем защиты информации корпоративных сетей любого уровня. Ранее были рассмотрены и подробно расшифрованы моменты, связанные с формализацией процессов современных систем деструкторов. Первый подготовительный этап в статье не рассматривается, так как достаточно подробно на примерах разбирается другими авторами в ведущих журналах (в статье приведено несколько ссылок с примерами). Используемые механизмы мультимножеств процессных моделей также рассматривались в многих публикациях, в данной статье приведено только разработанное решение при использовании заложенных ранее теоретических основах.

Приведенный пример поля решений использовался при проектировании контура безопасности в организации, охватывающей систему документооборота филиальной сети нескольких субъектов РФ. Используемые программные продукты (*KAV*, *Dallas*, *ViP Net* и др.) для организации защищенных сегментов подключены к единой системе мониторинга, так как кроссплатформенные системы позволяют подключать файлы отчетов к внешним системам. Прописанный код и результаты мониторинга за выделенный период по атакам и защитам будут представлены в последующих публикациях.

Литература

1. Ryzhenko A. A. Model of facet and hierarchical pyramidal system of support of management of information space of corporation. System analysis in economics – 2018: Proceedings of the V International research and practice conference-biennale (21-23 november 2018). – Moscow, Prometheus publishing house, 2018. – pp. 146-149.
2. Рыженко А. А. Формирование центров адаптации ресурсов как необходимого элемента международного сотрудничества / Большая Евразия: развитие, безопасность, сотрудничество. Ежегодник. – М.: ИНИОН РАН, 2018. – Вып.1, ч.1. – С. 327-328.
3. Julien Duchêne, Colas Le Guernic, Eric Alata, Vincent Nicomette & Mohamed Kaâniche State of the art of network protocol reverse engineering tools. Journal of Computer Virology and Hacking Techniques volume 14, pages53-68 (2018)
4. Razieh Eskandari, Mahdi Shajari & Mojtaba Mostafavi Ghahfarokhi ERES: an extended regular expression signature for polymorphic worm detection. Journal of Computer Virology and Hacking Techniques volume 15, pages177–194 (2019)
5. Hadis Ghanei, Farnoush Manavi & Ali Hamzeh, A novel method for malware detection based on hardware events using deep neural networks. Journal of Computer Virology and Hacking Techniques, volume 17, pages 319-331 (2021)

6. Varshini Reddy, Naimisha Kolli & N. Balakrishnan, Malware detection and classification using community detection and social network analysis. Journal of Computer Virology and Hacking Techniques, volume 17, pages 333-346 (2021)
7. Samanvitha Basole, Fabio Di Troia & Mark Stamp Multifamily malware models. Journal of Computer Virology and Hacking Techniques volume 16, pages79-92 (2020)
8. Mina Ebrahim & Seyed Alireza Hashemi Golpayegani, Anomaly detection in business processes logs using social network analysis. Journal of Computer Virology and Hacking Techniques, volume 18, pages 127-139 (2022)
9. Francesco Mercaldo & Antonella Santone, Audio signal processing for Android malware detection and family identification. Journal of Computer Virology and Hacking Techniques, volume 17, pages 139-152 (2021)
10. Рыженко А. А. Пирамидальная модель распределения информационных ресурсов госкорпораций на фасетно-иерархическом уровневом основании / А.А. Рыженко, Н.Ю. Рыженко // Анализ, моделирование, управление, развитие социально-экономических систем: сборник научных трудов XIII Всероссийской с международным участием школы-симпозиума АМУР-2019, Симферополь-Судак, 14-27 сентября 2019 / ред. совет: А.В. Сигал (предс.) и др. – Симферополь : ИП Корниенко А.А., 2019. – с. 346-353 ISBN 978-5-6042038-4-2
11. Рыженко А. А. Модифицированный алгоритм вируса полиморфика как основа деструктора информационной среды / Информатика: проблемы, методология, технологии: сборник материалов XVIII международной научно-методической конференции: в 7 т. / под редакцией Н.А. Тюкачева; Воронеж, Воронежский государственный университет, 14-15 февраля 2019 г. – Воронеж: Издательство «Научно-исследовательские публикации» (ООО «Вэлборн»), 2019. – Т. 5. – С. 857–861.
12. Jiaxing Cheng, Ying Li, Cheng Huang, Ailing Yu & Tao Zhang ACER: detecting Shadowsocks server based on active probe technology. Journal of Computer Virology and Hacking Techniques volume 16, pages217–227 (2020)
13. Рыженко А. А. Модель деструктора-полиморфа цифровой среды / Проблемы управления безопасностью сложных систем: материалы XXVI Междунар. конфер., 19 декабря 2018 г., Москва / под общ. ред. А. О. Калашникова, В.В. Кульбы. М.: ИПУ РАН, 2018. – с. 158–162.
14. Rizwan Ur Rahman & Deepak Singh Tomar, Threats of price scraping on e-commerce websites: attack model and its detection using neural network. Journal of Computer Virology and Hacking Techniques, volume 17, pages 75-89 (2021)
15. Рыженко А. А. Безопасность информации цифровой экономики / А. А. Рыженко, Н. Ю. Рыженко // Актуальные проблемы и перспективы развития экономики. Труды Юбилейной XX Всероссийской с международным участием научно-практической конференции. Симферополь, 2021. С. 289–291.
16. Systems of Systems Characterization and Types - NATO STO. URL: https://www.sto.nato.int/publications/STO%20Educational%20Notes/STO-EN-SCI-276/EN-SCI-276-01.pdf (дата обращения: 15.08.2023)

SMART BOTNET OR INTELLIGENT DESTRUCTOR MODEL

Ryzhenko A.A.[5]

The aim of the work *is to develop a model of an intelligent botnet destructor containing autonomous and semi-autonomous resources.*

Research method: *multiset methods, conceptual modeling, process algorithmizing.*

Research result: *a model for the formation of rules for the transition of states of intelligent destructors of a single network as an autonomous element and as part of a single network at the same time has been developed. A feature of the model is its adaptability to external disturbances using an agent-based model of the methodology of a system of systems and the semantics of connections between them using a single indestructible core of the rule base and multiple choice of the tree-like hierarchy of the decision field of association bases. Production-type rules are presented in a simplified algebraic form by analogy with modern algorithms for constructing a digital signature (organization of a trust zone with public keys). The resulting statement solves such a problem as the natural appearance of hermits and outcasts in the form of single-tasking autonomous, which was one of the key problems of polymorphic destructors.*

The scientific novelty *lies in the development of a new element of conceptual modeling of model destructors - an attributive process that allows you to adaptively change the rules for the transition of states.*

Keywords: *destructor, modeling, intelligent agent, facet, hierarchy, transition rules, autonomous, decision field, polymorphic.*

5 Aleksey A. Ryzhenko, Ph.D., Associate Professor, Financial University under the Government of the Russian Federation, Moscow, AARyzhenko@fa.ru

References

1. Ryzhenko A.A. Model of facet and hierarchical pyramidal system of support of management of information space of corporation. System analysis in economics – 2018: Proceedings of the V International research and practice conference-biennale (21-23 november 2018). – Moscow, Prometheus publishing house, 2018. – pp. 146-149.
2. Ryzhenko A.A. Formirovanie centrov adaptacii resursov kak neobhodimogo jelementa mezhdunarodnogo sotrudnichestva / Bol'shaja Evrazija: razvitie, bezopasnost', sotrudnichestvo. Ezhegodnik. – M.: INION RAN, 2018. – Vyp.1, ch.1. – S. 327-328.
3. Julien Duchêne, Colas Le Guernic, Eric Alata, Vincent Nicomette & Mohamed Kaâniche State of the art of network protocol reverse engineering tools. Journal of Computer Virology and Hacking Techniques volume 14, pages53-68 (2018)
4. Razieh Eskandari, Mahdi Shajari & Mojtaba Mostafavi Ghahfarokhi ERES: an extended regular expression signature for polymorphic worm detection. Journal of Computer Virology and Hacking Techniques volume 15, pages177–194 (2019)
5. Hadis Ghanei, Farnoush Manavi & Ali Hamzeh, A novel method for malware detection based on hardware events using deep neural networks. Journal of Computer Virology and Hacking Techniques, volume 17, pages 319-331 (2021)
6. Varshini Reddy, Naimisha Kolli & N. Balakrishnan, Malware detection and classification using community detection and social network analysis. Journal of Computer Virology and Hacking Techniques, volume 17, pages 333-346 (2021)
7. Samanvitha Basole, Fabio Di Troia & Mark Stamp Multifamily malware models. Journal of Computer Virology and Hacking Techniques volume 16, pages79-92 (2020)
8. Mina Ebrahim & Seyed Alireza Hashemi Golpayegani, Anomaly detection in business processes logs using social network analysis. Journal of Computer Virology and Hacking Techniques, volume 18, pages 127-139 (2022)
9. Francesco Mercaldo & Antonella Santone, Audio signal processing for Android malware detection and family identification. Journal of Computer Virology and Hacking Techniques, volume 17, pages 139-152 (2021)
10. Ryzhenko A.A. Piramidal'naja model' raspredelenija informacionnyh resursov goskorporacij na fasetno-ierarhicheskom urovnevom osnovanii / A.A. Ryzhenko, N.Ju. Ryzhenko // Analiz, modelirovanie, upravlenie, razvitie social'no-jekonomicheskih sistem: sbornik nauchnyh trudov XIII Vserossijskoj s mezhdunarodnym uchastiem shkoly-simpoziuma AMUR-2019, Simferopol'-Sudak, 14-27 sentjabrja 2019 / red. sovet: A.V. Sigal (preds.) i dr. – Simferopol' : IP Kornienko A.A., 2019. – s. 346-353 ISBN 978-5-6042038-4-2
11. Ryzhenko A.A. Modificirovannyj algoritm virusa polimorfika kak osnova destruktora informacionnoj sredy / Informatika: problemy, metodologija, tehnologii: sbornik materialov XVIII mezhdunarodnoj nauchno-metodicheskoj konferencii: v 7 t. / pod redakciej N.A. Tjukacheva; Voronezh, Voronezhskij gosudarstvennyj universitet, 14-15 fevralja 2019 g. – Voronezh: Izdatel'stvo «Nauchno-issledovatel'skie publikacii» (OOO «Vjelborn»), 2019. – T. 5. – S. 857-861.
12. Jiaxing Cheng, Ying Li, Cheng Huang, Ailing Yu & Tao Zhang ACER: detecting Shadowsocks server based on active probe technology. Journal of Computer Virology and Hacking Techniques volume 16, pages217–227 (2020)
13. Ryzhenko A.A. Model' destruktora-polimorfa cifrovoj sredy / Problemy upravlenija bezopasnost'ju slozhnyh sistem: materialy HHVI Mezhdunar. konfer., 19 dekabrja 2018 g., Moskva / pod obshh. red. A.O. Kalashnikova, V.V. Kul'by. M.: IPU RAN, 2018. – s. 158-162.
14. Rizwan Ur Rahman & Deepak Singh Tomar, Threats of price scraping on e-commerce websites: attack model and its detection using neural network. Journal of Computer Virology and Hacking Techniques, volume 17, pages 75-89 (2021)
15. Ryzhenko A.A. Bezopasnost' informacii cifrovoj jekonomiki / A.A. Ryzhenko, N.Ju. Ryzhenko // Aktual'nye problemy i perspektivy razvitija jekonomiki. Trudy Jubilejnoj XX Vserossijskoj s mezhdunarodnym uchastiem nauchno-prakticheskoj konferencii. Simferopol', 2021. S. 289-291.
16. Systems of Systems Characterization and Types - NATO STO. URL: https://www.sto.nato.int/publications/STO%20Educational%20Notes/STO-EN-SCI-276/EN-SCI-276-01.pdf (data obrashhenija: 15.08.2023)

МЕТОДОЛОГИЯ СБОРА ДАННЫХ ДЛЯ АНАЛИЗА БЕЗОПАСНОСТИ ПРОМЫШЛЕННЫХ КИБЕРФИЗИЧЕСКИХ СИСТЕМ

Котенко И.В. [1], Федорченко Е.В.[2], Новикова Е.С.[3], Саенко И.Б.[4], Данилов А.С.[5]

Цель исследования: *формирование методологии сбора и формирования наборов данных, используемых для разработки и тестирования эффективности подходов к выявлению аномалий и кибератак на основе машинного обучения, в т.ч. с применением моделей глубокого обучения.*

Методы исследования: *методы системного анализа и моделирования, машинное обучение, статистический анализ данных.*

Полученные результаты: *исследованы и систематизированы подходы к формированию обучающих наборов данных, используемых для разработки методов обнаружения аномалий и кибератак. Разработана методология сбора данных для анализа безопасности промышленных киберфизических систем, ключевые этапы проиллюстрированы на примере построения тестового стенда системы водоочистных сооружений, предназначенного для исследования ее защищенности от кибератак.*

Научная новизна: *представленная в работе методология специфицирует последовательность взаимосвязанных этапов, которые определяют действия, начиная от формализации промышленного процесса, заканчивая валидацией полученных данных. Последовательное выполнение этих этапов позволяет создавать наборы данных, которые содержат как сетевые данные, так и показания датчиков и актуаторов киберфизических систем, имеют четкую схему аннотирования и валидированы относительно реальных данных, обрабатываемых в подобных системах.*

Вклад: *Котенко И.В. и Федорченко Е.В. – общая концепция методологии сбора данных для исследования безопасности киберфизических систем; Котенко И.В., Федорченко Е.В. и Новикова Е.С. – проработка этапов методологии; Новикова Е.С. и Федорченко Е.В. – анализ положения дел по созданию обучающих наборов данных для разработки и тестирования аналитических моделей выявления аномалий и кибератак; Данилов А.А. и Саенко И.Б. – формализация флотационного процесса очистки воды и разработка тестового стенда в соответствии с формулированными требованиями к обучающему набору данных.*

Ключевые слова: *кибербезопасность, автоматизированные системы управления, выявление аномалий и кибератак, обучающие наборы, тестовый стенд, системы водоочистных сооружений.*

DOI:10.21681/2311-3456-2023-5-69-79

Введение

Анализ положения дел в области формирования наборов данных, используемых для оценки кибербезопасности промышленных систем показал, что в настоящее время отсутствует единая комплексная методология сбора данных и их валидации для тестирования методик обнаружения аномалий и кибератак на основе машинного обучения.

В настоящее время задача обнаружения аномалий и кибератак в промышленных киберфизических системах хорошо исследована [1, 2], и в научной литерату-

1 Котенко Игорь Витальевич, заслуженный деятель науки РФ, доктор технических наук, профессор, главный научный сотрудник и руководитель лаборатории проблем компьютерной безопасности, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: ivkote@comsec.spb.ru

2 Федорченко (Дойникова) Елена Владимировна, кандидат технических наук, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: doynikova@comsec.spb.ru

3 Новикова Евгения Сергеевна, кандидат технических наук, доцент, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: novikova@comsec.spb.ru

4 Саенко Игорь Борисович, доктор технических наук, профессор, ведущий научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: ibsaen@comsec.spb.ru

5 Данилов Александр Сергеевич, кандидат технических наук, доцент кафедры геоэкологии, Санкт-Петербургский Горный университет, старший научный сотрудник ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН) г. Санкт-Петербург, Россия. E-mail: aleksandrsdanilov@gmail.com

ре представлено большое число различных подходов к ее решению. Предложены методы на основе статистического анализа данных, анализа временных рядов [3, 4], на основе классического машинного обучения [5], и в последнее время активно исследуется применимость методов глубокого обучения [6-8]. Например, в [6] предложен подход к обнаружению и прогнозированию аномалий в потоках данных от промышленных систем с помощью рекуррентной сверточной нейронной сети с блоками долгой краткосрочной памяти (Long Short Term Memory, LSTM). В [7] был выполнен анализ различных архитектур глубоких нейронных сетей для обнаружения аномалий на примере реальных данных от датчиков промышленного лифта. Проведенные эксперименты показали, что эффективность обнаружения аномалий на несбалансированных данных достигает 0,899% для случая, когда в обучающем наборе содержится 3% аномальных данных, и 0,984% для случая, когда в тестовом наборе данных содержится 20% аномальных данных. В [8] представлена нейронная сеть с механизмом внимания, она состоит из энкодера, который кодирует входные многомерные данные временного ряда с помощью многослойных связанных модулей внимания, и двух декодеров, выполняющих операции по прогнозированию и реконструкции временных рядов. Исчерпывающий обзор подходов глубокого обучения к обнаружению атаки аномалий в киберфизических системах (КФС) можно найти в [9, 10].

Однако эффективное применение методов глубокого обучения для обнаружения аномалий связано с выполнением двух практических условий: наличие значительных вычислительных ресурсов и наличие больших объемов хорошо подготовленных данных. Под хорошо подготовленными данными подразумевается данные большого объема, которые имеют достоверную разметку. Разметка включает информацию об аномалиях или об их отсутствии.

В работе [11] авторы проанализировали наиболее часто используемые наборы данных, применяемые в исследованиях по обнаружению аномалий в киберфизических системах, и сформулировали основные требования к наборам данным, которые могут быть использованы в исследованиях кибербезопасности таких систем:

- набор данных должен включать данные от физических и цифровых компонентов киберфизической системы, т.е. включать данные сетевого трафика и журналы датчиков и актуаторов;
- набор данных должен иметь разметку, схема аннотации должна включать метки «норма» и/или «аномалия», аномалии и атаки должны быть описаны;
- набор данных должен быть максимально приближен к реальным данным, как с точки зрения моделируемых технологических процессов, так и выполняемых кибератак на систему.

Для исследования безопасности киберфизических систем и формирования соответствующих наборов данных, могут быть использованы следующие типы тестовых стендов [12]: виртуальные, аппаратные и гибридные.

Виртуальные стенды используют только методы программного моделирования и аппаратной эмуляции для моделирования работы промышленных устройств и сетевого взаимодействия между ними. Очевидным преимуществом таких стендов является низкая стоимость их разработки. Однако моделирование сложных процессов представляет собой нетривиальную задачу, и в результате программные имитаторы технологических процессов могут быть менее точными и надежными, чем их физические реализации. Пример такого стенда – виртуальный стенд, описанный в [13]. В ней моделируется небольшая электрическая сеть, состоящая из одной основной питающей ветви и трех подветвей – A, B и C. Данные от системы автоматизированного контроля и сбора данных (Supervisory Control and Data Acquisition, SCADA-системы) создаются с помощью специализированной “песочницы” SCADA-системы, кибератаки выполняются в реальности и имитируют кибератаки, совершенные на энергосистему Украины в декабре 2015 года.

Физические стенды разрабатываются с использованием реальных аппаратных и программных средств, применяемых в промышленных системах. Генерируемые с их помощью данные отличаются реалистичностью, кроме того, уязвимости конкретных устройств могут быть использованы для реализации кибератак. Примером такого стенда является стенд водоочистных сооружений SWaT [14].

Гибридные стенды представляют собой комбинацию программно-эмулируемых компонентов и физических устройств. Такой подход является компромиссным решением между дорогостоящими физическими стендами и более дешевыми, но иногда недостаточно реалистичными виртуальными стендами. Примером такого стенда является стенд HAI [15], в котором моделируются четыре различных процесса, три из которых реализуются с помощью программно-аппаратных средств, а четвертый является полностью программной моделью.

УДК 004.056.57

Разработка тестового промышленного стенда и, соответственно, формирование набора данных является сложной практической задачей [12, 16] в связи с наличием следующих проблем и ограничений:

- отсутствием единых рекомендаций по проектированию стендов площадок и формированию наборов данных;
- необходимостью реализацией реальных сценариев как на технологическом уровне, включающем программные и аппаратные средства, так и на уровне атак, заключающемся в моделировании атак, характерных для заданного промышленного процесса/системы;
- сложностью современных технологических процессов, что обусловливает необходимость участия в разработке стенда как специалистов в области промышленной автоматизации, в области моделируемых технологических процессов, так и в области кибербезопасности;
- масштабируемостью тестового стенда, поскольку, как правило, тестовые площадки моделируют некоторую упрощенную версию технологического процесса;
- сбором данных, включающем сбор как нормальных, так и аномальных данных;
- стоимостью в случае физических стендов;
- физической безопасностью в случае физических стендов, поскольку не все технологические процессы могут быть безопасно смоделированы в лабораториях в уменьшенном варианте;
- отсутствием документации;
- возможностью воспроизвести разработанные стенды, которая практически невозможна для физических стендов и очень ограничена для гибридных стендов.

В данной работе решается проблема, связанная с отсутствием единых рекомендаций по проектированию. Таким образом, новизна статьи и вклад авторов заключается в представленной методологии создания наборов данных, применимых для исследований в области кибербезопасности промышленных систем. В качестве примера авторы представляют макет стенда, разработанного для анализа защищенности системы управления водоочистными сооружениями.

Методология формирования набора данных

Предлагаемая методология создания наборов данных основывается на выполнении следующих этапов:

- определение технологического процесса;
- разработка соответствующего тестового стенда;
- формирование набора данных, соответствующих нормальному функционированию системы;
- разработка модели атак на рассматриваемый технологический процесс;
- разработка сценариев атакующих действий с учетом технологического стека, используемого для развертывания тестовой площадки;
- реализация атаки и сбор массива данных для атакуемой системы;
- валидация набора данных.

Рассмотрим данные этапы более подробно.

Определение технологического процесса. Данный этап предполагает определение технологического процесса и набора параметров, которые будут собираться во время экспериментов.

Технологический процесс может быть представлен в виде технологической схемы, которая описывает выполняемую последовательность технологических операций, в т.ч. графически в виде мнемосхем. На этом этапе определяются, какие параметры значимы для безопасного и эффективного выполнения заданного технологического процесса, а какими можно пренебречь. На их основе формируется перечень датчиков и актуаторов, необходимых для построения тестовой системы, и данные от этих устройств будут описывать технологический процесс на физическом уровне.

На этом этапе также определяется математическая модель процесса, если планируется использовать виртуальный или гибридный тестовый стенд. Формализованная модель технологического процесса также позволит определить возможные последствия атакующих действий с учетом связей между датчиками и исполнительными устройствами.

Создание тестового стенда. На этом же этапе определяется тип тестового стенда: гибридный, виртуальный или физический. В зависимости от типа стенда прорабатываются детали его реализации: определяется программное обеспечение для моделирования и эмуляции оборудования в случае виртуальной и гибридной тестовой площадки, выбираются сенсоры, датчики и соответствующее программное обеспечение для их подключения в случае физического стенда. Определяются протоколы взаимодействия между устройствами, а также механизмы автоматизированного сбора данных и управления. Специфицируется формат собираемых данных и интервал их получения, исходя их технических характеристик используемого программного и аппаратного обеспечения.

Другой важной задачей является выявление ключевых отличий технологических процессов, реализуе-

мых на тестовом стенде, от реальной системы: в частности, определяется, как влияет масштабирование и упрощение системы на ее функционирование, какие критические моменты необходимо учитывать при разработке аналитических моделей безопасности, предназначенных для реальных систем.

На этом этапе готовится документация по моделируемому технологическому процессу и разработанной экспериментальной площадке, она включает схему технологического процесса, детали аппаратной и программной реализации, доступные источники данных и их формат. Для виртуального стенда подготавливаются рекомендации по его воспроизводимости.

Формирование набора данных, соответствующих нормальному функционированию системы. Этот этап заключается в сборе данных с датчиков и сетей за определенный промежуток времени. Интервал времени определяется специалистом предметной области с учетом переходного периода, необходимого для достижения системой нормального рабочего состояния. Сбор данных должен осуществляться с использованием инструментов и механизмов, определенных на предыдущем этапе. Подготавливается документация, касающаяся продолжительности функционирования стенда, его особенностей, например, переходного периода.

Разработка модели атак на рассматриваемый технологический процесс. Этот этап включает в себя определение модели атакующего, его возможностей. Таким образом, необходимо определить объем ресурсов, доступных злоумышленникам, и их осведомленность об атакуемой системе.

Кибератаки могут осуществляться как на сетевом, так и на физическом уровне и могут включать в себя различные типы атак. Однако они должны коррелировать с выбранным технологическим процессом и быть максимально похожими на реальные случаи. К наиболее распространенным сетевым атакам относятся атаки сбора данных (разведки), атака «человек посередине» (MitM), атака с вбросом ложных данных или команд (False Data Injection), атака повторного воспроизведения (Replay Attack) и атака “отказ в обслуживании” (DoS-атака). Атаки на физическом уровне направлены на физические устройства и имеют целью изменение их показаний (Device Manumission Attack) или их физическое повреждение (Direct Damage Attack). Учитывая многообразие возможных атакующих действий, рекомендуется сформировать упрощенную формальную модель, связывающую скомпрометированные датчики и ожидаемый результат атаки в виде изменений в функционировании системы. Такая формализация полезна при оценке эффективности атак. Информация о модели злоумышленника и атаках должна быть включена в документацию.

Разработка сценариев атакующих действий с учетом технологического стека экспериментального стенда. Технологический стек, включающий протоколы сетевого взаимодействия, систему SCADA и настройки контроля доступа на рабочей станции оператора, определяет выбор средств и методов атаки.

При использовании SCADA-систем рекомендуется проводить атаки непосредственно на инфраструктуру тестового стенда, что обеспечивает максимальную точность определения времени атаки и реакции системы на нее [13].

На этом этапе также необходимо определить, каким образом будет фиксироваться процесс выполнения атаки для получения размеченных данных, а также определен формат протокола фиксации атакующих действий. Обычно этот процесс реализуется вручную. В протокол необходимо включить следующую информацию об атаке: точка входа атаки (IP-адрес), цели атаки (атакующий IP-адрес, датчик), тип атаки, времени начала и окончания атаки, используемые программного-аппаратные средства, наблюдаемые изменения в работе системы. Для автоматизированного разбора протокола атак, рекомендуется хранить их в формате JSON[6].

Реализация атак и сбор массива данных для атакуемой системы. Данный этап заключается в проведении кибератак, фиксации их реализации и записи результатов. Частота выполняемых атакующих действий должна быть определена с учетом времени реакции системы на воздействие, поскольку это позволит избежать наложения реакции системы на разные виды атак, возникновению аномальных периодов функционирования системы, не связанных с выполнением атакующих действий.

Наблюдаемые изменения в работе стенда могут быть получены путем анализа данных с сервера-историка SCADA. Подготовленная на этом этапе документация включает заполненные протоколы атак.

Валидация набора данных по имеющимся реальным данным. В настоящее время не существует устоявшихся процедур проверки и валидации сгенерированных наборов данных. Однако проверка набора данных может быть проведена через проверку выбранного тестового стенда, т.е. на основе оценки

6 https://javaee.github.io/tutorial/jsonp001.html

Рис. 1. Основные этапы методологии сбора данных для анализа защищенности киберфизических систем

соответствия реализованной модели технологического процесса и реального процесса. В случае доступа к реальным данных от подобных систем валидация набора данных может быть осуществлена путем статистического анализа данных на основе оценки корреляции между параметрами искусственного и реального наборов данных, и относительной энтропии между двумя наборам.

На рис. 1 представлены этапы предложенной методологии и промежуточные результаты каждого этапа.

В следующем разделе представлена реализация двух этапов предлагаемой методологии.

Разработка тестового стенда для моделирования процессов очистки воды

Определение технологического процесса. В качестве примера технологического процесса очистки воды был выбран процесс флотации. Данный процесс является одним из наиболее современных и безопасных методов очистки сточных вод и относится к механическим процессам очистки воды. Для него характерна высокая способность удалять жировые отложения из воды, что исключает необходимость устранения засоров в трубах на выходе очищенной воды. Он недорог в эксплуатации и надежен, поскольку все элементы флотационной установки представляют собой простые механизмы. Флотационный процесс также обладает высокой скоростью очистки воды от органических загрязнений (по сравнению с отстаиванием воды) и эффективно снижает количество болезнетворных бактерий, ПАВ и легко окисляемых веществ и микропластика [17]. Благодаря этим факторам данный процесс водоочистки часто используется в составе городских систем водоподготовки.

Процесс флотационной очистки воды состоит из двух основных этапов: 1) процесс флокуляции; 2) про-

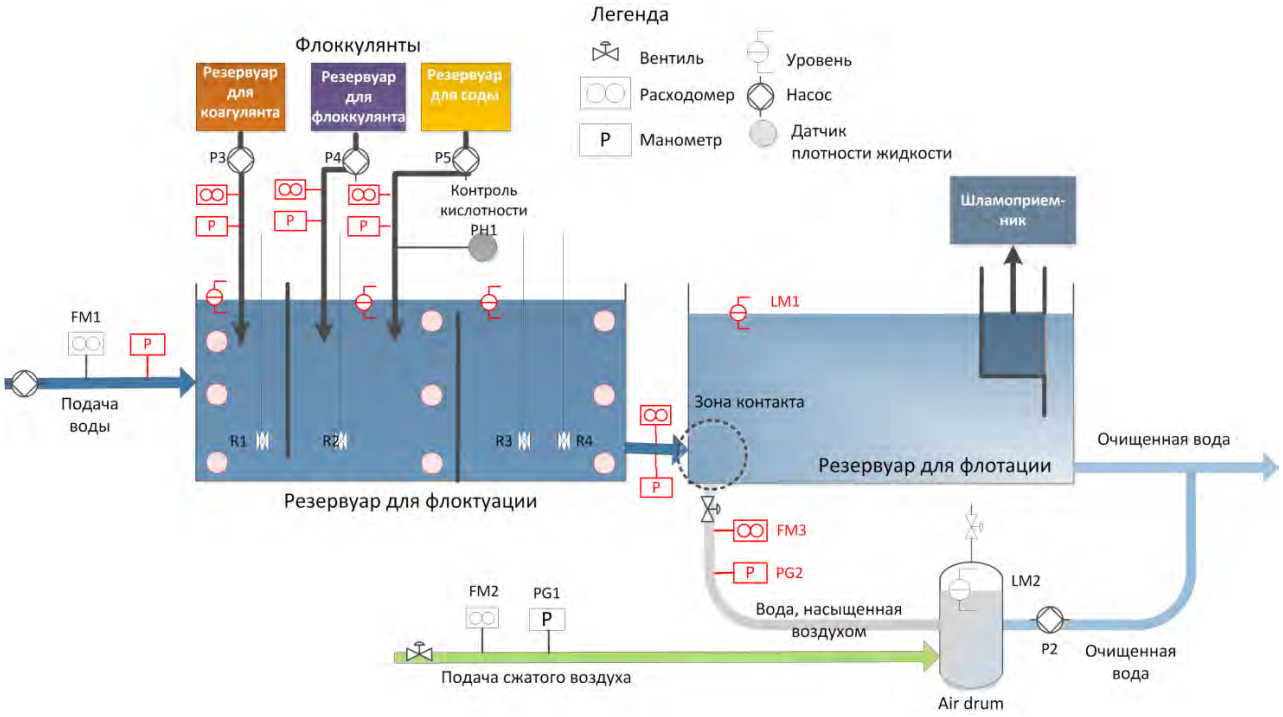

Рис. 2. Технологическая схема процесса флотации

цесс флотации. Упрощенная схема технологического процесса представлена на рис. 2.

Флокуляция является предварительным этапом флотации. Она заключается в объединении тонкодисперсных твердых частиц в более крупные агломераты. Такие агломераты более эффективно удаляются из воды, чем мелкие частицы. Процесс флокуляции происходит в трехкамерном резервуаре для флокуляции (рис. 2) и включает несколько подпроцессов: коагуляцию, непосредственно флокуляцию и созревание флокулы и регулирование уровня кислотности (pH).

Процесс коагуляции инициируется коагулянтами, объем добавляемых реагентов контролируется насосами для их подачи, четыре погружные мешалки перемешивает воду с коагулянтами. В первой камере резервуара формируются микрофлокулы. Во второй камере подается флокулянт и сода, которые перемешиваются с водой и микрофлокулами, постоянно увеличивающимися в размерах. Необходимая эффективная концентрация коагулянта – флокулянта и соды – зависит от соотношения скорости подачи воды, определяемой насосом P1, и скорости дозирования коагулянтов, которые контролируются насосами-дозаторами (P3, P4, P5). Затем суспензия с крупными флокулами поступает в третью камеру резервуара для флокуляции. В этой камере мешалки R3 и R4 препятствуют осаждению флокулы, постоянно перемешивая воду. Скорость их вращения должна быть несколько ниже скорости вращения мешалок в первых двух камерах, чтобы предотвратить разрушение флокулы. Коагулянт имеет низкое значение кислотности, что может снизить эффективность флокуляции. По этим причинам необходимо контролировать уровень pH в емкости для флокуляции.

Из резервуара для флокуляции грязная вода через зону контакта поступает в резервуар для флотации. Смесь воды и пузырьков воздуха подается в резервуар для флотации через три отверстия в зоне контакта. Образование мелких воздушных пузырьков является важной составляющей процесса флотации. Мелкие пузырьки получаются за счет снижения давления: воздух, растворенный в воде под высоким давлением, переходит в газообразное состояние за счет быстрой декомпрессии. Для образования пузырьков используется чистая вода, получаемая на выходе, и которая подается обратно в систему под давлением с помощью насоса P2.

Время контакта флокулы с воздухом зависит от соотношения скоростей подачи загрязненной и очищенной воды. Пузырьки воздуха обволакивают флокулу, в результате плотность агломератов, связанных с воздухом, становится меньше плотности воды во флотационном резервуаре, и они всплывают на поверхность воды. Грязная пена с поверхности резервуара

для флотации удаляется скребком в шламоприемник, а очищенная вода выходит через правый нижний угол резервуара.

Математическое описание этого процесса достаточно сложно, оно требует рассмотрения гидродинамических процессов с большим числом управляющих, и контролируемых переменных с учетом физико-химического взаимодействия [18], поэтому в данном случае было решено разрабатывать физический стенд.

Создание экспериментального стенда. Основой для экспериментального стенда послужил учебный стенд CE 587 (G.U.N.T., GmbH, Германия). Он оснащен следующими датчиками:

- подача воды на установку - датчики используются в реальной системе водоочистки для контроля объема воды: расходомер воды FM1, устанавливаемый после насоса P1, расходомер давления внутри насоса P2, расходомер перед входным резервуаром (после насосов P2 и P6);
- коагуляция — расходомер воды FM1 перед резервуаром флокуляции;
- флокуляция - нет датчиков;
- созревание флокул - расходомер воздуха FM2 и манометр PG1 перед воздушным барабаном, измеритель уровня воды LM2 устройства генерации пузырьков, измеритель давления внутри насоса P2;
- контроль уровня кислотности pH: датчик кислотности PH1.

На рис. 2 эти датчики отмечены черным цветом.

Учебный стенд CE 587 имеет весьма ограниченные возможности контроля и управления технологическим процессом. Блок управления стенда достаточно прост, он позволяет включать/выключать насосы P1, P2, P3, P4 и P5, мешалки R1-R4 (см. рис. 2), задавать параметры скорости вращения мешалок, однако настройки управления всех задаются непосредственно на устройства. Централизованный модуль управления процессом и модуль сбора данных отсутствуют. С целью автоматизации процесса управления и сбора информации было предложено: 1) расширить набор датчиков; 2) дополнить учебный стенд SCADA-системой, осуществляющей управление компонентами стенда и сбор данных.

Для дальнейшего анализа данных авторы предлагают расширить набор датчиков следующими датчиками:

- подача воды на установку - датчик уровня, измеритель давления после водохранилища;
- коагуляция — датчики плотности жидкости на разных уровнях резервуара для флокуляции в первой камере (верхний, нижний, средний), расходомер и манометр после насоса P3, расходомер в первой камере резервуара для флокуляции, датчик уровня;
- флокуляция — датчики плотности на разных уровнях резервуара для флокуляции во второй камере (сверху, снизу, посередине), расходомер и измеритель давления после насоса P4, расходомер во второй камере резервуара для флокуляции;
- созревание флокул — уровнемер, измеритель давления в резервуаре для флокуляции, расходомер и измеритель давления перед резервуаром для флокуляции;
- контроль pH — датчики плотности на разных уровнях резервуара для флотации в третьей камере (верхний, нижний, средний), расходомер и измеритель давления после насоса P5, расходомер в третьей камере резервуара для флокуляции.

На рис. 2 новые датчики выделены красным цветом.

Расширение тестового стенда включает в себя не только добавление новых датчиков, но и компонентов, которые объединяют их в единую систему: нормализаторов сигналов для аналоговых датчиков, программируемого логического контроллер (ПЛК) и всех необходимых модулей его расширения для подключения датчиков и актуаторов, OPC сервера (Open Platform Communication, OPC), базы данных архив и пульта управления оператора.

Программируемый логический контроллер – это микропроцессорные электронные устройства реального времени, которые соединяют между собой датчики и исполнительные механизмы, реализуют логику технологического процесса, передают данные и получают команды от оператора процесса. Датчики и исполнительные устройства могут подключаться непосредственно к цифровым или аналоговым входам/выходам ПЛК, обычно ПЛК поддерживает различные модули расширения, позволяющие подключать различные типы устройств и использовать различные протоколы связи, такие как RS-232, CAN, Modbus и Industrial Ethernet. В качестве ПЛК может выступать отладочная плата Arduino Mega 2560 R3, однако для приближения тестового стенда к реальным системам управления водоочистными сооружениями будет использован промышленный контроллер.

OPC-технология – это набор принятых во всем мире спецификаций, обеспечивающих универсальный механизм обмена данными в промышленных

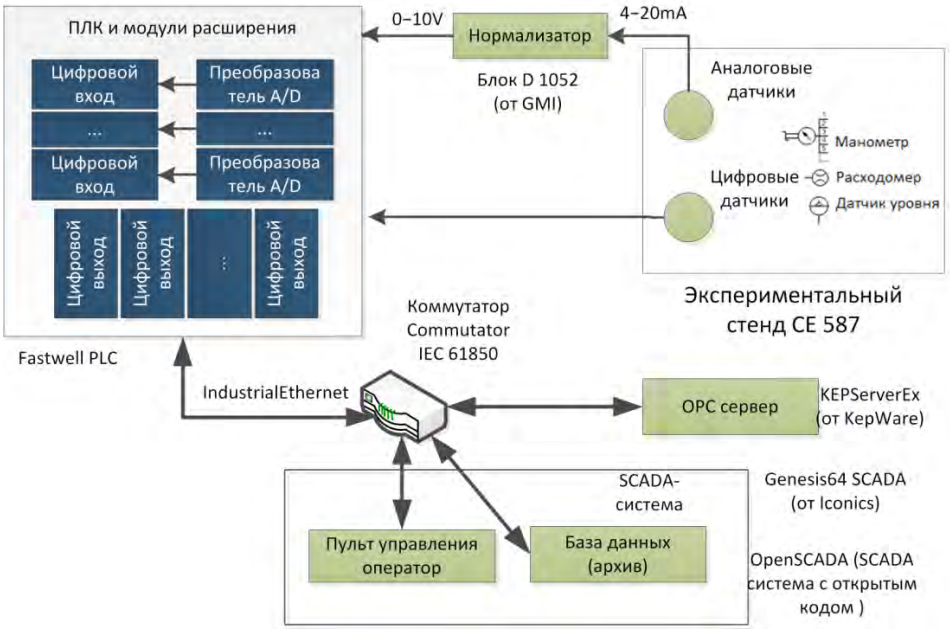

Рис. 3. Программно-аппаратное расширение учебного стенда CE 587 (G.U.N.T., GmbH, Германия)

системах управления, OPC-сервер – это программа, которая получает данные во внутреннем формате устройства или системы и преобразует их в формат OPC. Таким образом, OPC-сервер – это своего рода универсальный драйвер физического оборудования, который обеспечивает взаимодействие с любым OPC-клиентом, причем любые изменения в программных решениях на уровне OPC-клиентов не приводят к изменениям в контролируемом оборудовании.

Для организации хранения данных и мониторинга состояния тестового стенда естественным решением является использование SCADA-системы, которая собирает, обрабатывает и хранит данные, поступающие от ПЛК; предоставляет текущую и архивную информацию в удобной для оператора форме (мнемосхемы, графики, тренды, журналы сообщений); предоставляет утилиты для ввода команд оператора и передачи их в ПЛК; поддерживает отчетность по результатам технологического процесса. На рис. 3 показано предлагаемое расширение флотационного стенда CE 587 с соответствующими программными и аппаратными решениями.

Заключение

В последнее время для обнаружения атак и аномалий в промышленных киберфизических системах было предложено большое число методов, в основе которых лежат модели машинного обучения, включая глубокие нейронные сети. Однако их применение требует наличия реалистичных размеченных обучающих наборов данных. Формирование подобных наборов данных представляет собой сложную практическую задачу, и одной из причин, обуславливающих ее сложность, является отсутствие единых рекомендаций по их созданию.

В данной статье авторы решают эту проблему и предлагают методологию создания размеченных наборов данных для исследования кибербезопасности промышленных систем. Для иллюстрации разработанной методологии подробно рассмотрены первые два этапа - проектирование и создание тестового стенда водоочистных сооружений.

Будущие работы включают реализацию дальнейших этапов методики и создание самого набора данных.

Благодарность. *Работа выполнена при поддержке гранта Российского научного фонда № 23-11-20024, https://rscf.ru/project/23-11-20024/, и Санкт-Петербургского научного фонда.*

Рецензент: *Лаута Олег Сергеевич, доктор технических наук, профессор кафедры комплексного обеспечения информационной безопасности Государственного университета морского и речного флота имени адмирала С.О. Макарова, Санкт-Петербург, Россия.*

E-mail: laos-82@yandex.ru

Литература

1. Котенко И.В., Ушаков И.А. Технологии больших данных для мониторинга компьютерной безопасности // Защита информации. Инсайд, № 3, 2017. С. 23-33.
2. Котенко И.В., Левшун Д.С., Чечулин А.А., Ушаков И.А., Красов А.В. Комплексный подход к обеспечению безопасности киберфизических систем на системе микроконтроллеров // Вопросы кибербезопасности. 2018. № 3 (27). С. 29-38. DOI: 10.21681/2311-3456-2018-3-29-38.
3. Котенко И.В., Саенко И.Б., Лаута О.С., Крибель А.М. Метод раннего обнаружения кибератак на основе интеграции фрактального анализа и статистических методов // Первая миля. 2021. № 6 (98). С. 64-71. DOI: 10.22184/2070-8963.2021.98.6.64.70.
4. Котенко В.И., Саенко И.Б., Коцыняк М.А., Лаута О.С. Оценка киберустойчивости компьютерных сетей на основе моделирования кибератак методом преобразования стохастических сетей // Труды СПИИРАН. 2017. № 6(55). С. 160-184. DOI: 10.15622/sp.55.7.
5. Branitskiy A., Kotenko I., Saenko I. Applying Machine Learning and Parallel Data Processing for Attack Detection in IoT // IEEE Transactions on Emerging Topics in Computing, 2021, vol. 9, no. 4, pp. 1642-1653. DOI: 10.1109/TETC.2020.3006351.
6. Wu Z., Guo Y., Lin W., Yu S., Ji Y. A weighted deep representation learning model for imbalanced fault diagnosis in cyber-physical systems // Sensors, vol. 18, no. 4, 2018, 1096. DOI: 10.3390/s18041096.
7. Canizo M., Triguero I., Conde A., Onieva E. Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study // Neurocomputing, vol. 363, pp. 246-260, 2019, pp. 246-260. DOI: 10.1016/j.neucom.2019.07.034.
8. Xia F., Chen X., Yu S., Hou M., Liu M., and You L. Coupled attention networks for multivariate time series anomaly detection // Arxiv. 2023. URL: https://arxiv.org/pdf/2306.07114.pdf (дата обращения: 29.08.2023).
9. Luo Y., Xiao Y., Cheng L., Peng G., Yao D. D., Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities // ACM Comput. Surv., 2021, vol. 54, no. 5, Article 106, 36 p. DOI: 10.1145/3453155.
10. Kotenko I., Gaifulina D., Zelichenok I. Systematic Literature Review of Security Event Correlation Methods // IEEE Access, 2022, vol. 10, pp. 43387-43420. DOI: 10.1109/ACCESS.2022.3168976.
11. Tushkanova O., Levshun D., Branitskiy A., Fedorchenko E., Novikova E., Kotenko I. Detection of cyber attacks and anomalies in cyber-physical systems: approaches, data sources, evaluation // Algorithms, vol. 16, no. 2, 2023, 85. DOI: 10.3390/a16020085.
12. Conti M., Donadel D., Turrin F. A survey on industrial control system testbeds and datasets for security research // IEEE Communications Surveys & Tutorials, vol. 23, no. 4, pp. 2248–2294, 2021.
13. Lemay A., Fernandez J. M. Providing SCADA network data sets for intrusion detection research // 9th Workshop on Cyber Security Experimentation and Test (CSET 16). Austin, TX: USENIX Association, Aug. 2016.
14. Goh J., Adepu S., Junejo K. N., Mathur A. A dataset to support research in the design of secure water •treatment systems // Critical Information Infrastructures Security. Cham: Springer International Publishing, 2017, pp. 88–99. DOI: 10.1007/978-3-319-71368-7_8.
15. Shin H.-K., Lee W., Yun J.-H., Kim H. HAI 1.0: HIL-based augmented ICS security dataset // Proceedings of the 13th USENIX Conference on Cyber Security Experimentation and Test, 2020, pp. 1–1.
16. Dominguez M., Fuertes J.J., Prada M.A., Alonso S., Moran A., Perez D. Design of platforms for experimentation in industrial cybersecurity // Applied Sciences, vol. 12, no. 13, 2022, 6520. DOI: 10.3390/app12136520.
17. Kyzas G.Z., Matis K.A. Flotation in water and wastewater treatment // Processes, vol. 6, no. 8, 2018, 116. DOI: 10.3390/pr6080116.
18. Антонова Е.С. Моделирование процесса очистки сточных вод во флотационной установке с эжекционной системой аэрации с диспергатором // Безопасность в техносфере. 2017. №. 1. С. 43-50. DOI: 10.12737/article590199b9952dc2.23575176 (дата обращения: 29.08.2023).

DATA COLLECTION METHODOLOGY FOR SECURITY ANALYSIS OF INDUSTRIAL CYBER-PHYSICAL SYSTEMS

Kotenko I.V.[7], Fedorchenko E.V.[8], Novikova E.S.[9], Saenko I.D.[10], Danilov A.S.[11]

7 Igor V. Kotenko, Honored Worker of Science of the Russian Federation, Dr.Sc., Professor, Chief Scientist and Head of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: ivkote@comsec.spb.ru

8 Elena V. Fedorchenko, Ph.D., Senior researcher of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: doynikova@comsec.spb.ru

9 Evgenia S. Novikova, Ph.D., Associate Professor, Senior researcher of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: novikova@comsec.spb.ru

10 Igor B. Saenko, Dr.Sc., Professor, Leading researcher of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: ibsaen@comsec.spb.ru

11 Aleksandr S. Danilov, Ph.D., Associate Professor of Geoecology department at St. Petersburg Mining University, Senior researcher of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: aleksandrsdanilov@gmail.com

The purpose of the study: *formation of methodology for collecting and generating datasets used to develop and test the effectiveness of anomaly and cyber attack detection approaches based on machine learning, including deep learning models.*

Research methods: *methods of system analysis and modeling, machine learning, statistical data analysis.*

Results obtained: *approaches to the formation of training data sets used for the development of anomaly and cyber attack detection methods were investigated and systematized. The methodology of data collection for analyzing the security of industrial cyber-physical systems is developed, the key stages are illustrated on the example of building a test bench of a water treatment plant system designed to study its security against cyber attacks.*

Scientific novelty: *The analysis of the state of arts in the field of forming datasets used to assess the cyber security of industrial systems has shown that there is currently no unified methodology for data collection and validation for testing anomaly and cyber attack detection techniques based on machine learning. The methodology presented in this paper specifies a sequence of interrelated steps that define actions ranging from the formalization of the industrial process to the validation of the acquired data. The sequential execution of these steps will allow the creation of datasets that contain both network data and readings from sensors and actuators of the cyber-physical system, have a clear annotation scheme and are validated against real data from similar systems.*

Contribution: *Igor Kotenko and Elena Fedorchenko - general concept of data collection methodology for cyber-physical systems security research; Igor Kotenko, Elena Fedorchenko and Evgenia Novikova - elaboration of methodology stages; Evgenia Novikova and Elena Fedorchenko - analysis of the state of affairs on the creation of training data sets for the development and testing of analytical models of anomaly and cyber attack detection; Aleksandr Danilov and Igor Saenko - formalization of the flotation process of water treatment development and development of a test bench in accordance with the formulated requirements for the training data set.*

Keywords: *cyber security, automated control systems, anomaly and cyber attack detection, training sets, test bed, water treatment facilities.*

References

1. Kotenko I.V., Ushakov I.A. [Big data technologies for monitoring computer security] Технологии больших данных для мониторинга компьютерной безопасности. Information security. Inside [Защита информации. Инсайд], No. 3, 2017. pp. 23-33.
2. Kotenko I.V., Levshun D.S., Chechulin A.A., Ushakov I.A., Krasov A.V. [Integrated approach to provide security of cyber-physical systems based on microcontrollers] Комплексный подход к обеспечению безопасности киберфизических систем на основе микроконтроллеров. Cybersecurity issues [Вопросы кибербезопасности]. 2018. No 3 (27). pp.29-38. DOI: 10.21681/2311-3456-2018-3-29-38.
3. Kotenko I., Saenko I., Lauta O., Kribel. [A method for early detection of cyberattacks based on the integration of fractal analysis and statistical methods] Метод раннего обнаружения кибератак на основе интеграции фрактального анализа и статистических методов. Pervaya milya [Первая миля]. 2021. № 6 (98). pp. 64-71. DOI: 10.22184/2070-8963.2021.98.6.64.70
4. Kotenko V.I., Saenko I.B., Kotsynyak M.A., Lauta O.S. [Assessment of Cyber-Resilience of Computer Networks based on Simulation of Cyber Attacks by the Stochastic Networks Conversion Method] Оценка киберустойчивости компьютерных сетей на основе моделирования кибератак методом преобразования стохастических сетей. SPIIRAS Proceedings [Труды СПИИРАН]. 2017. No 6(55). pp.160-184. DOI: https://doi.org/10.15622/sp.55.7.
5. Branitskiy A., Kotenko I., Saenko I. Applying Machine Learning and Parallel Data Processing for Attack Detection in IoT // IEEE Transactions on Emerging Topics in Computing, 2021, vol. 9, no. 4, pp. 1642-1653. DOI: 10.1109/TETC.2020.3006351.
6. Wu Z., Guo Y., Lin W., Yu S., Ji Y. A weighted deep representation learning model for imbalanced fault diagnosis in cyber-physical systems. Sensors, vol. 18, no. 4, 2018, 1096. DOI: 10.3390/s18041096.
7. Canizo M., Triguero I., Conde A., Onieva E. Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study. Neurocomputing, vol. 363, pp. 246–260, 2019, pp. 246-260. DOI: 10.1016/j.neucom.2019.07.034.
8. Xia F., Chen X., Yu S., Hou M., Liu M., You L. Coupled attention networks for multivariate time series anomaly detection. Arxiv. 2023. URL: https: arxiv.org/pdf/2306.07114.pdf (accessed on: 29.08.2023).
9. Luo Y., Xiao Y., Cheng L., Peng G., Yao D.D., Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities. ACM Comput. Surv., 2021. vol. 54, no. 5, Article 106, 36 p. DOI: 10.1145/3453155.
10. Kotenko I, Gaifulina D., Zelichenok I. Systematic Literature Review of Security Event Correlation Methods. IEEE Access, 2022, vol. 10, pp. 43387-43420. DOI: 10.1109/ACCESS.2022.3168976.
11. Tushkanova O., Levshun D., Branitskiy A., Fedorchenko E., Novikova E., Kotenko I. Detection of cyber attacks and anomalies in cyber-physical systems: approaches, data sources, evaluation. Algorithms, vol. 16, no. 2, 2023, 85. DOI: 10.3390/a16020085.
12. Conti M., Donadel D., Turrin F. A survey on industrial control system testbeds and datasets for security research. IEEE Communications

Surveys & Tutorials, vol. 23, no. 4, pp. 2248-2294, 2021.
13. Lemay A., Fernandez J.M. Providing SCADA network data sets for intrusion detection research. 9th Workshop on Cyber Security Experimentation and Test (CSET 16). Austin, TX: USENIX Association, Aug. 2016.
14. Goh J., Adepu S., Junejo K.N., Mathur A.A dataset to support research in the design of secure water treatment systems. Critical Information Infrastructures Security. Cham: Springer International Publishing, 2017, pp. 88-99. DOI: 10.1007/978-3-319-71368-7_8.
15. Shin H.-K., Lee W., Yun J.-H., Kim H. HAI 1.0: HIL-based augmented ICS security dataset. Proceedings of the 13th USENIX Conference on Cyber Security Experimentation and Test, 2020, pp. 1–1.
16. Dominguez M., Fuertes J.J., Prada M.A., Alonso S., Moran A., Perez D. Design of platforms for experimentation in industrial cybersecurity. Applied Sciences, vol. 12, no. 13, 2022, 6520. DOI: 10.3390/app12136520.
17. Kyzas G.Z., Matis K.A. Flotation in water and wastewater treatment. Processes, vol. 6, no. 8, 2018, 116. DOI: 10.3390/pr6080116.
18. Antonova E.S. [Modeling of wastewater treatment process in a flotation plant with an induction aeration system with dispersant] Моделирование процесса очистки сточных вод во флотационной установке с эжекционной системой аэрации с диспергатором. Bezopsnost' v technosphere [Безопасность в техносфере]. 2021. № 6 (98). p. 64-71. DOI: 10.22184/2070-8963.2021.98.6.64.70

МЕТОД ГЕНЕРАЦИИ СЕМАНТИЧЕСКИ КОРРЕКТНОГО КОДА ДЛЯ ФАЗЗИНГ-ТЕСТИРОВАНИЯ ИНТЕРПРЕТАТОРОВ JAVASCRIPT

Козачок А.В.[1], Спирин А.А.[2], Ерохина Н.С.[3]

Цель работы: *разработка метода генерации входных данных для фаззинг-тестирования интерпретаторов JavaScript и его оценка.*

Метод исследования: *изучение закономерностей генерации данных и процента покрытия кода с целью его повышения. Предложенный метод позволяет генерировать входные данные для выявления большего количества уязвимостей при последующем фаззинг-тестировании, за счет повышения процента покрытия кода.*

Результаты исследования: *интерпретатор JavaScript является наиболее уязвимым блоком архитектуры веб-браузера, как следствие возникает необходимость постоянного наращивания объемов анализа/тестирования его исходного кода. Фаззинг-тестирование интерпретатора веб-браузера на основе сложноструктурированных входных данных, например, программного кода JavaScript, является актуальной задачей. В работе приведены уязвимости современных веб-браузеров, а также ключевые проблемы, возникающие при тестировании интерпретаторов JavaScript. Наиболее существенными проблемами являются: отсутствие общедоступных синтаксически и семантически корректных входных данных для фаззинг-тестирования, проблема преодоления внутренних механизмов фильтрации входных данных, выбор рационального алгоритма мутации данных, а также проблема повышения степени покрытия тестируемого кода. Авторами предложен метод генерации входных данных для фаззинг-тестирования интерпретаторов JavaScript, который позволяет повысить качество и скорость фаззинг-тестирования.*

Научная и практическая значимость *результатов исследования заключаются в разработке нового метода генерации входных данных для фаззинг-тестирования интерпретаторов JavaScript веб-браузеров, на основе применения нейросетевых языковых моделей, повышающий покрытие исходного кода.*

Ключевые слова: *веб-браузер, интерпретатор JavaScript, покрытие кода, уязвимости программного обеспечения, информационная безопасность.*

DOI: 10.21681/2311-3456-2023-5-80-88

Введение

В 2023 году согласно отчёту аналитического агентства Meltwater[4] в мире насчитывается 5,16 миллиарда пользователей сети Интернет, что составляет 64,4% мирового населения. По сравнению с 2022 годом количество интернет-пользователей выросло на 1,9%. 92,3% и 65,6% пользователей сети интернет используют мобильные устройства и персональные компьютеры и планшеты соответственно. На каждом из этих устройств работает веб-браузер или аналогичная программа, способная обрабатывать и отображать контент веб-сайтов. Веб-браузеры совершенствуются и становятся все более сложными, осуществляя обработку не только открытого текста и HTML, но и изображений, видео и других форматов данных.

Наибольшую угрозу безопасности веб-браузера представляют интерпретаторы JavaScript (англ. JavaScript engines). Каждый интерпретатор подобен языковому модулю, который позволяет приложению поддерживать определенное подмножество стандартов языка JavaScript. Развитие технологий приводит к постоянному усложнению структуры интерпретаторов JavaScript и увеличению их исходного кода. Данный

4 https://www.meltwater.com/en/global-digital-trends.

1 Козачок Александр Васильевич, доктор технических наук., доцент, Академия ФСО России, г. Орел, Россия, E-mail: a.kozachok@academ.msk.rsnet.ru, https://orcid.org/0000-0002-6501-2008

2 Спирин Андрей Андреевич, кандидат технических наук, Академия ФСО России, г. Орел, Россия, E-mail: spirin_aa@bk.ru, https://orcid.org/ 0000-0002-7231-5728

3 Ерохина Наталья Сергеевна, сотрудник, Академия ФСО России, г. Орел, Россия, E-mail: ens@secdev.space, https://orcid.org/0000-0002-4878-0865

факт негативно влияет на безопасность, что, в свою очередь, активизирует деятельность авторов вредоносных программ.

В последнее время большинство обнаруживаемых ошибок в программном обеспечении (ПО), связанных с удаленным выполнением кода и повышением привилегий, обнаруживаются при помощи фаззинг-тестирования [1]. Вследствие имеющихся ограничений фаззеров, осуществляющих тестирование интерпретаторов JavaScript веб-браузеров, фаззинг-тестирование может быть недостаточно эффективным. Одним из путей повышения эффективности данного процесса является совместное использование алгоритмов машинного обучения и анализа покрытия кода при тестировании с целью преодоления ключевых проблем существующих методов фаззинг-тестирования.

Безопасность интерпретаторов JavaScript веб-браузеров

Среди многих компонентов веб-браузеров интерпретаторы JavaScript представляет особый интерес для злоумышленников, поскольку их полная по Тьюрингу природа позволяет злоумышленникам создавать сложный код, содержащий уязвимости. В частности, интерпретаторы JavaScript оказались в центре внимания исследователей безопасности по разным причинам: во-первых, из соображений производительности они часто реализуются на небезопасных для памяти языках, что влечет за собой уязвимости, приводящие к повреждению памяти. Во-вторых, продолжающаяся гонка за производительностью и постоянное усложнение структуры увеличивает вероятность ошибок при разработке.

Задача интерпретатора JavaScript – анализировать и выполнять код JavaScript. В отличие от большинства других сред, интерпретатор JavaScript, встроенный в веб-браузер, должен безопасно обрабатывать ненадежные сценарии. Кроме того, он разработан с большим акцентом на производительность, чтобы обеспечить интерактивность клиентским веб-приложениям. Как это часто бывает, повышение производительности связано с увеличением сложности кода, что, в свою очередь, приводит к ошибкам программирования, которые иногда являются критическими с точки зрения безопасности. Согласно Национальной базе данных уязвимостей (NVD[5]), 43% всех уязвимостей, обнаруженных в веб-браузерах Microsoft Edge и Google Chrome, были уязвимостями интерпретатора JavaScript [2].

5 https://nvd.nist.gov/.

Хотя дизайн и реализация каждого интерпретатора JavaScript сильно различаются, все они имеют общую архитектуру и два общих свойства: во-первых, они служат стандартизированной средой выполнения для JavaScript кода; во-вторых, обеспечивают JIT-компиляцию для повышения производительности.

В то время как «классические» уязвимости, такие как переполнение буфера или использование динамической памяти после освобождения, редко встречаются в механизмах сценариев, их заменили сложные и специфичные для предметной области уязвимости.

Наиболее часто встречающиеся проблемы в обработчиках сценариев, обнаруженных за последние годы:

- ошибки, связанные с целочисленным переполнением, обычно приводящие к несанкционированному доступу к буферу памяти;
- ошибки из-за неожиданных обратных вызовов при реализации некоторых встроенных функций;
- ошибки использования после освобождения из-за того, что сборщик мусора не находит объект на этапе маркировки;
- уязвимости, возникающие из-за того, что внутренний объект или функция интерпретатора «проникают» в код приложения из-за какой-либо логической проблемы;
- уязвимости, возникающие из-за неправильной оптимизации JIT-компилятора. На сегодняшний день уязвимости были обнаружены, по крайней мере, в реализации устранения проверки границ, анализа выхода и исключении проверки типов [3].

В дополнение к описанным классам уязвимостей существует большое количество различных уязвимостей, которые в настоящее время не могут быть однозначно отнесены к какой-либо категории.

Существующие проблемы при тестировании интерпретаторов JavaScript

Фаззинг – методика тестирования, при которой на вход программы подаются невалидные, непредусмотренные или случайные данные, которые могут привести ее к аварийному завершению или неопределенному поведению. Этот метод автоматического тестирования охватывает множество граничных, а также ложных значений и использует их в качестве входных данных тестируемой программы [4].

Существующие ограничения методов фаззинг-тестирования приводят к тому, что в настоящее время этот процесс недостаточно эффективен. Фаззинг ПО со сложноструктурированными входными данными,

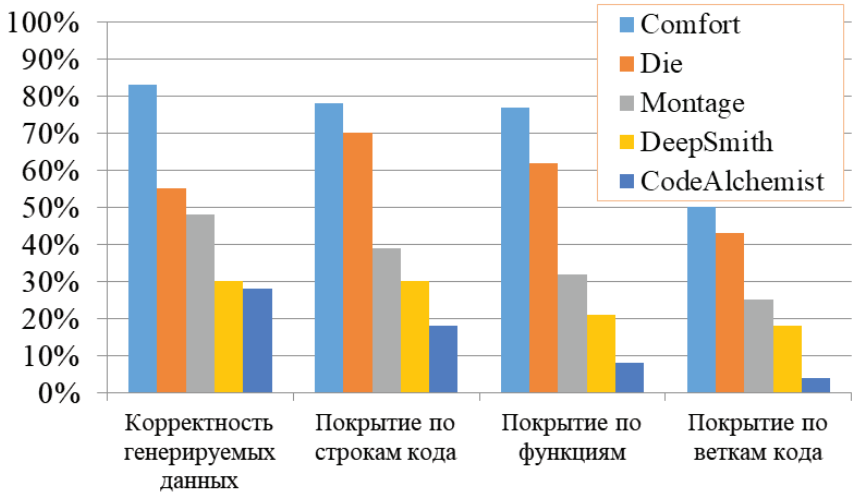

Рис. 1. Оценка существующих фаззеров интерпретаторов JavaScript

такими как программный код, сталкивается с несколькими ключевыми проблемами.

Первая проблема заключается в отсутствии общедоступных синтаксически и семантически корректных входных данных для проведения тестирования. Наличие общедоступного входного корпуса, изначально обеспечивающего высокое покрытие кода тестируемой программы, может значительно повысить эффективность всего процесса фаззинг-тестирования.

Второй является проблема преодоления внутренних механизмов фильтрации входных данных. Существующие фаззеры при генерации тестовых данных, как правило, разрушают тонкую семантику или условия, закодированные во входном корпусе, однако повышают процент покрытия кода программы. Такие тестовые данные отбрасываются тестируемой программой еще перед синтаксическим анализом и обработкой. Проверка выполняется с целью защиты программы от сбоев, вызванных некорректными входными данными. При фаззинге программного кода важным критерием при генерации или мутации входных файлов является сохранение синтаксической и семантической корректности. Эффективный фаззер должен полностью собирать тонкие условия, закодированные в высококачественном входном корпусе, таком как известные PoC-файлы, подтверждающие эксплуатацию уязвимостей [5] или модульные тесты интерпретаторов JavaScript.

Третья проблема заключается в поиске оптимального алгоритма изменения исходных данных. Стратегия генерации на основе мутаций широко используется современными фаззерами. Тем не менее, ключевой проблемой являются ответы на вопросы: как изменять и генерировать тестовые примеры, охватывающие больше программных путей как быстрее обнаруживать ошибки [6]. В частности, при выполнении мутации необходимо ответить на два вопроса: как и что мутировать. Одна мутация в нескольких ключевых позициях повлияет на поток управления выполнением. Кроме того, еще одна ключевая проблема заключается в том, как фаззеры изменяют ключевые позиции, то есть, как определить значение, которое могло бы направить тестирование на новые, еще не исследованные трассы в программе. Стратегии мутации должны быть направлены на создание высококачественных тестовых случаев, а не просто на увеличение охвата кода, чтобы можно было найти значимые, трудно обнаруживаемые ошибки. Отсутствие обратной связи по коду приводит к потере качества фаззинг-тестирования, а изменение стратегии мутации может значительно повысить эффективность фаззинга.

Четвертая проблема – это проблема повышения степени покрытия тестируемого кода. Повышение охвата кода означает повышение охвата состояний выполнения программы и повышение качества тестирования. Известно, что большее покрытие приводит к повышению вероятности обнаружения дефектов. Это подтверждается отчетом Миллера[6], который показал, что увеличение покрытия кода на 1% увеличивает процент обнаруженных ошибок на 0,92%. Однако большинство тестовых примеров охватывают только

6 C. Miller, Fuzz by number: More data about fuzzing than you ever wanted to know // in Proceedings of the CanSecWest – 2008.

некоторое, ограниченное число путей, в то время как большая часть кода не достигается. Однако существующие методы обычно фокусируются на покрытии кода, а не на уязвимом коде. Эти методы направлены на то, чтобы охватить как можно больше путей, а не исследовать пути, которые с большей вероятностью будут уязвимы. При выборе начальных значений для тестирования существующие фаззеры обычно обрабатывают все начальные входные данные одинаково, игнорируя тот факт, что пути, реализуемые различными начальными входными данными, не одинаково уязвимы. Это приводит к трате времени на тестирование безопасных, а не уязвимых путей, что снижает эффективность обнаружения уязвимостей. Как сообщается в [7], распределение ошибок в программах часто бывает несбалансированным, т. е. примерно 80% ошибок находятся примерно в 20% программного кода. В результате существующие фаззеры тратят много времени на тестирование «неуязвимых» путей, тем самым снижая эффективность фаззинга.

Метод генерации входных данных для фаззинга интерпретаторов

Метод генерации входных данных для фаззинг-тестирования интерпретаторов JavaScript веб-браузеров отличается использованием нейросетевой языковой модели, а также возможностью отслеживания информации о покрытии исходного кода.

Фаззинг на основе покрытия – широко используемый метод обнаружения ошибок и уязвимостей безопасности в программном обеспечении. Основная идея заключается в улучшении генерации будущих тестовых данных путем сбора обратной связи о текущем образце. С целью повышения эффективности подходы управляемого фаззинга требуют использования разумных мутаций, которые сохраняют особенности существующих образцов, слегка изменяя их семантику. С практической точки зрения это один из самых эффективных на сегодня типов фаззеров. На этой основе работают AFL++ [8], libFuzzer [9]. Несколько исследователей работали над разработкой различных стратегий мутации, основанных на различном поведении программы (например, фокусировке на редких ветвях, контексте вызова и т. д.) [10, 11]. Однако поведение программы резко меняется не только в разных программах, но и в разных частях одной и той же программы. Таким образом, поиск общей надежной стратегии мутации все еще остается важной открытой проблемой.

Анализ покрытия кода – действенный способ повышения эффективности фаззинга, однако нельзя отталкиваться только от него. Добившись высокой степени покрытия кода, все равно возможно пропустить критически важные участки [12].

Применение управляемого фаззинга к интерпретаторам JavaScript нетривиально, поскольку оно требует определения разумных изменений в программном коде. Вследствие чего, результаты фаззинга интерпретаторов сильно уступают результатам, достигнутым в других областях.

Эффективность этого подхода может быть повышена за счет исходного корпуса, который реализует более крупные части целевой программы, предоставляя более широкие границы для поиска входных данных.

Применение методов машинного обучения во многих исследованиях в области кибербезопасности как для обнаружения уязвимостей [13-14]; так и в фаззинг-тестировании [15-18] демонстрирует впечатляющие результаты. Преимуществом нейронных сетей является возможность обработки больших объемов данных с целью выявления закономерностей, что может быть применено для генерации сложноструктурированных данных для фаззинга.

Машинное обучение не может работать напрямую с кодом, ему необходимо подавать на вход числа, поэтому для работы с языковыми моделями необходима процедура токенизации – преобразование кода в последовательность чисел. В общем случае, на вход нейронной сети подаются не фрагменты кода или его структурные единицы, а токены – результат разбиения строки кода на непересекающиеся подстроки. Процедура токенизации для работы с нейросетями широко изучалась при решении задач завершения кода [19-20]. Однако генерация исполняемого теста является более сложной задачей, чем проблема завершения кода, которая предсказывает ограниченное количество семантически корректных лексических токенов.

В работе [2] предложена идея работы обучения языковых моделей не токенами, а AST-фрагментами и их последовательностями.

Фрагмент $T = (N, E, n|\ \ |0)$ – это поддерево $T = E_i, n_i)$, где

$n_i \in NC(n_i) \neq 0;$

$N_i = \{n_i\} \cup C(n_i);$

$E_i = \{(n_i; n') \vee n' = C(n_i)\}.$

где: N – множество узлов, E – множество ребер, n_0 – корневой узел, $C(n_i)$ – непосредственный потомок n_i, где n_i – узел в T.

Данный способ сохраняет семантику в обучающем наборе, разбивая узлы AST-дерева на фрагменты, ко-

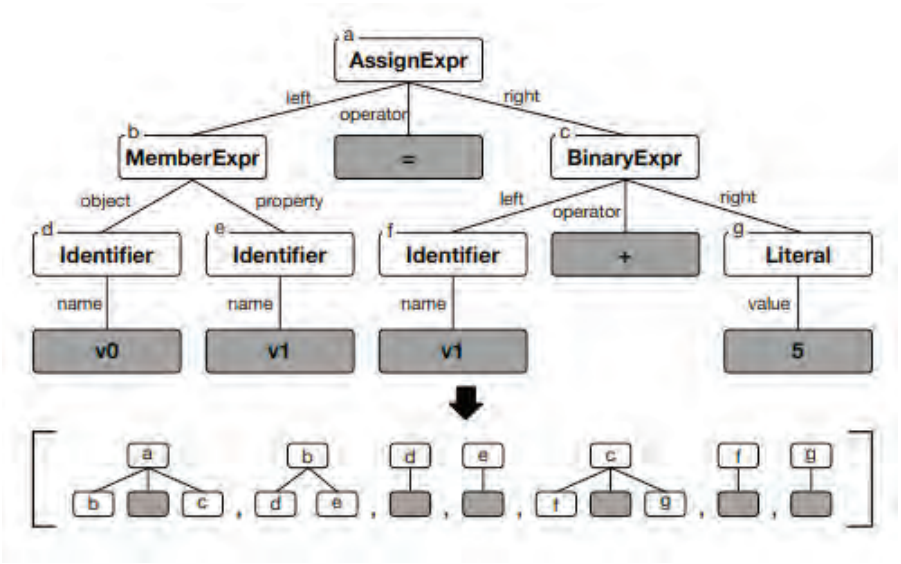

Рис. 2. Процесс фрагментации AST-дерева

торые используются в качестве лексикона для генерации кода JS. Процесс фрагментации AST-дерева представлен на рисунке Рис. 2. Процесс фрагментации AST-дерева. Используя фрагменты, инкапсулирующие структурные связи AST-деревьев, AST-дерево кодируется в последовательности фрагментов. Таким образом, возможно фиксировать глобальные отношения композиции между фрагментами кода для выбора следующего фрагмента.

Также были выявлены следующие закономерности:

1. Уязвимости интерпретаторов JavaScript часто возникают из-за того, что один и тот же js-файл может быть повторно исправлен в случае наличия нескольких ошибок;

2. Более 95% AST-фрагментов синтаксически перекрываются между регрессионными тестами интерпретаторов и PoC-фрагментами кода, запускающими CVE.

Данные факты подразумевают, что вероятность обнаружения новой уязвимости безопасности путем сборки фрагментов кода из существующих наборов регрессионных тестов, гораздо выше. Таким образом, возможно сгенерировать новый, обеспечивающий хорошее покрытие тестируемого кода, набор входных данных для проведения на нем дальнейшего фаззинг-тестирования.

Процесс фаззинг-тестирования интерпретатора предложено разделить на два больших этапа:

— генерацию качественных входных данных для фаззинга;
— непосредственно сам процесс фаззинг-тестирования.

Данное разделение способствует повышению покрытия тестируемого кода тестами, а также ускорению процесса фаззинг-тестирования. Так как генерация базы входных данных является длительным процессом, более эффективно с точки зрения временных затрат проделать этот этап один раз до начала тестирования. Предварительно сгенерированный набор входных данных для фаззинга может использоваться далее при тестировании различных интерпретаторов и за счет дальнейших мутаций сгенерированных файлов выполнять тестирование на наличие уязвимостей.

Данный подход позволяет:

— ускорить процесс фаззинг-тестирования;
— удалить избыточность из входных данных;
— обеспечить большее покрытие тестируемого кода.

Для реализации метода генерации входных данных для фаззинг-тестирования интерпретаторов JavaScript веб-браузеров, отличающегося использованием нейросетевых языковых моделей, а также управляемого информацией о покрытии исходного кода была разработана следующая архитектура (Рис. 3).

Данная архитектура позволяет генерировать новые входные данные, обеспечивающие высокое начальное покрытие кода, на основе базы регрессионных тестов для интерпретаторов.

Процесс генерации также подразделяется на два этапа:

— обучение нейронной сети;
— генерация входных данных.

1. Первый этап состоит из следующих шагов:

2. Фильтрация ошибок и формирование AST-деревьев;

3. Нормализация идентификаторов;

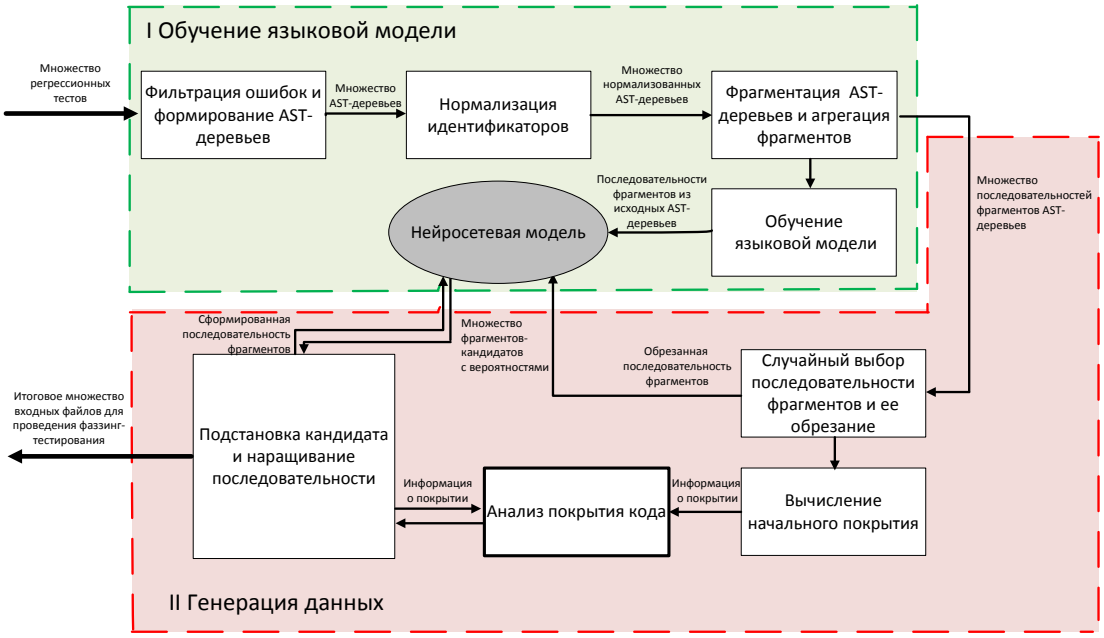

Рис. 3. Архитектура генератора входных данных

4. Фрагментация AST-деревьев и агрегация фрагментов;

Подбор гиперпараметров языковой модели и обучение нейросетевой модели.

Результатом первой фазы генерации являются:

— обученная нейросетевая модель;

— сформированное множество последовательностей фрагментов AST-деревьев.

Второй этап включает в себя следующие шаги:

1. Случайный выбор последовательности фрагментов из множества, сформированного в 1 фазе;
2. Выбор случайного фрагмента из последовательности;
3. Удаление поддерева из AST-дерева, для которого выбранный фрагмент является корневым;
4. Фиксация типа корневого фрагмента;
5. Подача на вход языковой модели обрезанной последовательности фрагментов;
6. Получение множества фрагментов-кандидатов от нейросетевой модели;
7. Отсев некорректных кандидатов по типу;
8. Наращивание последовательности по каждому фрагменту-кандидату;
9. Отсев полученных последовательностей, не прошедших валидацию;
10. Анализ прироста покрытия кода сгенерированными последовательностями;
11. Добавление в итоговое множество последовательностей, повышающих исходное покрытие кода;
12. Выбор новой последовательности из множества и повтор пунктов 3-11.
13. Минимизация полученного множества последовательностей;

В результате работы генератора формируется набор входных данных, с помощью которого можно провести более эффективное последующее фаззинг-тестирование.

Оценка покрытия тестируемого кода интерпретаторов JavaScript сгенерированными входными данными

В рамках данного исследования в качестве исходных данных была собрана база регрессионных тестов различных интерпретаторов JavaScript, была выбрана нейросеть долгой краткосрочной памяти (англ. Long Short Term Memory, LSTM) и JavaScript интерпетатор ChakraCore.

Для оценки покрытия рассматривались две широко используемые метрики: покрытие строк кода, а также покрытие функций. Две метрики соответственно измеряют среднее соотношение строк кода и функций тестируемой программы, а именно интерпретатора ChakraCore, которые выполняются во время тестового прогона. Для сбора информации о покрытии кода используются утилиты gcov [21] и lcov [22].

Используя вышеизложенный метод генерации входных данных для фаззинга интерпретаторов, удалось сгенерировать новый набор входных данных,

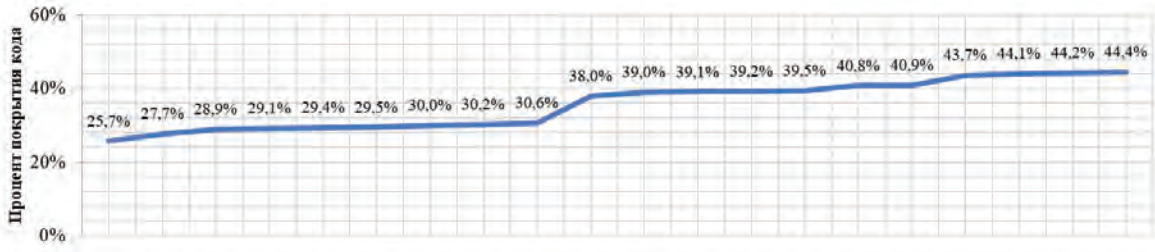

Рис. 4. График роста покрытия при запуске метода генерации

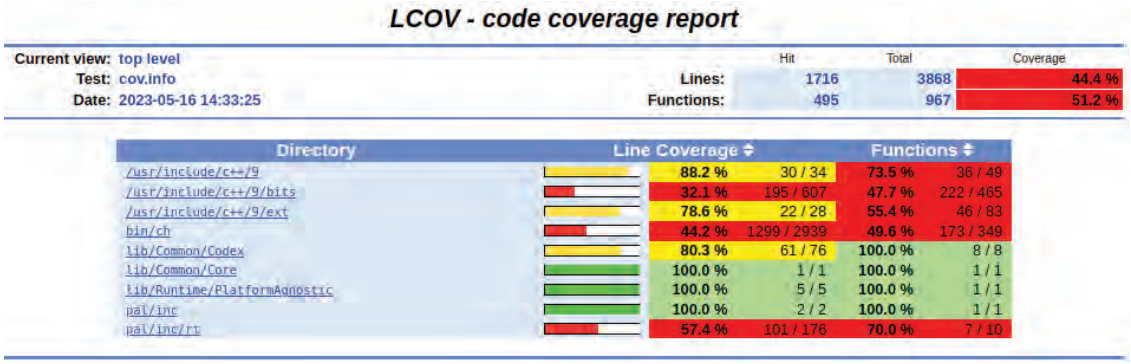

Рис. 5. Отчет утилиты lcov об увеличении покрытия кода интерпретатора

обеспечивающий покрытие 44.4 % по строкам кода и 51.2 % по функциям (Рис. 4).

Данные результаты демонстрируют возможность успешной генерации сложноструктурированных входных данных, таких как JavaScript код, для последующего фаззинг-тестирования. Преимуществом данного метода является ускорение процесса фаззинга, за счет разделения процесса тестирования на два этапа.

Выводы

Фаззинг-тестирование сложного программного обеспечения, такого как интерпретатор языка JavaScript, со сложно структурированными входными данными является актуальной и трудоемкой задачей. В работе приведены актуальные уязвимости веб-браузеров, а также ключевые проблемы, возникающие при тестировании интерпретаторов веб-браузеров. Наиболее актуальными проблемами являются: отсутствие общедоступных синтаксически и семантически корректных входных данных, проблема преодоления внутренних механизмов фильтрации входных данных, выбор рационального алгоритма мутации данных, а также проблема повышения степени покрытия тестируемого кода. Авторами предложен метод генерации входных данных для фаззинг-тестирования интерпретаторов JavaScript, который позволяет повысить качество и скорость последующего фаззинг-тестирования.

Литература

14. Bytes A. et al. Field Fuzz: In Situ Blackbox Fuzzing of Proprietary Industrial Automation Runtimes via the Network //Proceedings of International Symposium on Research in Attacks, Intrusions and Defenses (RAID). – 2023 DOI: 10.1145/3607199.3607226.
15. Lee S. et al. Montage: A neural network language model-guided javaScript engine fuzzer //Proceedings of the 29th USENIX Conference on Security Symposium. – 2020. – С. 2613-2630, DOI: 10.48550/arXiv.2001.04107.
16. Groß S. Fuzzil: Coverage guided fuzzing for JavaScript engines // Department of Informatics, Karlsruhe Institute of Technology, 2018.
17. Козачок А. В. и др. Обзор исследований по применению методов машинного обучения для повышения эффективности фаззинг-тестирования // Вестник Воронежского государственного университета, серия: системный анализ и информационные технологии. – 2021. – №. 4. – С. 83–106., DOI: 10.17308/sait.2021.4/3800.
18. C. Han. js-vuln-db, A collection of JavaScript engine CVEs with PoCs,2019. https://github.com/tunz/js-vuln-db.
19. Huang W. et al. testrnn: Coverage-guided testing on recurrent neural networks //arXiv preprint arXiv:1906.08557. – 2019, DOI: 10.48550/arXiv.1906.08557.

20. Gouveia I. P., Völp M., Esteves-Verissimo P. Behind the last line of defense: Surviving SoC faults and intrusions //Computers & Security. – 2022. – Т. 123. – С. 102920, DOI: 10.1016/j.cose.2022.102920.
21. Fioraldi A. et al. AFL++ combining incremental steps of fuzzing research //Proceedings of the 14th USENIX Conference on Offensive Technologies. – 2020. – С. 10-10.
22. Chao W. C. et al. Design and Implement Binary Fuzzing Based on Libfuzzer //2018 IEEE Conference on Dependable and Secure Computing (DSC). – IEEE, 2018. – С. 1-2.
23. Peng Chen and Hao Chen. 2018. Angora: Efficient fuzzing by principled search.
24. Österlund S. et al. Parmesan: Sanitizer-guided greybox fuzzing //Proceedings of the 29th USENIX Conference on Security Symposium. – 2020. – С. 2289-2306.
25. Козачок А. В., Николаев Д. А., Ерохина Н. С. Подходы к оценке поверхности атаки и фаззингу веб-браузеров //Вопросы кибербезопасности. – 2022. – №. 3 (49). – С. 32–43, DOI: 10.21681/2311–3456-2022-3-32-43.
26. Lin G. et al. Software vulnerability detection using deep neural networks: a survey //Proceedings of the IEEE. – 2020. – Т. 108. – №. 10. – С. 1825-1848, DOI: 10.1109/JPROC.2020.2993293.
27. Hanif H. et al. The rise of software vulnerability: Taxonomy of software vulnerabilities detection and machine learning approaches // Journal of Network and Computer Applications. – 2021. – Т. 179. – С. 103009, DOI: 10.1016/j.jnca.2021.103009.
28. Chernis B, Verma R. Machine Learning Methods for Software Vulnerability Detection. In: Proceedings of the Fourth ACM International. Workshop on Security and Privacy Analytics ACM. 2018. p. 31–39, DOI: 10.1145/3180445.3180453.
29. Zhu X. et al. Fuzzing: a survey for roadmap // ACM Computing Surveys (CSUR). – 2022. – Т. 54. – №. 11s. – С. 1-36, DOI: 10.1145/3512345.
30. Kaloudi N., Li J. The ai-based cyber threat landscape: A survey //ACM Computing Surveys (CSUR). – 2020. – Т. 53. – №. 1. – С. 1-34, DOI: 10.1145/3372823.
31. She D, Pei K, Epstein D, Yang J, Ray B, Jana S. NEUZZ: Efficient Fuzzing with Neural Program Smoothing; IEEE Symposium on Security & Privacy; 2019 – с. 38, DOI: 10.1109/SP.2019.00052.
32. Allamanis M. et al. A survey of machine learning for big code and naturalness //ACM Computing Surveys (CSUR). – 2018. – Т. 51. – №. 4. – С. 1-37, DOI: 10.1145/3212695.
33. Karampatsis R. M. et al. Big code! = big vocabulary: Open-vocabulary models for source code //Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering. – 2020. – С. 1073-1085, DOI: 10.1145/3377811.3380342.
34. Hu Jr Z. A Software Package for Generating Code Coverage Reports with Gcov, 2021.
35. Beyer D., Lemberger T. TestCov: Robust test-suite execution and coverage measurement // 34th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2019, pp. 1074-1077, DOI: 10.1109/ASE.2019.00105.

METHOD FOR SEMANTICALLY CORRECT CODE GENERATION FOR FUZZING TESTING JAVASCRIPT ENGINES

Kozachok A.V.[7], Spirin A.A.[8], Erokhina N.S.[9]

Purpose of the work *is to develop of a method for input data generation for fuzzing testing of JavaScript engines and its evaluation.*

Research method *studying the patterns of data generation and the percentage of code coverage in order to increase it. The proposed method allows you to generate input data to identify more vulnerabilities during subsequent fuzzing testing, by increasing the percentage of code coverage.*

Results of the research: *the JavaScript engines is the most vulnerable block of the web-browser architecture, as a result, there is a need to constantly increase the volume of analysis/testing of its source code. Fuzzing testing of a web-browser engine based on complexly structured input data, such as JavaScript code, is an urgent task. The paper presents the vulnerabilities of modern web-browsers, as well as key problems that arise when testing JavaScript engines. The most significant problems are: the lack of publicly available syntactically and semantically*

7 Alexander V. Kozachok, Dr.Sc., Associate Professor, Academy of the Federal Guard Service of the Russian Federation, Orel, Russia. E-mail: a.kozachok@academ.msk.rsnet.ru, https://orcid.org/0000-0002-6501-2008

8 Andrey A. Spirin, Ph.D., Academy of the Federal Guard Service of the Russian Federation, Orel, Russia. E-mail: spirin_aa@bk.ru, https://orcid.org/ 0000-0002-7231-5728

9 Natalya S. Erokhina, Academy of the Federal Guard Service of the Russian Federation, Orel, Russia. E-mail: ens@secdev.space, https://orcid.org/0000-0002-4878-0865

correct input data for fuzzing testing, the problem of overcoming internal mechanisms for filtering input data, the choice of a rational data mutation algorithm, and the problem of increasing the degree of coverage of the code under test. The authors propose a method for generating input data for fuzzing testing of JavaScript engines, which improves the quality and speed of fuzzing testing.

Scientific and practical significance: *the results lies in the development of a new method for generating input data for fuzzing testing of JavaScript engines of web-browsers, based on the use of neural network language models, which increases the coverage of the source code.*

Keywords: *web-browser, JavaScript engine, code coverage, software defects, software vulnerabilities, fuzzing testing, information security.*

References

1. Bytes A. et al. FieldFuzz: In Situ Blackbox Fuzzing of Proprietary Industrial Automation Runtimes via the Network //Proceedings of International Symposium on Research in Attacks, Intrusions and Defenses (RAID). – 2023 ¬ DOI: 10.1145/3607199.3607226.
2. Lee S. et al. Montage: A neural network language model-guided javascript engine fuzzer //Proceedings of the 29th USENIX Conference on Security Symposium. – 2020. – S. 2613-2630, DOI: 10.48550/arXiv.2001.04107.
3. Groß S. Fuzzil: Coverage guided fuzzing for javascript engines // Department of Informatics, Karlsruhe Institute of Technology, 2018.
4. Kozachok A. V. i dr. Obzor issledovanij po primeneniju metodov mashinnogo obuchenija dlja povyshenija jeffektivnosti fazzing-testirovanija // Vestnik Voronezhskogo gosudarstvennogo universiteta, serija: sistemnyj analiz i informacionnye tehnologii. – 2021. – №. 4. – S. 83–106., DOI: 10.17308/sait.2021.4/3800.
5. C. Han. js-vuln-db, A collection of JavaScript engine CVEs with PoCs, 2019. https://github.com/tunz/js-vuln-db.
6. Huang W. et al. testrnn: Coverage-guided testing on recurrent neural networks //arXiv preprint arXiv:1906.08557. – 2019, DOI: 10.48550/arXiv.1906.08557.
7. Gouveia I. P., Völp M., Esteves-Verissimo P. Behind the last line of de-fense: Surviving SoC faults and intrusions //Computers & Security. – 2022. – T. 123. – S. 102920, DOI: 10.1016/j.cose.2022.102920.
8. Fioraldi A. et al. AFL++ combining incremental steps of fuzzing research //Proceedings of the 14th USENIX Conference on Offensive Technologies. – 2020. – S. 10-10.
9. Chao W. C. et al. Design and Implement Binary Fuzzing Based on Libfuzz-er //2018 IEEE Conference on Dependable and Secure Computing (DSC). – IEEE, 2018. – S. 1-2.
10. Peng Chen and Hao Chen. 2018. Angora: Efficient fuzzing by principled search.
11. Österlund S. et al. Parmesan: Sanitizer-guided greybox fuzzing //Proceedings of the 29th USENIX Conference on Security Symposium. – 2020. – S. 2289-2306.
12. Kozachok A. V., Nikolaev D. A., Erohina N. S. Podhody k ocenke po-verhnosti ataki i fazzingu veb-brauzerov //Voprosy kiberbezopasnosti. – 2022. – №. 3 (49). – S. 32–43, DOI: 10.21681/2311-3456-2022-3-32-43.
13. Lin G. et al. Software vulnerability detection using deep neural networks: a survey //Proceedings of the IEEE. – 2020. – T. 108. – №. 10. – S. 1825-1848, DOI: 10.1109/JPROC.2020.2993293.
14. Hanif H. et al. The rise of software vulnerability: Taxonomy of software vulnerabilities detection and machine learning approaches // Journal of Network and Com-puter Applications. – 2021. – T. 179. – S. 103009, DOI: 10.1016/j.jnca.2021.103009.
15. Chernis B, Verma R. Machine Learning Methods for Software Vulnerabil-ity Detection. In: Proceedings of the Fourth ACM International. Workshop on Security and Privacy Analytics ACM. 2018. p. 31–39, DOI: 10.1145/3180445.3180453.
16. Zhu X. et al. Fuzzing: a survey for roadmap // ACM Computing Surveys (CSUR). – 2022. – T. 54. – №. 11s. – S. 1-36, DOI: 10.1145/3512345.
17. Kaloudi N., Li J. The ai-based cyber threat landscape: A survey //ACM Computing Surveys (CSUR). – 2020. – T. 53. – №. 1. – S. 1-34, DOI: 10.1145/3372823.
18. She D, Pei K, Epstein D, Yang J, Ray B, Jana S. NEUZZ: Efficient Fuzzing with Neural Program Smoothing; IEEE Symposium on Security & Privacy; 2019 – s. 38, DOI: 10.1109/SP.2019.00052.
19. Allamanis M. et al. A survey of machine learning for big code and natural-ness //ACM Computing Surveys (CSUR). – 2018. – T. 51. – №. 4. – S. 1-37, DOI: 10.1145/3212695.
20. Karampatsis R. M. et al. Big code! = big vocabulary: Open-vocabulary models for source code //Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering. – 2020. – S. 1073-1085, DOI: 10.1145/3377811.3380342.
21. Hu Jr Z. A Software Package for Generating Code Coverage Reports with Gcov, 2021.
22. Beyer D., Lemberger T. TestCov: Robust test-suite execution and coverage measurement // 34th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2019, pp. 1074-1077, DOI: 10.1109/ASE.2019.00105.

МЕТОДЫ ЗАЩИТЫ ВЕБ-ПРИЛОЖЕНИЙ ОТ ЗЛОУМЫШЛЕННИКОВ

Боровков В.Е.[1], Ключарев П.Г.[2]

Цель статьи: *аналитический обзор методов защиты веб-приложений.*

Метод исследования: *анализ научных публикаций по теме статьи.*

Полученные результаты: *в обзорной статье проанализирована литература, посвященная защите веб-приложений от уязвимостей, а также такого программного обеспечения, как веб-бэкдоры, которые встраиваются злоумышленником для выполнения нелегитимных операций. Высокая угроза последних обусловлена тем, что они могут загружаться на веб-сервер через уязвимости, а также через другие доступные для злоумышленника пути. К тому же исходный код веб-бэкдора может быть различным. Все это усложняет процесс эффективного обнаружения. В статье приведена классификация методов защиты и дана их сравнительная характеристика. Особое внимание уделяется интеллектуальным методам защиты и проблемам, которые возникают при обучении моделей. Основными проблемами являются некачественные, неполные наборы данных, а также отсутствие проверки многими исследователями своих результатов в реальных условиях.*

Научная новизна *заключается в систематизации и достаточно обширном обзоре работ в области защиты веб-приложений от уязвимостей и бэкдоров, которые могут быть использованы злоумышленниками. Работа выявляет проблемы, связанные с этой областью, что подчеркивает важность и актуальность данной темы.*

Ключевые слова: *веб-уязвимости, веб-бэкдоры, веб-шеллы, машинное обучение.*

DOI:10.21681/2311-3456-2023-5-89-99

Введение

В настоящее время наблюдается глобальный процесс информатизации общества, который сопровождается ростом числа онлайн-сервисов и веб-приложений. В свою очередь, этот процесс также увеличивает количество уязвимостей, которые часто используют злоумышленники. Под злоумышленником мы будем понимать нарушителя, преднамеренно совершающего попытки выполнения запрещенных операций с данными, следствием чего может являться нарушение информационной безопасности (ИБ). По данным компании Positive Technologies[3] 17% от общего числа атак связаны с уязвимостями и недостатками защиты веб-приложений, которые могут быть использованы для проникновения в локальный сетевой периметр организации или распространения вредоносного программного обеспечения.

Для защиты веб-приложений используют различные методы и средства, начиная от ручного анализа логов специалистами и применения антивирусов до использования облачных файрволлов. Системы защиты, такие как «Security Information and Event Management» (SIEM), как правило, генерируют большие объемы данных, что затрудняет их анализ. Использование математических расчетов и научных достижений во многом ускорило процесс анализа данных и принятия необходимого решения. Одним из актуальных направлений развития стало использование в системах кибербезопасности искусственного интеллекта, а в частности такого подкласса, как глубокое машинное обучение. Его преимущества по сравнению с традиционными методами машинного обучения, такие как низкий коэффициент ошибок, непрерывное усовершенствование и оптимизация различных алгоритмов обучения и упрощения сетей позволяют автоматизировать процесс анализа данных, прогнозирования и классификации.

3 Уязвимости и угрозы веб-приложений в 2020-2021 г.г. // Positive Technologies [Электронный ресурс]. 2022. – URL: https://www.ptsecurity.com/ru-ru/research/ analytics/web-vulnerabilities-2020-2021/ (дата обращения: 20.09.2022)

1 Боровков Владислав Евгеньевич, аспирант кафедры «Информационная безопасность» МГТУ им Н.Э. Баумана, Москва, Россия. E-mail: vbscience@yandex.ru

2 Ключарёв Петр Георгиевич, доктор технических наук, доцент кафедры «Информационная безопасность», МГТУ им. Н.Э. Баумана, Москва, Россия. E-mail: pk.iu8@yandex.ru

В данной работе будут рассмотрены теоретические основы нелегитимных действий злоумышленника в сфере веб-приложений, а также представлен обзор современных исследований в области методов их защиты.

1. Веб-приложения

Необходимо отметить, что существует множество определений таких понятий, как «веб-приложение», «веб-сайт», «веб-ресурс». Зачастую некоторые нативные приложения (приложения, которые разработаны для использования на определённой платформе) определяются как веб-приложения и наоборот. Некоторые приложения могут использовать веб-технологии для связи, обработки данных, хранения, и тогда такие системы можно рассматривать как веб-приложения. В данном исследовании мы будем использовать определение, предложенное в работе [1]: «веб-приложение» — это система с компонентами на стороне клиента (компоненты клиента), которые взаимодействуют с компонентами на веб-сервере (компоненты сервера) для обработки данных. Они используют веб-службу, основанную на клиент-серверной архитектуре, модели «запрос-ответ», стандартном HTTP и других связанных с ним методах и технологиях.

В качестве основы для классификации веб-приложений используем тип клиентского компонента. Мы можем разделить веб-приложения на основе этого на две группы – приложения, использующие браузер, и приложения, которые его не используют.

Для работы браузерных приложений необходимо специальное программное обеспечение (ПО), называемое браузером. В нем происходит выполнение всех клиентских компонентов. В качестве примера могут служить различные сайты, которые загружаются посредством браузеров Yandex, Google и т.д.

Приложения, не использующие браузер, во многом похожи на обычные настольные приложения, но взаимодействуют с веб-серверами на основе HTTP-запросов. Этот метод не является широко используемым для построения полноценной архитектуры приложения из-за недостатков с управлением и обслуживанием [1]. В свою очередь, эти приложения являются составной частью некоторых программных решений, таких как гибридные веб-приложения, которые сочетают в себе функции нативных и веб-приложений. Благодаря такой комбинации они могут быть легко адаптированы и развернуты на различных программных платформах, таких как IOS, Android, Windows и т.д. [2].

Некоторым промежуточным вариантом выступают прогрессивные веб-приложения (Progressive Web Application – PWA). Это технология позволяет клиентам устанавливать сайт как мобильное приложение. Она также сочетает в себе нативные функции операционной системы и стратегии веб-разработки, является альтернативой другим подходам из-за дополнительных преимуществ, таких как автономность и фоновая синхронизация [3].

Как можно увидеть, главным составляющим каждого из подходов к веб-разработке является использование HTTP-запросов для взаимодействия между клиентом и сервером. Наличие недостатков в логике этого взаимодействия зачастую приводит к атакам на веб-серверы и использованию их в нелегитимных целях.

2. Использование злоумышленниками уязвимостей веб-приложений

Открытый проект обеспечения безопасности веб-приложений (Open Web Application Security Project – OWASP) публикует известный список угроз веб-приложений, ранжированных в порядке от одного до десяти. На момент написания данной работы актуальным является список OWASP Top 10 – 2021[4]:

A01:2021 – Нарушение контроля доступа (Broken Access Control).

A02:2021 – Ошибки в криптографии (Cryptographic Failures).

A03:2021 – Внедрение кода (Injection).

A04:2021 – Небезопасный дизайн (Insecure Design).

A05:2021 – Неправильная конфигурация (Security Misconfiguration).

A06:2021 – Уязвимые и устаревшие компоненты (Vulnerable and Outdated Components).

A07:2021 – Ошибки идентификации и аутентификации (Identification and Authentication Failures).

A08:2021 – Нарушение целостности данных и программного обеспечения (Software and Data Integrity Failures).

A09:2021 – Журнал безопасности и сбои мониторинга (Security Logging and Monitoring Failures).

A10:2021 – Подделка запросов со стороны сервера или же SSRF (Server-Side Request Forgery).

В отчете от компании Positive Technologies[5] представлен подробный анализ веб-уязвимостей, основан-

4 OWASP Top 10:2021 // OWASP [Электронный ресурс]. 2021. – URL: https://owasp.org/Top10/ (дата обращения 01.10.2022).

5 Web application vulnerabilities and threats: statistics for 2019 // Positive Technologies [Электронный ресурс]. 2020. – URL: https://www.ptsecurity.com/ww-en/analytics/web-vulnerabilities-2020/ (дата обращения 05.10.2022).

ный на предыдущем списке OWASP Top 10 – 2017. Каждая из представленных угроз приводит к той или иной степени компрометации веб-приложения, что может привести к утечке чувствительных данных или получению контроля над управлением веб-сервером. Также используя некоторые уязвимости (например, межсайтовый скриптинг XSS, который теперь входит в категорию A03:2021), злоумышленники могут проводить атаки на клиентов, организовывать фишинговые атаки, например, для получения учетных данных, и выполнять действия, выдавая себя за другого пользователя.

Сама категория «Внедрение кода» всегда являлась критически опасной. Это метод, используемый злоумышленниками для внедрения вредоносного кода в уязвимые веб-приложения. Атака происходит чаще всего, когда администратор не добавляет правила, ограничивающие использование определенных символов. Некоторые из наиболее распространенных инъекций — SQL, NoSQL, команда ОС, реляционное сопоставление объектов (ORM), LDAP и внедрение языка выражений (EL) или библиотеки навигации по графу объектов (OGNL). Концепция одинакова среди всех интерпретаторов[6].

В крупных хакерских атаках на организации веб-сервер часто не является конечной целью. В качестве злоумышленников в таком случае могут выступать APT-группировки (Advanced Persistent Threat). В таких атаках используются различные изощренные методы и приемы с целью кражи конфиденциальной информации. Одной из приоритетных задач таких группировок является получение быстрого нелегитимного доступа к атакуемой системе, и веб-сервера часто выступают в качестве первой входной точки во внутреннюю сеть организации.

3. Использование злоумышленниками веб-бэкдоров

Получив доступ к веб-серверу, злоумышленники могут встроить в него *бэкдор*. Бэкдор – это средство доступа к компьютерной системе или зашифрованным данным в обход обычных механизмов безопасности системы. В качестве компьютерной системы в контексте исследования выступает веб-сервер, поэтому мы будем использовать понятие *веб-бэкдор*. В качестве бэкдора зачастую на таких серверах могут использоваться *веб-шеллы*.

Веб-шеллом называют скрипт, целью которого является обеспечение постоянного доступа к зараженным серверам для выполнения злонамеренных действий. Злоумышленники часто загружают хорошо продуманные сценарии веб-шеллов с помощью SQL-инъекций, уязвимостей неограниченной загрузки файлов и атак с использованием межсайтовых сценариев [4]. Также необходимо отметить, что веб-шеллы могут быть установлены и другими способами, например, злоумышленники, подобрав пароли от SSH или FTP, встраивают веб-шеллы для создания резервных каналов доступа к атакуемой системе.

Как отмечают исследователи в работе [5] в зависимости от функции и размера языка сценариев веб-шеллы можно условно разделить на три категории:

1) Большой троян (Big Trojan). Он имеет большой размер и обладает комплексными функциями для выполнения команд, работы с сетями, проведения операций с базами данных и осуществления других вредоносных намерений. Кроме того, к таким веб-шеллам применен дружественный графический интерфейс.

2) Однословный троян (One Word Trojan). Это веб-шелл с одной строкой кода, его часто встраивают в обычные файлы или картинки из-за своей компактности. Он может выполнять множество функций, подобно большому веб-шеллу, при этом код, который нужно выполнить, вместе с командами зачастую передается через HTTP запрос, что увеличивает объём полезной нагрузки.

3) Маленький троян *(Small Trojan).* Он имеет небольшой размер и его легко скрыть, но обычно он имеет только функцию загрузки файлов, поэтому его часто называют загрузчиком. Поскольку большинство веб-сайтов имеют ограничения по размеру при загрузке файлов, злоумышленники обычно сначала устанавливают загрузчик, а затем загружают на веб-сайт полноценный веб-шелл для выполнения ключевых функций.

Один из примеров пользовательского интерфейса веб-шелла представлен на рис. 1.

Сложнее всего обнаружить веб-шеллы второго и третьего типа. При этом для однословных троянов зачастую могут разрабатываться целые приложения, которые генерируют полезную нагрузку, и они по своей функциональности ничем не уступают большим троянам. Таким является китайский инструмент для управления веб-шеллами AntSword[7]. Этот инструмент может работать с директориями и файлами системы, подключаться к базам данных, создавать терминал и

6 A03:2021 – Injection // OWASP [Электронный ресурс]. 2021. – URL: https://owasp.org/Top10/A03_2021-Injection/ (дата обращения 25.10.2022).

7 Github. AntSword // Github [Электронный ресурс]. 2022. – URL: https://github.com/AntSwordProject/antSword (дата обращения 20.11.2022).

Рис.1. Пример пользовательского интерфейса большого трояна

использовать различные плагины, которые помогают обходить ограничения безопасности. Для его использования достаточно всего лишь одной строки кода на языке PHP – *eval(@$_POST['ant'])*.

Часто злоумышленники могут использовать веб-приложение для компрометации паролей пользователей, если оно включает авторизацию. Для этого они могут внедрить вредоносный код, называемый *стиллером* (от англ. *steal* – красть), который позволяет сохранять пароли пользователей в открытом виде. После получения паролей злоумышленники могут использовать их для авторизации в данном веб-приложении и других сервисах, что может привести к повышению привилегий злоумышленника.

Обнаружение уязвимостей веб-приложений и вредоносных средств, таких как веб-бэкдоры, может быть очень затруднено. Во ввиду этого методы защиты веб-приложений являются предметом изучения многих исследователей.

4. Методы защиты веб-приложений от веб-уязвимостей и веб-бэкдоров

Многие категории угроз могут быть обнаружены при использовании систем логирования и проверкой их администраторами. Однако проверка всех журналов безопасности может быть физически невозможна, особенно в случае, если администраторы должны обслуживать десятки или сотни серверов. В связи с этим для защиты веб-приложений используются различные средства, такие как межсетевые экраны веб-приложений, сканеры уязвимостей, системы обнаружения и предотвращения вторжений и т.д. В следующих разделах мы рассмотрим классификацию методов защиты веб-приложений и приведем обзор исследований в данной области.

4.1. Классификация методов защиты веб-приложений

Методы защиты веб-приложений могут быть классифицированы по нескольким факторам. Такие фак-

торы могут включать в себя тип защищаемого веб-приложения, класс атак, которые метод защиты обнаруживает или предотвращает, используемые техники и т.д. В статье [6] была предложена классификация методов защиты, достаточно полно охватывающая все современные классы атак, направленные на веб-ресурсы.

Для классификации методов защиты веб-приложений мы условно используем два независимых набора свойств: «*Анализируемые данные*» и «*Основание обнаружения*».

Методы защиты для анализа могут использовать:

1) Входные данные веб-приложения. Сюда входят значения полей HTTP-запросов.

2) Данные веб-приложения. Для анализа может использоваться исходный код веб-приложения, а также данные, сгенерированные веб-сервером и хранящиеся на нем.

3) Выходные данные веб-приложения. К ним относятся сгенерированные HTTP-ответы, их заголовки и поля.

По основаниям обнаружения методы можно разделить на следующие категории:

1) Метод на основе политик и правил. Предполагает анализ данных, выявление сигнатур и установление определенных правил.

2) Метод на основе намерений. Предполагает анализ и сравнение работы приложения с установленной заранее спецификацией веб-приложения.

3) Статистический метод. Предполагает использование вероятностного подхода к выявлению веб-атаки.

Данные классификации являются независимыми, и каждая из них может использоваться для описания метода защиты. Например, защита может контролировать входные данные веб-приложения и принимать решения на основе политик и правил. К таким средствам защиты можно отнести межсетевые экраны веб-приложений (Web application firewall – WAF), которые проверяют поступающий на приложение трафик HTTP/HTTPS, после чего принимают решения на основании заданных правил (блокировать, разрешить, отправить уведомление).

В следующем разделе приведем более подробное техническое описание классификаций и рассмотрим работы исследователей в области защиты веб-приложений от уязвимостей.

4.2. Методы защиты от веб-уязвимостей

Средства защиты веб-приложений зависят от данных, которые они использует для анализа, и могут быть размещены на веб-сервере или за его пределами. Объем доступных данных может отличаться в зависимости от выбранного варианта. При использовании веб-приложения как «черного ящика» у средства защиты есть возможность анализировать только входные и/или выходные данные из него, а само средство находится за пределами веб-сервера. Так работают сканеры веб-приложений такие как Acunetix, AppScan, BurpSuite, Arachni, W3af т.д. Также к таким средствам относятся облачные файрволлы, такие как Cloudflare.

Если используется метод «белого ящика», то защитное средство имеет возможность анализировать программный код и/или процессы, происходящие внутри сервера. Хотя данный метод позволяет выявлять более широкий круг проблем, так как обладает гораздо большим количеством информации, он менее универсален, чем метод «черного ящика».

Согласно вышеописанной классификации на основе «Анализируемых данных», можно установить, что методы, основанные на «черном ящике», используют только входные и выходные данные веб-приложений. Методы «белого ящика», кроме этого, могут также использовать данные самого веб-приложения.

В работе [7] исследователи анализируют различные методы тестирования программного обеспечения на каждом этапе жизненного цикла, которые связаны как с тестированием «черного», так и «белого ящика».

В недавней работе [8] исследователи провели сравнение генерации тестовых сценариев для REST API (Representational State Transfer Application Programming Interface) на основе «черного» и «белого ящиков» и показали, что комбинация обоих подходов дает наилучшие результаты в большинстве исследований с точки зрения поиска ошибок.

Выбор набора данных, необходимого и достаточного тому или иному методу защиты является одной из основных задач. Другой, не менее важной задачей, является правильная манипуляция этими данными, выбор техники и основания для обнаружения уязвимостей.

4.2.1 Методы на основе политик и правил

Метод на основе политик и правил является часто используемым методом защиты веб-приложений. Брандмауэры веб-приложений, или WAF, просматривают каждый запрос и/или ответ на различных уровнях обслуживания, таких как HTTP и HTTPS. Большинство существующих WAF основаны на правилах, ис-

пользующие регулярные выражения и сигнатуры, для обнаружения ключевых синтаксических конструкций, которые могут являться признаком проведения атаки.

Анализом сигнатурных методов защиты занимались многие исследователи. В работе [9] представлено исследование производительности трех систем обнаружения вторжений на основе сигнатур (Signature-based Intrusion Detection Systems – SIDS) – Snort, ModSecurity и Nemesida в контексте веб-атак. Результаты показали, что заранее заданные конфигурации этих систем не обеспечивают достаточно высокой производительности и способности обнаруживать известные атаки даже в самых чувствительных конфигурациях. Менее чувствительные конфигурации обеспечивали очень низкую скорость обнаружения, а наиболее чувствительные – неприемлемую точность и частоту срабатывания. Тем самым, исследователи поставили под сомнение вопрос о роли ненастроенных SIDS с открытым исходным кодом в качестве основных элементов защиты в контексте веб-служб.

Основным недостатком WAF на основе сигнатур является постоянная необходимость обновления их баз. Тем не менее, невозможно включить все сигнатуры в набор политик безопасности. Кроме того, наличие избыточного количества ненужных сигнатур также увеличивает вероятность блокировки законного трафика, повышая количество ложных срабатываний. При этом необходимо понимать, что они не защищают от атак «*нулевого дня*» (0-day). Обычно рекомендуемой контрмерой против таких кибератак является своевременное применение исправлений ПО, распространяемых поставщиком. Однако существует период от выявления уязвимости до выхода исправления.

В работе [10] авторы предлагают уменьшить воздействие атак «нулевого дня» на веб-приложения путем создания системы, которая собирает информацию об уязвимостях, связанных с веб-приложениями, в режиме реального времени в Интернете и генерирует сигнатуры брандмауэра веб-приложений. Для получения актуальной информации об уязвимостях система использует потоки данных, такие как социальные сети и веб-сайты для обсуждения технологий безопасности. В статье авторы использовали Twitter и базу данных уязвимостей NVD. После очистки собранных данных система проверяет наличие связанных уязвимых веб-приложений и создает для них сигнатуры WAF в виде виртуального исправления.

Более модернизированным методом является генерация обобщенного правила на основе шаблонов атак, что позволяет определять такое же количество атак, что и сотни сигнатур. В зависимости от среды приложения политика безопасности на основе сигнатур работает с более 2000–8000 сигнатур, в то время как политика безопасности на основе правил требует всего несколько десятков правил для обнаружения эквивалентного количества атак. Примером такого файрволла является Wapples[8].

В работе [11] был представлен гибридный метод, который использует обнаружение на основе сигнатур (signature-based detection – SBD) и аномалий (anomaly-based detection – ABD). Для обнаружения аномалий использовался байесовский классификатор. Запросы, классифицированные как ненормальные, помещаются в базу сигнатур. При повторном получении таких запросов они блокируется на этапе проверки сигнатур. Исследователями был сделан вывод, что сочетание обоих подходов будет более эффективным в процессе обнаружения и предотвращения веб-атак.

4.2.2 Методы на основе намерений

В отличие от предыдущего данный метод учитывает первоначальные намерения разработчика веб-приложения. Под «*намерением*» подразумевается функциональность, которая должна быть заложена в приложении с учетом целей и решаемых задач. Намерения определяют требования к приложению.

Данный метод использует информацию о том, какие процессы и операции заложены и разрешены разработчиком веб-приложения. Как правило, такая информация находится в технической документации на программный продукт. Также ее можно получить из анализа исходного кода приложения.

На первом этапе происходит обработка полученной информации и формирование «намерений» разработчика. Этот этап может существенно усложниться, если связи с разработчиком нет. В итоге должен получиться список «намерений», который указывает, как должно работать веб-приложение при штатных условиях.

На следующем этапе происходит мониторинг веб-приложения и выявление непредвиденных «намерений». К примеру, в HTTP-запросе было передано неожиданное количество аргументов, или запрос оказался неожиданно большим.

Данный метод является достаточно трудным в реализации, так как зачастую очень сложно учитывать все намерения разработчика. Как правило, данный метод используют для защиты некоторых аспектов

8 Wapples – The logical web application // PentaSecurity [Электронный ресурс]. 2022. – URL: https://www.pentasecurity.com/product/wapples (дата обращения 17.11.2022).

веб-приложения (например, потоков SQL-запросов) в сочетании с другими методами защиты.

4.2.3 Статистические методы

Ввиду сложности создания эффективной системы, основанной на сигнатурах, системы защиты, использующие статистические методы стали предметом множества научных исследований. Данные методы предполагают использование вероятностного подхода к обнаружению атак и направлены в первую очередь на обнаружение уязвимостей «нулевого дня».

Статистические подходы основаны на моделях. Это означает, что для данных создается модель или прототип, и объекты оцениваются с точки зрения того, насколько хорошо они вписываются в модель. Обнаружение атак строится на поиске «аномальных» объектов, которые не соответствуют модели и связаны с недостатками или редкими событиями.

В последние годы алгоритмы обнаружения аномалий на основе машинного обучения применяются для разнообразных задач. Особую популярность получили алгоритмы, использующие глубокое машинное обучение. В статье [12] авторами представлен широкий обзор методов глубокого обучения для задач по кибербезопасности. Они охватили широкий спектр типов атак, включая вредоносные программы, спам, инсайдерские угрозы, сетевые вторжения, ввод ложных данных и др.

В статье [13] исследователи проводят структурированный обзор методов исследования в области обнаружения аномалий на основе глубокого обучения, а также оценивают эффективность внедрения методов глубокого обучения в различные области применения. В частности они указывают, что варианты моделей глубокого обучения без учителя, основанные на глубокой нейронной сети и рекуррентной нейронной сети с памятью LSTM (Long Short-Term Memory), превосходят традиционные методы, такие как метод главных компонент (PCA), метод опорных векторов (SVM) и Isolation Forest в таких областях приложений, как здравоохранение и кибербезопасность.

В статье [14] авторами исследована эффективность WAF на основе машинного обучения. В некоторых приложениях оцениваемая система достигла точности 98,8% (во избежание недопониманий, под точностью (accuracy) будем понимать долю объектов, для которых был правильно предсказан класс. Она определяется формулой (1), где y и y^{pred} – настоящие и предсказанные метки классов соответственно, N – общее количество объектов, $\mathbb{I}\left[y_i = y_i^{pred}\right]$ возвращает единицу, если метки классов совпадают). Они показали, что методы, основанные на машинном обучении, имеют преимущества перед сигнатурными методами, так как могут предотвратить атаки нулевого дня, проще настраиваются и поддерживаются в актуальном состоянии.

$$Accuracy\left(y, y^{pred}\right) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\left[y_i = y_i^{pred}\right] \quad (1)$$

Все чаще исследователи показывают эффективность гибридных методов защиты. А. Текерек и др. в исследовании [15] продолжили усовершенствовать гибридную модель брандмауэра веб-приложений, используя SBD и ABD. Всего алгоритм включает три стадии обнаружения:

— 1 стадия – обнаружение известных типов атак по сигнатурам;
— 2 стадия – проверка через список HTTP-запросов, которые раньше были идентифицированы как аномальные;
— 3 стадия – проверка ABD и обновление списка аномальных HTTP-запросов.

Обнаружитель ABD реализуется с использованием искусственных нейронных сетей (Artificial Neural Networks – ANN). Предлагаемая модель была протестирована с использованием наборов данных WAF2015, CSIC2010 и ECML-PKDD. Согласно результатам теста, средний процент точности составлял 96,59 %.

Развитие машинного обучения также может отражаться на создании новых подходов к проведению атак. Авторами статьи [16] был представлен инструмент WAF-A-MoLE, который предназначен для обхода различных WAF, основанных на машинном обучении, путем поэтапного изменения полезной нагрузки – мутации кода. Сам инструмент использует состязательное машинное обучение (Adversarial Machine Learning – AML). Сутью AML является генерация таких входных данных, которые вводят в заблуждение модели машинного обучения, что приводит к неправильной классификации. Несмотря на то, что брандмауэры включали в себя различные алгоритмы машинного обучения, такие как рекуррентные нейронные сети, SVM, случайные леса и т.д., инструмент WAF-A-MoLE показал возможность их обхода.

Методы машинного обучения показывают свою эффективность, но при этом также имеют ряд проблем [17]:

1) *Недостаточный размер обучающих данных*. При обучении даже для решения простых задач необходимо большое количество данных, а для таких задач как

распознавание изображений или речи, понадобится миллионы образцов.

2) *Нерепрезентативные обучающие данные*. Для эффективного обучения необходимо использовать репрезентативные обучающие данные, которые могут быть обобщены на новые примеры.

3) *Данные плохого качества*. Чем больше в данных будет ошибок и шума, тем сложнее алгоритмам будет выявить закономерности в данных.

4) *Несущественные признаки*. Важно, чтобы обучающие данные имели как можно больше существенных признаков на входе.

5) *Переобучение обучающими данными*. При чрезмерном обучении может возникнуть случай, когда модель переобучается – хорошо выполняется на обучающих данных, но не обобщается при тестировании и использовании модели. Для ограничения модели с целью ее упрощения и снижения риска переобучения используют регуляризацию.

6) *Недообучение обучающими данными*. В данном случае для обучения используется слишком простая модель, чтобы узнать лежащую в основе структуру данных.

Кроме того, авторами статьи [18] была выявлена проблема, что модели машинного обучения, показывающие высокую эффективность в тестовой среде, при исследовании на реальном сетевом трафике демонстрируют низкую эффективность. Это говорит о плохих исходных данных, которые не вполне соответствуют реальности. Поэтому при разработке средства защиты успешные эксперименты должны проверяться в реальных условиях.

4.3. Методы защиты от веб-бэкдоров

Веб-бэкдоры могут быть загружены через уязвимости в веб-приложениях или слабую конфигурацию веб-сервера. Злоумышленник может встраивать веб-бэкдор в различные места, использовать обфускацию и шифрование, создавая уникальный код. Эти факторы затрудняют обнаружение веб-бэкдоров.

Подходы к обнаружению веб-бэкдоров могут существенно различаться и включать в себя анализ исходного кода, журналов безопасности и логирования, а также сетевого трафика.

С анализом исходного кода связано много исследований. В работе [4] авторами предложен метод обнаружен веб-шеллов на языке PHP, основанный на выделении признаков на различных уровнях представления (лексические, синтаксические и абстрактные). Затем данные признаки использовались в модели машинного обучения, основанной на методе опорных векторов, для обнаружения веб-шеллов. В работе [19] модель обнаружения основана на внутристрочной ассоциации слов и использовании Word2vec для векторного представления слов, которое затем используются в управляемом рекуррентном блоке (Gated Recurrent Units – GRU) для выполнения процесса обучения. Авторами работы [20] был предложен метод обнаружения с использованием модели сверточной нейронной сети при анализе кода на языках PHP, ASP, JSP.

Вместо того, чтобы производить анализ исходных кодов и содержимого пакетов HTTP-трафика, авторами работы [5] был предложен метод, который предполагает выделение и анализ сеансов, полученных из веб-журналов. Функции были извлечены из необработанных данных в веб-журналах, и для точного определения сеансов был применен статистический метод, основанный на временном интервале. Эксперимент показал, что модель на основе рекуррентной сети LSTM может достигать точности 95.97% при показатели полноты в 96,15%.

Значимых результатов достигли исследователи в работе [21]. Основная идея заключалась в том, что они производили анализ исходного кода PHP и выделяли 3 группы характеристик. Первая группа включала в себя статические признаки, такие как информационная энтропия, длина самого длинного слова и др. Вторая группа характеристик связана с выделением кода операций с использованием отладчика PHP PHPDBG. Третья группа связана с построением абстрактного синтаксического дерева [22] кода PHP для извлечения характеристик исполняемых данных с помощью PHP-Parser[9]. Эксперименты проводились на наборе данных, состоящих из 2917 образцов веб-шеллов. Результаты показали эффективность модели, достигнув точности обнаружения 99,66% без изучения оптимизации алгоритма машинного обучения. Идеи с выделением кода операций также использовали исследователи в работе [23].

Обобщенные результаты рассмотренных исследований представлены в табл. 1.

Главной трудностью, с которой сталкиваются исследователи, является наборы данных низкого качества. Веб-шеллы могут быть как отдельными файлами, так и встраиваться в легитимный программный код. Кроме того, они могут состоять из нескольких файлов, использовать различные поля и заголовки в пакетах

9 PHP Parser // Github [Электронный ресурс]. 2022. – URL: https://github.com/nikic/PHP-Parser (дата обращения 05.12.2022).

HTTP. Все это усложняет поиск и составление хорошего набора данных. Другой проблемой является то, что эксперименты не проводились на реальных системах, а результаты предоставлялись только на тех наборах, которые были извлечены исследователями. Так результаты, представленные в табл. 1, не могут являться объективными, так как эксперименты не были протестированы в реальных условиях. Наконец, многие исследования не являются универсальными, т.е. алгоритм для обнаружения веб-шелла, который создавался под конкретный язык программирования, неприменим или сложен в адаптации к другим языкам.

Таблица 1

Точность обнаружения веб-шеллов в различных исследованиях

Исследование	Язык программирования	Точность (accuracy)
[4]	PHP	92,18%
[19]	PHP, ASP, ASPX, JSP	99,19 %
[20]	PHP	99,5%
	JSP	97,5%
	ASP	98,3%
[5]	-	95.97%
[21]	PHP	99,66%
[23]	PHP	97,1%

5. Заключение

Современные веб-приложения – важная цель для злоумышленников. Через них они могут украсть конфиденциальную информацию и проникнуть в локальную сеть организации. «Закрепление» на веб-сервере является одним из методов сохранения доступа.

В статье рассматриваются основные угрозы, связанные с веб-приложениями, такие как уязвимости и использование злоумышленниками веб-бэкдоров. Уязвимости могут возникнуть из-за системных недостатков или ошибок при разработке приложения, и их эксплуатация может дать злоумышленникам первоначальный доступ к веб-серверу. Веб-бэкдор является следствием получения доступа, и встраивается злоумышленником для удобного выполнения несанкционированных действий. Чаще всего для таких целей злоумышленники использует веб-шеллы, а также веб-стиллеры для кражи учетных данных пользователей. Стоит отметить, что веб-бэкдоры могут быть загружены не только через уязвимости, но и через другие доступы к системе.

Методы защиты от угроз веб-приложений в последние время стали больше основываться на интеллектуальном анализе данных. Это связано с тем, что сигнатурные методы, несмотря на свою эффективность, не позволяют выявлять новые, заранее неизвестные, угрозы. К тому же злоумышленники часто производят обфускацию и шифрование для обхода таких методов защиты. Методы, основанные на машинном обучении, в свою очередь имеют свои проблемы, такие как низкое качество данных и их небольшой объем. К тому же многие исследователи в своих работах не проверяют эффективность методов на реальных системах, а ориентируются на результаты экспериментов, проводимых на выбранных наборах данных. Также отягчающими факторами являются трудная настройка системы и неуниверсальность методов. Ввиду этого исследования в данной области продолжают быть актуальными.

Литература

1. Dissanayake N., Dias. K. Web-based Applications: Extending the General Perspective of the Service of Web // 10th International Research Conference of KDU (KDU-IRC 2017) on Changing Dynamics in the Global Environment: Challenges and Opportunities, 2017.
2. Navyashree S., Rashmi R. Hybrid Web Application using Content Management System App used by all platforms // Advances in Computational Sciences and Technology. 2019. Vol. 12. P. 23-36.
3. Adetunji O., Ajaegbu C., Nzechukwu O. Dawning of Progressive Web Applications (PWA): Edging Out the Pitfalls of Traditional Mobile Development // American Scientific Research Journal for Engineering, Technology, and Sciences. 2020. Vol. 68(1). P. 85–99.
4. Zhu T., Weng Z., Fu L., Ruan L. A Web Shell Detection Method Based on Multiview Feature Fusion // Applied Sciences. 2020. Vol. 10(18). P. 1-16.
5. Wu Y., Sun Y., Huang C., Jia P., Liu L. Session-Based Webshell Detection Using Machine Learning in Web Logs // Security and Communication Networks. 2019. Vol. 2019. P. 1-11.
6. Лесько С. А. Модели и методы защиты веб–ресурсов: систематический обзор // CLOUD OF SCIENCE. 2020. № 3. С. 577-610.
7. Nidhra S. Black Box and White Box Testing Techniques - A Literature Review // International Journal of Embedded Systems and Applications. 2021. Vol. 2. P. 29-50.
8. Martin-Lopez A., Arcuri A., Segura S., Ruiz-Cortes A. Black-Box and White-Box Test Case Generation for RESTful APIs: Enemies or Allies? // Conference: International Symposium on Software Reliability Engineering. 2021. P.1-11.
9. Diaz-Verdejo J., Munoz-Calle J., Estepa A., Estepa R., Madinabeitia G. On the Detection Capabilities of Signature-Based Intrusion Detection Systems in the Context of Web Attacks // Applied Sciences. 2022. Vol. 12. P.1-16.
10. Kumazaki M., Yamaguchi Y., Shimada H., Hasegawa H. WAF Signature Generation with Real-Time Information on the Web // The

Fourteenth International Conference on Emerging Security Information, Systems and Technologies. 2020. P. 40-45.
11. Adem T., Cemal G., Omer F. Web tabanlı saldırı önleme sistemi tasarımı ve gerçekleştirilmesi: yeni bir hibrit model // Journal of the Faculty of Engineering and Architecture of Gazi University. 2016. Vol. 31(3). P. 645-653.
12. Berman D., Buczak A., Chavis J., Corbett C. A Survey of Deep Learning Methods for Cyber Security // Information (Switzerland). 2019. Vol. 10. P. 1-35.
13. Chawla S., Chalapathy R. Deep Learning for Anomaly Detection: A Survey // Computer Science. 2019. P.1-47.
14. Applebaum S., Gaber T., Ahmed A. Signature-based and Machine-Learning-based Web Application Firewalls: A Short Survey // Procedia Computer Science. 2021. Vol. 189. P. 359-367.
15. Tekerek A., Bay O. Design and Implementation of an Artificial Intelligence-Based Web Application Firewall Model // Neural Network World. 2019. Vol. 29. P. 189-206.
16. Demetrio L., Valenza A., Costa G., Lagorio G. WAF-A-MoLE: Evading Web Application Firewalls through Adversarial Machine Learning // Conference: 35th Annual ACM Symposium on Applied Computing. 2020. P. 1745-1752.
17. Жерон О., Прикладное машинное обучение с помощью Scikit-Learn, Keras и Tensorflow: концепции, инструменты и техники для создания интелектуальных систем, 2-е изд.: Пер. с англ. / О. Жерон. – СПб.: ООО «Диалектика», 2020. – 1040 с.
18. 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 Института системного программирования РАН. – 2020. – № 32(5). – С. 81-94.
19. Tingting L., Chunhui R., Yusheng F., Jie X., Jinhong G., Xinyu C. Webshell Detection Based on the Word Attention Mechanism // IEEE Access. Deep Learning: Security and Forensics Research Advances and Challenges. 2019. P. 185140 – 185147.
20. Zhuo-Hang L., Han-Bing Y., Rui M. Automatic and Accurate Detection of Webshell Based on Convolutional Neural Network // 15th International Annual Conference, CNCERT 2018. 2018. P. 73-85.
21. Pan Z., Chen Y., Chen Y., Shen Y., Guo X. Webshell Detection Based on Executable Data Characteristics of PHP Code // Wireless Communications and Mobile Computing. 2021. P.1-12.
22. Neamtiu I., Foster J., Hicks M. Understanding source code evolution using abstract syntax tree matching // in Proceedings of the 2005 international workshop on Mining software repositories - MSR '05. 2005. P. 1–5.
23. Fu J., Li L., Wang Y. Webshell Detection Based on Convolutional Neural Network // Journal of Zhengzhou University (Natural Science Edition). 2019. Vol 51(2). P. 1-8.

METHODS OF PROTECTING WEB APPLICATIONS FROM ATTACKER

Borovkov V.E.[10], Klyucharev P.G.[11]

The purpose of the article *is an analytical review of web application protection methods.*

Research method: *an analysis of scientific publications on the topic of the article.*

Results: *the review article analyzes the literature devoted to the protection of web applications from vulnerabilities, as well as software such as web backdoors that are embedded by an attacker to perform illegitimate operations. The high threat of the latter is due to the fact that they can be uploaded to the web server through vulnerabilities, as well as through other paths available to an attacker. In addition, the source code of the web backdoor may be different. All this complicates the process of effective detection. The article provides a classification of protection methods and their comparative characteristics. Special attention is paid to intellectual methods of protection and problems that arise when training models. The main problems are poor-quality, incomplete datasets, as well as the lack of verification by many researchers of their results in real conditions.*

The scientific novelty *lies in the systematization and a fairly extensive review of works in the field of protecting web applications from vulnerabilities and backdoors that can be used by attackers. The work identifies problems related to this area, which emphasizes the importance and relevance of this topic.*

Keywords: *web vulnerabilities, web backdoors, web shells, machine learning.*

10 Vladislav Borovkov, postgraduate student of Information Security department, Bauman Moscow State Technical University, Moscow, Russia. E-mail: vbscience@yandex.ru

11 Petr G. Klyucharev, Grand PhD, associate professor of Information Security department, Bauman Moscow State Technical University, Moscow, Russia. E-mail: pk.iu8@yandex.ru

References

1. Dissanayake N., Dias. K. Web-based Applications: Extending the General Perspective of the Service of Web // 10th International Research Conference of KDU (KDU-IRC 2017) on Changing Dynamics in the Global Environment: Challenges and Opportunities, 2017.
2. Navyashree S., Rashmi R. Hybrid Web Application using Content Management System App used by all platforms // Advances in Computational Sciences and Technology. 2019. Vol. 12. P. 23-36.
3. Adetunji O., Ajaegbu C., Nzechukwu O. Dawning of Progressive Web Applications (PWA): Edging Out the Pitfalls of Traditional Mobile Development // American Scientific Research Journal for Engineering, Technology, and Sciences. 2020. Vol. 68(1). P. 85–99.
4. Zhu T., Weng Z., Fu L., Ruan L. A Web Shell Detection Method Based on Multiview Feature Fusion // Applied Sciences. 2020. Vol. 10(18). P. 1-16.
5. Wu Y., Sun Y., Huang C., Jia P., Liu L. Session-Based Webshell Detection Using Machine Learning in Web Logs // Security and Communication Networks. 2019. Vol. 2019. P. 1-11.
6. Lesko S. A. Modeli i metody zashchity web–resursov: sistematicheskiy obzor // CLOUD OF SCIENCE. 2020. № 3. P. 577-610.
7. Nidhra S. Black Box and White Box Testing Techniques - A Literature Review // International Journal of Embedded Systems and Applications. 2021. Vol. 2. P. 29-50.
8. Martin-Lopez A., Arcuri A., Segura S., Ruiz-Cortes A. Black-Box and White-Box Test Case Generation for RESTful APIs: Enemies or Allies? // Conference: International Symposium on Software Reliability Engineering. 2021. P.1-11.
9. Diaz-Verdejo J., Munoz-Calle J., Estepa A., Estepa R., Madinabeitia G. On the Detection Capabilities of Signature-Based Intrusion Detection Systems in the Context of Web Attacks // Applied Sciences. 2022. Vol. 12. P.1-16.
10. Kumazaki M., Yamaguchi Y., Shimada H., Hasegawa H. WAF Signature Generation with Real-Time Information on the Web // The Fourteenth International Conference on Emerging Security Information, Systems and Technologies. 2020. P. 40-45.
11. Adem T., Cemal G., Omer F. Web tabanlı saldırı önleme sistemi tasarımı ve gerçekleştirilmesi: yeni bir hibrit model // Journal of the Faculty of Engineering and Architecture of Gazi University. 2016. Vol. 31(3). P. 645-653.
12. Berman D., Buczak A., Chavis J., Corbett C. A Survey of Deep Learning Methods for Cyber Security // Information (Switzerland). 2019. Vol. 10. P. 1-35.
13. Chawla S., Chalapathy R. Deep Learning for Anomaly Detection: A Survey // Computer Science. 2019. P.1-47.
14. Applebaum S., Gaber T., Ahmed A. Signature-based and Machine-Learning-based Web Application Firewalls: A Short Survey // Procedia Computer Science. 2021. Vol. 189. P. 359-367.
15. Tekerek A., Bay O. Design and Implementation of an Artificial Intelligence-Based Web Application Firewall Model // Neural Network World. 2019. Vol. 29. P. 189-206.
16. Demetrio L., Valenza A., Costa G., Lagorio G. WAF-A-MoLE: Evading Web Application Firewalls through Adversarial Machine Learning // Conference: 35th Annual ACM Symposium on Applied Computing. 2020. P. 1745-1752.
17. Zheron O. Prikladnoye mashinnoye obucheniye s pomoshchyu Scikit-Learn. Keras i Tensorflow: kontseptsii. instrumenty i tekhniki dlya sozdaniya intelektualnykh sistem. 2 izd.: Per. s angl. / O. Zheron. – SPb.: OOO «Dialektika». 2020. – P. 1040.
18. Goryunov. M.N. Sintez modeli mashinnogo obucheniya dlya obnaruzheniya kompyuternykh atak na osnove nabora dannykh CICIDS2017 / M.N. Goryunov. A.G. Matskevich. D.A. Rybolovlev // Trudy Instituta sistemnogo programmirovaniya RAN. – 2020. – № 32(5). – P. 81-94.
19. Tingting L., Chunhui R., Yusheng F., Jie X., Jinhong G., Xinyu C. Webshell Detection Based on the Word Attention Mechanism // IEEE Access. Deep Learning: Security and Forensics Research Advances and Challenges. 2019. P. 185140 – 185147.
20. Zhuo-Hang L., Han-Bing Y., Rui M. Automatic and Accurate Detection of Webshell Based on Convolutional Neural Network // 15th International Annual Conference, CNCERT 2018. 2018. P. 73-85.
21. Pan Z., Chen Y., Chen Y., Shen Y., Guo X. Webshell Detection Based on Executable Data Characteristics of PHP Code // Wireless Communications and Mobile Computing. 2021. P.1-12.
22. Neamtiu I., Foster J., Hicks M. Understanding source code evolution using abstract syntax tree matching // in Proceedings of the 2005 international workshop on Mining software repositories - MSR '05. 2005. P. 1–5.
23. Fu J., Li L., Wang Y. Webshell Detection Based on Convolutional Neural Network // Journal of Zhengzhou University (Natural Science Edition). 2019. Vol 51(2). P. 1-8.

ОБ ОДНОМ КЛАССЕ АЛГОРИТМОВ АНАЛИЗА ПОВЕДЕНИЯ КОМПОНЕНТОВ УСТРОЙСТВ С ПРОГРАММИРУЕМЫМИ ПОЛЬЗОВАТЕЛЕМ ВЕНТИЛЬНЫМИ МАТРИЦАМИ

Титов А.С.[1], Гордеев Э.Н.[2]

Цель исследования: *исследование возможностей повышения защищённости аппаратных средств путём обнаружения участков схемы уровня регистровых передач, находящихся под угрозой нарушения конфиденциальности.*

Метод исследования: *математическое моделирование схемы уровня регистровых передач и применение к модели классических вероятностных алгоритмов проверки свойств булевых функций для обнаружения потенциально уязвимых мест внутренней логики микросхемы.*

Результаты исследования: *на основе построения комбинационных и последовательностных схем, выражающих внутреннюю логику устройств через совокупности булевых функций, построена конкретная модель конвейерной микроархитектуры с разделёнными состояниями и передачей данных, позволяющая проводить исследования путём применения выбранного математического аппарата.*

Выделена модель нарушителя конфиденциальности обрабатываемых специальными видами вычислительных устройств данных. Для конкретной модели конвейерной микроархитектуры и нарушителя конфиденциальности, который может быть расположен на любой стадии производственного процесса, рассмотрена задача минимизации размерности считываемых из схемы данных.

Проведён анализ одного класса алгоритмов в рамках исследуемой модели. По результатам анализа предложены модификации некоторых из них.

Построенные в работе алгоритмы позволяют за счёт выделения индексов аргументов булевых функций уточнять расположение тех входов-выходов смоделированных устройств, которые потенциально уязвимы к нарушению конфиденциальности элементов последовательностных и комбинационных схем.

Научная новизна: *заключена в анализе применимости одного класса вероятностных алгоритмов к задаче обнаружения уязвимых участков устройств, использующих схемную логику, и построении на их основе модификаций с целью улучшения точности определения входов уязвимых элементов.*

Ключевые слова: *программируемая логическая интегральная схема, регистровые передачи, конвейерная микроархитектура, модель нарушителя, булевы функции, схемы из функциональных элементов, существенные переменные.*

DOI:10.21681/2311-3456-2023-5-100-112

1. Введение

Программируемая пользователем вентильная матрица (далее – ППВМ) – интегральная микросхема, внутренняя логика (схема) которой проектируется разработчиком в зависимости от функций конструируемого устройства, используемого ППВМ.

Комбинационная схема (комбинационная логика) – схема на основе функциональных элементов (булевых функций), используемая при математическом моделировании.

Последовательностная схема (последовательностная логика) – схема с определенным видом структуры. (См. ниже точное определение).

Уровень регистровых передач – способ описания комбинационной и последовательностной схемы с ис-

1 Титов Анатолий Сергеевич, студент кафедры ИУ-8 «Информационная безопасность» МГТУ им. Н.Э. Баумана, Москва, Россия. E-mail: toliakpurple@gmail.com

2 Гордеев Эдуард Николаевич, доктор физико-математических наук, профессор кафедры ИУ-8 «Информационная безопасность» МГТУ им. Н.Э. Баумана, Москва, Россия. E-mail: werhorn@yandex.ru

пользованием логических операций, применяемых к данным, которые передаются между элементами схемы (см., например, [1]).

Компонент – это описанная на уровне регистровых передач составная часть устройства, которая выполняет заданную его разработчиком функцию.

Аппаратным трояном (аппаратной недекларированной возможностью) называется внедрённая злоумышленником возможность, которая влияет на информационную безопасность устройства. Участок схемы, который может быть использован злоумышленником для размещения аппаратного трояна, принято называть *уязвимым участком* [2].

Области применения ППВМ: анализ сетевых данных[3], цифровая обработка аналоговых сигналов[4], построение роботизированных систем, ускорение вычислений [3–5] и другие. В большинстве случаев актуальна необходимость обеспечения информационной безопасности и вследствие этого исследование угроз и анализа возможных атак злоумышленников на устройства с ППВМ [6].

К целям злоумышленника относятся, например, внедрение аппаратных троянов [7] и другие действия по нарушению конфиденциальности или целостности обрабатываемой информации[5].

Для исследования поведения компонентов устройств с ППВМ и локализации аппаратных троянов авторы работы [8] строят модель передачи данных внутри схемы уровня регистровых передач.

В [9–12] предложены методы анализа информационной безопасности таких устройств с использованием машинного обучения и нейронных сетей.

В работах [13, 14] рассмотрен инструментарий для автоматизированного выявления уязвимых мест схемы.

В статьях [15, 16] описана методика выделения уязвимых участков схемы уровня регистровых передач на основе, в частности, математического моделирования устройства с использованием систем булевых функций.

При анализе свойств и способов конструкции ППВМ с точки зрения информационной безопасности рассматривается несколько разных моделей нарушителей. В настоящей работе нарушителем считается участвующий в конструировании устройства с ППВМ разработчик, у которого есть доступ к описанию внутренней логики на уровне регистровых передач.

Его целью является закладка в конструкцию возможностей раскрытия конфиденциальных данных, которые обрабатываются внутренней логикой устройства. Для этого нарушитель может, например, внедрить в устройство компонент, считывающий, сохраняющий и передающий обрабатываемые этим устройством данные.

Сокрытие такого компонента обеспечивается уменьшением его физических размеров и потребляемых ресурсов (например, электропитания).

Одним из факторов, способствующих снижению использования ресурсов компонента, является уменьшение количества считываемых из устройства сигналов, а, следовательно, обеспечение возможности для хранения и дальнейшей передачи несанкционированной к распространению информации.

Для анализа наличия признаков уменьшения количества считываемых сигналов используется математическая модель устройства. В частности, для моделирования последовательностной логики применяются конечные автоматы Мура и Мили[6]. Конечные автоматы здесь используются для описания *последовательностной логики* с целью дальнейшего преобразования этого описания в схему моделирующую устройство.

Так как в настоящий работе рассматривается анализ уже существующей схемы, то он должен базироваться на свойствах самой ППВМ. А в этом случае конечные автоматы мы не применяем.

Мы будем использовать математическую модель устройства на базе систем булевых функций, а затем к ее анализу нами применён вероятностный алгоритм проверки свойств булевых функций. Подобные исследования осуществлены, в частности, в работах [17, 18].

В статьях [19, 20] рассматривается алгоритм проверки существенной зависимости булевой функции от не более, чем k переменных. Поводом для этого является тот факт, что сигналы, соответствующие фиктивным аргументам математической модели, могут быть исключены из считывания, уменьшив таким образом необходимые ресурсы компонента для хранения и передачи данных и высвободив их для других целей.

Отметим, что приведённые в этих работах вероятностные алгоритмы возвращают результат в форме распознавания («да» или «нет») без указания конкрет-

3 Trimberger S. Three Ages of FPGAs: A Retrospective on the First Thirty Years of FPGA Technology. Proceedings of the IEEE, vol. 103, no. 3, pp. 318-331, 2015. DOI: 10.1109/JPROC.2015.2392104

4 Romoth J., Porrmann M. Survey of FPGA applications in the period 2000 – 2015. 2017. DOI: 10.13140/RG.2.2.16364.56960

5 Li H., Liu Q., Zhang J. A survey of hardware Trojan threat and defense. // Integr. VLSI J., 2016

6 Pedroni V. Finite State Machines in Hardware: Theory and Design (with VHDL and SystemVerilog). The MIT Press. 2013. DOI: 10.7551/mitpress/9657.001.0001

ных существенных аргументов. В ряде случаев знание этих аргументов повышает эффективность методов обнаружения возможных закладок злоумышленника, о которых говорилось выше.

Статья состоит из введения и трех разделов. В разделе 2 описана модель комбинационной и последовательностной логики, построена модель конвейерной микроархитектуры и поставлена задача минимизации размерности считываемых из схемы данных.

Раздел 3 посвящён решению этой задачи. Для этого используемые ранее алгоритмы проверки существенной зависимости булевой функции от не более, чем k переменных, модифицированы так, чтобы возвращаемым значением было множество индексов существенных аргументов.

В разделе 4 рассмотрена сравнительная характеристика исходных и модифицированных алгоритмов.

Полученные здесь алгоритмы могут быть использованы для определения исключаемых или нарушаемых злоумышленником сигналов.

С точки зрения защиты информации, применение такого анализа предоставит данные для определения уязвимых участков схемы. К ним могут быть отнесены те участки, например, где количество считываемых сигналов отличается от того, которое может быть получено на базе проведенного анализа.

2. Математическая модель комбинационной и последовательностной логики

Определение 1. *Моделью комбинационной логики* названа тройка $\langle f, n, m\rangle$, где: $f\colon \{0,1\}^n \to \{0,1\}^m$ – система булевых функций.

Пусть $f_i\colon\{0,1\}^n \to \{0,1\}$, $i = \overline{1,m}$. Тогда элемент комбинационной логики задаётся вектор-функцией $f(\vec{X}) = \left(f_1(\vec{X}), \ldots, f_m(\vec{X})\right)$.

При описании *последовательностной логики* обычно используется понятие тактирования – пошагового задания входных и выходных данных. Вектор данных $\vec{X}$ такта (шага) $\mathbf{t}$ обозначается как $\vec{X}^t$.

Определение 2. *Моделью последовательностной логики* названа четвёрка $\langle g, I, n, m\rangle$, где g – отображение $\{0,1\}^{n+k} \to \{0,1\}^{m+k}$, а I – множество (мощности k) индексов аргументов, задающих внутреннее состояние последовательностной схемы.

На вход отображения g подаётся вектор $\vec{\tilde{X}}^t = \left(X_1^t, \ldots, X_n^t, S_1^t, \ldots, S_k^t\right)$, содержащий значения входов $X_1^t, \ldots, X_n^t$ и состояния $S_1^t, \ldots, S_k^t$ схемы. Выходом отображения является (1) конкатенация вектора выходных значений $\left(Y_1^{t+1}, \ldots, Y_m^{t+1}\right)$ и вектора состояния на следующем такте $\left(S_1^{t+1}, \ldots, S_k^{t+1}\right)$:

$$g\left(X_1^t, \ldots, X_n^t, S_1^t, \ldots, S_k^t\right) = \\ = \left(Y_1^{t+1}, \ldots, Y_m^{t+1}, S_1^{t+1}, \ldots, S_k^{t+1}\right) \quad (1)$$

Отображение g модели $\langle g, I, n, m\rangle$ представимо в виде (2) совокупности булевых функций преобразования данных $g_i\colon\{0,1\}^{n+k} \to \{0,1\}$ и функций перехода состояний $s_j\colon\{0,1\}^{n+k} \to \{0,1\}$:

$$\begin{cases} g_1\left(X_1^t, \ldots, X_n^t, S_1^t, \ldots, S_k^t\right) = Y_1^{t+1} \\ \ldots \\ g_m\left(X_1^t, \ldots, X_n^t, S_1^t, \ldots, S_k^t\right) = Y_m^{t+1} \\ s_1\left(X_1^t, \ldots, X_n^t, S_1^t, \ldots, S_k^t\right) = S_1^{t+1} \\ \ldots \\ s_k\left(X_1^t, \ldots, X_n^t, S_1^t, \ldots, S_k^t\right) = S_k^{t+1} \end{cases} \quad (2)$$

Приведенное ниже обозначение $\vec{X}^{(\pi)}$ указывает на применение подстановки π к индексам значений $\vec{X}$: $\vec{X}^{(\pi)} = \left(X_{\pi(1)}, \ldots, X_{\pi(n)}\right)$.

Определение 3. Отображение $\hat{g}\colon\{0,1\}^{n+k} \to \{0,1\}^{m+k}$ называется эквивалентным описанием для модели $\langle g, I, n, m\rangle$ тогда и только тогда, когда существует подстановка π_X на множестве индексов аргументов функции; и подстановка π_Y на множестве индексов вектора выходных данных, для которой верно (3):

$$\forall \vec{X} \in \{0,1\}^n, \vec{S} \in \{0,1\}^k: \\ \left[\hat{g}\left(\vec{X}^{(\pi_X)}, \vec{S}^{(\pi_X)}\right)\right]^{(\pi_Y)} = g\left(\vec{X}, \vec{S}\right) \quad (3)$$

Определение 4. Модель $\langle g, I, n, m\rangle$ эквивалентна модели $\langle \hat{g}, \hat{I}, \hat{n}, \hat{m}\rangle$ при выполнении следующих условий:

1) $\hat{n} = n$;

2) $\hat{m} = m$;

3) существует такая подстановка σ на множестве состояний $\hat{S} \subseteq \{0,1\}^k$, что $\hat{g}$ является *эквивалентным* (с подстановкой π_X и π_Y) *описанием* элемента с состояниями $\sigma\left(\vec{S}\right)$, и $\hat{I} = I^{(\pi_X)}$.

Пусть задано множество $M \subseteq \{1,n\}$ и $\overline{M} = \{1,n\} \setminus M$.

Определение 5. Модель $\langle \hat{g}, \hat{K}, \hat{n}, \hat{m}\rangle$ называется составной частью модели $\langle g, K, n, m\rangle$ тогда и только тогда, когда существует $J_P = \{j_1, \ldots, j_{\hat{n}}, s_1, \ldots, s_{|\hat{K}|}\} \subseteq \{1, \ldots, n + |K|\}$ и суще-

ствует $I_P = \{i_1, \ldots, i_{\hat{m}}\} \subseteq \{1, \ldots, m + |K|\}$ такое, что для любого входного $\vec{X} \in \{0,1\}^{\hat{n}+|\hat{K}|}$ и для любого $\vec{\bar{X}} \in \{0,1\}^{n+|K|-\hat{n}-|\hat{K}|}$ выполнено: $\hat{\vec{g}}(\vec{X}) = (Y_{i_1}, \ldots, Y_{i_{\hat{m}}})$, где $\vec{Y} = (Y_{i_1}, \ldots, Y_{i_m}) = g(\vec{X}_{(J_P)}, \vec{\bar{X}}_{(\bar{J}_P)})$.

Для краткости здесь введено обозначение: $\vec{X}_{(J_P)} = (X_{J_{P_1}}, \ldots, X_{J_{P_n}})$.

Теперь рассмотрим частный случай *последовательностной логики – конвейерную микроархитектуру* (*конвейеризацию*), на основе которой ниже будет формализована задача минимизации считываемых злоумышленником данных.

Конвейеризация (см., например, [21]) – это временной параллелизм, предполагающий разбиение составной операции таким образом, чтобы в процессе выполнения одной операции начинали выполняться следующие, согласно приведённой ниже схеме:

Вектор $\vec{Z}$ назовём *вектором передаваемых данных*. Его смысл проиллюстрирован на следующем рисунке:

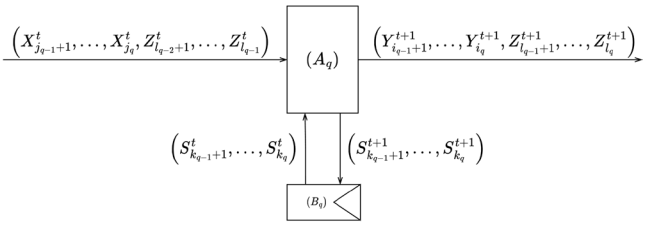

Рис. 2. Модель последовательностной логики с выделенным для передачи данных вектором $\vec{Z}$

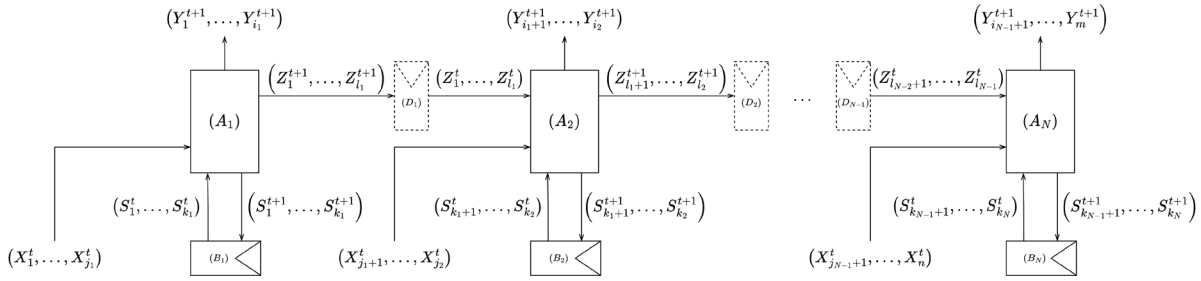

Рис. 1. Модель конвейерной микроархитектуры

Формально это определяется следующим образом.

Определение 6. *Моделью конвейерной микроархитектуры* (рис.1) это такая *последовательностная модель* $\langle g, I, n, m \rangle$, что в ней:

— вектор состояния $\vec{S} \in \{0,1\}^{|I|}$ имеет вид:

$$\left(S_1, \ldots, S_{\sum_{q=1}^{N}|I_q|}, Z_1, \ldots, Z_{\sum_{q=1}^{N-1}\alpha_q}\right);$$

— функция $g\left(\overrightarrow{X^t}, \overrightarrow{S^t}\right) = \left(\overrightarrow{Y^{t+1}}, \overrightarrow{S^{t+1}}\right)$ задана так, как показано ниже (4):

$$g\left(\overrightarrow{X^t}, \overrightarrow{S^t}\right) = \begin{bmatrix} g_1\begin{pmatrix} X^t_1, \ldots, X^t_{j_1}, \\ S^t_1, \ldots, S^t_{k_1} \end{pmatrix}, \\ \cdots \\ g_q\begin{pmatrix} X^t_{j_{q-1}+1}, \ldots, X^t_{j_q}, \\ Z^t_{l_{q-2}+1}, \ldots, Z^t_{l_{q-1}}, \\ S^t_{k_{q-1}+1}, \ldots, S^t_{k_q} \end{pmatrix}, \\ \cdots \\ g_N\begin{pmatrix} X^t_{j_{N-1}+1}, \ldots, X^t_{j_N}, \\ Z^t_{l_{N-2}+1}, \ldots, Z^t_{l_{N-1}}, \\ S^t_{k_{N-1}+1}, \ldots, S^t_{k_N} \end{pmatrix} \end{bmatrix} \quad (4)$$

Определение 7. *Вспомогательным элементом* (рис.2) конвейерной микроархитектуры является модель $g_q, I_q, n_q + \alpha_{q-1}, m_q + \alpha_q$. Где α_{q-1} – размерность вектора входной передачи данных, а $\pm_q$ – выходной. Остальные параметры обладают следующими свойствами: $l_q - l_{q-1} = \alpha_q$, $k_q - k_{q-1} = |I_q|$, $j_q - j_{q-1} = n_q$ и $i_q - i_{q-1} = m_q$. А отображение g_q имеет вид:

$$g_q\begin{pmatrix} X^t_{j_{q-1}+1}, \ldots, X^t_{j_q}, \\ Z^t_{l_{q-2}+1}, \ldots, Z^t_{l_{q-1}}, \\ S^t_{k_{q-1}+1}, \ldots, S^t_{k_q} \end{pmatrix} = \begin{pmatrix} Y^{t+1}_{i_{q-1}+1}, \ldots, Y^{t+1}_{i_q}, \\ Z^{t+1}_{l_{q-1}+1}, \ldots, Z^{t+1}_{l_q}, \\ S^{t+1}_{k_{q-1}+1}, \ldots, S^{t+1}_{k_q} \end{pmatrix} \quad (5)$$

При этом, для задания элемента A_1 полагаем $\alpha_0 = 0$ (так как у A_1 отсутствует входной вектор передаваемых данных), а для элемента A_N полагаем $\alpha_N = 0$ (так как для него отсутствует выходной вектор передаваемых данных).

Очевидно, что модель конвейерной микроархитектуры и её вспомогательные элементы обладают следующими свойствами (6):

$$\begin{cases} |I| = \sum_{q=1}^{N} |I_q| + \sum_{q=1}^{N-1} \alpha_q \\ n = \sum_{q=1}^{N} n_q \\ m = \sum_{q=1}^{N} m_q \end{cases} \quad (6)$$

Теперь мы подошли к возможности формализовать задачу, о которой шла речь во введении.

Задача минимизации собираемых злоумышленником данных. **(Задача 1)**.

Пусть задана модель конвейерной микроархитектуры g, I, n, m, составными частями которой являются g_q, I_q, n_q, m_q. На каждом такте вредоносный компонент E считывает векторы X^t, S^t, Y^t (рис.3).

Пусть $E^t = \left[X^t, S^t, Y^t \right]$, тогда считанные за T тактов данные обозначим матрицей E (7). Размерность $\vec{E}$ равна $e = n + m + |I|$, а размерность матрицы E – $\mathrm{T} \times \mathrm{e} = T \times \left(n + m + |I| \right)$.

$$E = \begin{bmatrix} \overrightarrow{E_1} \\ \dots \\ \overrightarrow{E_T} \end{bmatrix} = \begin{bmatrix} \overrightarrow{X^0} & \vec{0} & \vec{0} \\ \overrightarrow{X^1} & \overrightarrow{S^1} & \overrightarrow{Y^1} \\ \dots & \dots & \dots \\ \overrightarrow{X^{T-1}} & \overrightarrow{S^{T-1}} & \overrightarrow{Y^{T-1}} \end{bmatrix} \quad (7)$$

Как было отмечено ранее, для уменьшения потребляемых внедрённым компонентом ресурсов необходимо минимизировать считываемые данные (матрицу E). Следовательно, необходимо снизить размерность вектора $\vec{X}$ и $\vec{Z}$, а для этого нужно выделить существенные переменные булевых функций, образующих g_q.

Чтобы проиллюстрировать поведение составных элементов конвейерной микроархитектуры (рис.3), на рис.4 показана диаграмма переходов векторов входных, выходных, передаваемых данных и состояний между составными частями $g_q, I_q, n_q + \alpha_{q-1}, m_q + \alpha_q$ модели g, I, n, m.

Работа, согласно этой схеме, выглядит следующим образом. Допустим, некоторые данные поступили в A_1 на шаге $\mathrm{t} = 1$. Обработанные данные будут записаны в A_2 на шаге $t = 2$, при этом в A_1 поступят новые данные. В результате, на шаге $t = N$ вектор $\left(Y_{i_{N-1}}^N, \dots, Y_m^N \right)$ будет содержать прошедшие N шагов обработки данные $\left(X_1^1, \dots, X_{j_1}^1 \right)$. При этом на шаге $t = N$ каждый элемент A_q содержит данные, прошедшие q шагов обработки.

Таким образом осуществляется *временной параллелизм* с разбиением одной составной операции на несколько.

3. Применение вероятностных алгоритмов проверки свойств булевых функций

Для решения *Задачи 1* мы применим классические вероятностные алгоритмы проверки свойств булевых функций. А именно алгоритмы проверки существенной зависимости булевой функции от не более, чем k переменных.

Пусть: $[n] = \{1, \dots, n\}$ и $\bar{S} = [n] \setminus S$.

Определение 8. *Свойством* $\mathcal{P}$ *булевой функции* от n переменных называется множество $\mathcal{P} = \left\{ g : \{0,1\}^n \to \{0,1\} \right\}$.

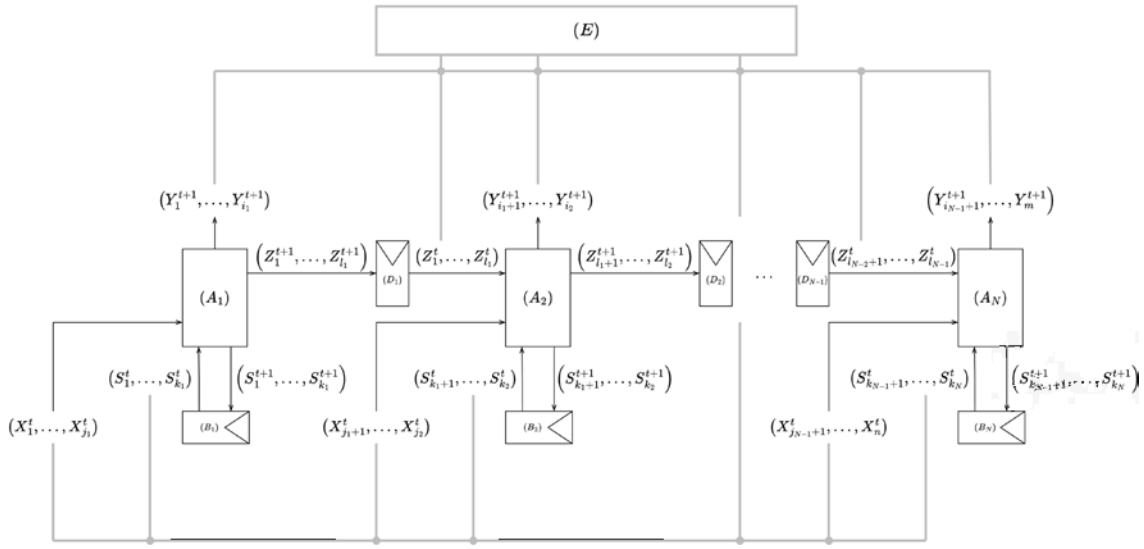

Рис. 3. Модель конвейерной микроархитектуры с внедрённым злоумышленником архитектурным компонентом

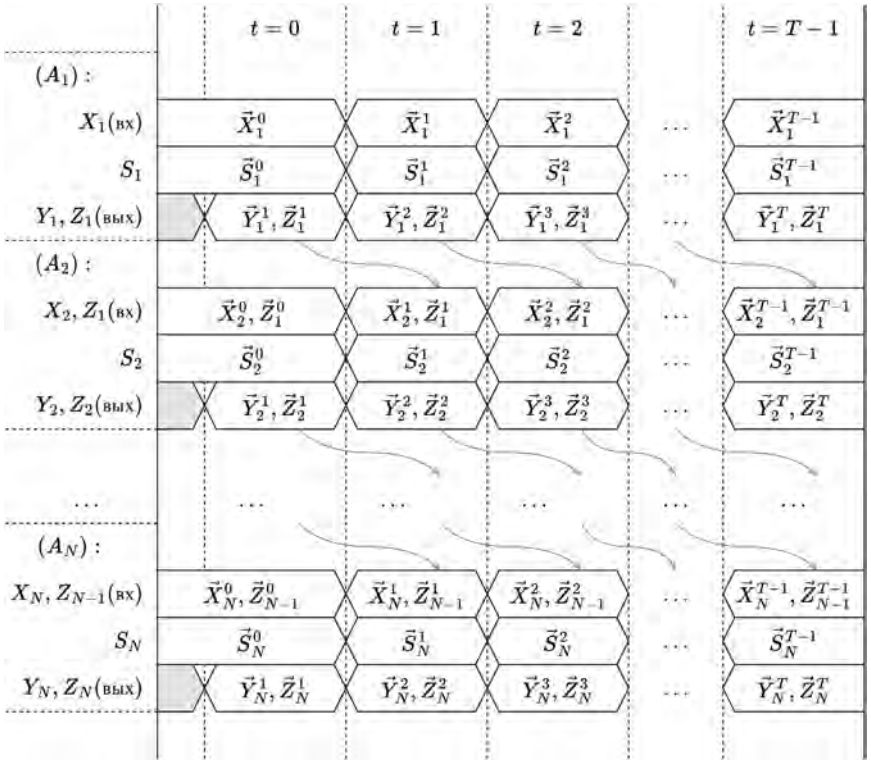

Рис. 4. Диаграмма потока данных модели конвейерной микроархитектуры

Определение 9. *Расхождением булевой функции* f *от булевой функции* g называется такое $\rho(f,g)$, что (8):

$$\rho(f,g) = P_{\forall x\in\{0,1\}^n}\left(f(x)\neq g(x)\right) = \frac{\left|\left\{f(x)\neq g(x) \;:\; x\in\{0,1\}^n\right\}\right|}{2^n} \quad (8)$$

Определение 10. *Расхождением булевой функции* f *от свойства* $\mathcal{P}$ называется такое $\rho(f,\mathcal{P})$, что (9):

$$\rho(f,\mathcal{P}) = \min_{g\in\mathcal{P}} \rho(f,g) \quad (9)$$

Вероятностный алгоритм проверки свойства $\mathcal{P}$ *булевой функции* f – это алгоритм, возвращающий ответ в форме распознавания, с параметрами:

- s – мощность множества S входных данных функции f со случайным распределением D;
- q – количество входных векторов, значение функции на которых необходимо проверить;
- μ (точность) – максимальное расстояние булевой функции от свойства.

Такой алгоритм проверки свойства $\mathcal{P}$ *«принимает»* булеву функцию f с вероятностью не менее $\frac{2}{3}$ тогда и только тогда, когда $f\in\mathcal{P}$. И *«отвергает»* булеву функцию f с вероятностью не менее $\frac{2}{3}$ тогда и только тогда, когда $\rho(f,\mathcal{P})\geq\varepsilon$. То есть достоверность результата выполнения алгоритма составляет не менее $\frac{2}{3}$.

Пусть f – булева функция, которая вычисляется в процессе работы алгоритма.

В качестве *меры сложности* используется функция, зависящая от длины входа алгоритма и результатом которой является количество вычислений заданной выше функции f.

Используется подход, который, в частности, применялся в работах [22, 23].

Определение 11. *Избыточное множество индексов* существенных переменных I – такое множество $I\subseteq[n]$, для которого верно, что $J\subseteq I$, где J – множество индексов существенных аргументов.

То есть избыточное множество индексов содержит индексы существенных переменных, однако может содержать, в том числе, индексы фиктивных.

Определение 12. Множество индексов существенных переменных I *с недостатком* – такое множество $I\subseteq[n]$, для которого верно, что $I\subseteq J$, где J – множество индексов существенных аргументов.

То есть, I не содержит индексы фиктивных переменных, однако, может содержать не все индексы существенных переменных.

Теперь в рамках сформулированной в предыдущем разделе *Задачи 1* необходимо для каждой составной части g_q, I_q, n_q, m_q выделить количество существенных аргументов булевых функций g_{q_i}, образующих g_q, и определить индексы этих аргументов.

Для определения количества существенных переменных используется следующий подход.

По заданной функции g_{q_i} выбирается минимальное по размеру подмножество индексов аргументов J так, чтобы существовала существенно зависимая от $|J|$ переменных функция h, для которой выполнено: $\rho\left(g_{q_i}, h\right) < \varepsilon$.

Для определения индексов аргументов в нашем случае нельзя использовать общепринятую формулировку вероятностного алгоритма проверки существенной зависимости функции не более чем от k переменных.

Это связано с тем, что алгоритмы, например, построенные в [22, 23] возвращают *вероятностный ответ* в форме распознавания без указания конкретного множества индексов существенных аргументов.

Поэтому предложены модификации алгоритмов из работы [22] (обозначен как А1) и [23] (обозначен как А2) с целью обеспечения этого свойства.

Опишем ниже последовательность действий алгоритма А1 с сопровождающей его процедурой П1.

Входными данными процедуры П1 является функция f и множество $S \subseteq [n]$.

Процедура П1 состоит из следующих шагов.

а) Случайно задаются два вектора $\vec{x}, \vec{y} \in \{0,1\}^n$. Распределение равномерное.

б) Формируется вектор $\vec{x} = \left(x_{\bar{S}}, y_S\right)$; если $f\left(\vec{x}\right) = f\left(\vec{x}\right)$, то результат выполнения процедуры положительный. Иначе – процедура выполнена отрицательно.

Теперь приведём шаги алгоритма А1.

а) Произвольно разбивается множество индексов аргументов $[n]$ на $\mathrm{O}\left(\mathrm{k}^2\right)$ подмножеств S_i.

б) Для всех $\mathrm{i}: O\left(k^2/\varepsilon\right)$ раз выполняется процедура П1 для f и S_i.

в) Если для не более k подмножеств S_i в предыдущем пункте получен отрицательный результат, то заданная f существенно зависит не более чем от k переменных.

Далее описан алгоритм А2 с используемой процедурой П2.

Входными данными процедуры П2 являются:

— булева функция f от n переменных;

— анализируемый вектор $\vec{x}$;

— множество S анализируемых индексов переменных;

— множества $S_1, \ldots, S_s$, которые являются разбиением $[n]$.

Процедура П2 состоит из следующих шагов.

а) Если $|S| = 1$, то процедура возвращает тот S_j, которому принадлежит индекс из S.

б) Выполняется разбиение S на два множества $\widehat{S_1}$ и $\widehat{S_2}$. Их мощности, соответственно, равны $\left\lfloor \frac{|S|}{2} \right\rfloor$ и $\left\lceil \frac{|S|}{2} \right\rceil$.

в) Составляется вектор $\vec{y} = \vec{x}$ с инверсией битов множества $\widehat{S_2}$.

г) Если $f\left(\vec{x}\right) \neq f\left(\vec{y}\right)$, то выполняется процедура П2 от f, $\vec{x}$, $\widehat{S_2}$ и разбиения $S_1, \ldots, S_s$.

д) Иначе, выполняется процедура П2 от f, $\vec{z}$, $\widehat{S_1}$ и разбиения $S_1, \ldots, S_s$. За вектор $\vec{z}$ принят $\vec{x}$ с инверсией битов множества S.

Шаги алгоритма А2 приведены ниже.

а) Инициализируются переменные $S \leftarrow [n]$ и $\psi \leftarrow 0$.

б) Индексы $[n]$ случайно разбиваются на s подмножеств: $S_1, \ldots, S_s$.

1) Выполняется цикл из r итераций, в котором создаётся случайная пара $\left(\vec{x}, \vec{y}\right) \in \{0,1\}^n \times \{0,1\}^n$ и составляется вектор $\vec{x} = \left(x_{\bar{S}}, y_S\right)$.

2) Проверяется $f\left(\vec{x}\right) = f\left(\vec{x}\right)$ – если верно, то выполняется переход к следующей итерации цикла; если ложно – то к следующему пункту.

3) Выполняется процедура П2 от f, $\vec{x}$, $\left\{i : x_i \neq \widehat{x_i}\right\}$ (множества анализируемых индексов) и разбиения $S_1, \ldots, S_s$. Результат процедуры – I_j.

4) Обновляется переменная $S \leftarrow S \setminus I_j$ и $\psi \leftarrow \psi + 1$.

5) Если счётчик E превысил значение k, то алгоритм «отвергает» f.

в) В случае успешного прохождения цикла алгоритм «принимает» f как функцию, существенно зависящую не более чем от k переменных. Иначе – «отвергает».

Рассмотрим модифицированный алгоритм А1*, который включает изменённую процедуру П1*.

Входными данными процедуры П1* является функция f и множество $S \subseteq [n]$. Процедура П1* состоит из следующих шагов.

а) Случайно задаются два вектора $\vec{x}, \vec{y} \in \{0,1\}^n$. Распределение равномерное.

б) Формируется вектор $\vec{\tilde{x}} = \left(x_{\bar{S}}, y_S\right)$; если $f\left(\vec{x}\right) = f\left(\vec{\tilde{x}}\right)$, то процедура выполнена успешно.

в) (Модификация) в случае отрицательного результата, полученного в пункте (б), процедура возвращает отказ и множество индексов $\left\{i : \; x_i \neq \widehat{x_i}\right\}$.

Шаги алгоритма А1*:

а) Произвольно разбивается множество индексов аргументов $[n]$ на $O(k^2)$ подмножеств S_i.

б) Для всех $i: O(k^2/\varepsilon)$ раз выполняется процедура П1* для f и S_i.

в) Если для не более k подмножеств S_i в предыдущем пункте получен отрицательный результат, то заданная f существенно зависит не более чем от k переменных.

г) (Модификация) если функция f «принята», то возвращается множество $I = \bigcup_i I_i$ – объединение результатов процедуры П1*, которые были не пройдены для S_i.

Таким образом, алгоритм А1* возвращает множество индексов аргументов функции f в случае её «принятия».

Ниже рассмотрена модификация алгоритма А2 – А2.1*. Она включает процедуру П2.1*, набор входных данных и действий которой аналогичен процедуре П2.

Шаги модифицированного алгоритма А2.1* приведены ниже.

а) Инициализируются переменные $S \leftarrow [n]$ и $\psi \leftarrow 0$.

б) Индексы $[n]$ случайно разбиваются на s подмножеств: $S_1, \ldots, S_s$.

1) Выполняется цикл из r итераций, в котором создаётся случайная пара $(\vec{x}, \vec{y}) \in \{0,1\}^n \times \{0,1\}^n$ и составляется вектор $\vec{\hat{x}} = (\vec{x}_{\bar{S}}, \vec{y}_S)$.

2) Проверяется $f(\vec{x}) = f(\vec{\hat{x}})$ – если верно, то выполняется переход к следующей итерации цикла; если ложно – то к следующему пункту.

3) Выполняется процедура П2.1* от f, $\vec{x}$, $\{i: x_i \neq \hat{x}_i\}$ (множества анализируемых индексов) и разбиения $S_1, \ldots, S_s$. Результат процедуры – I_j.

4) Обновляется переменная $S \leftarrow S \setminus I_j$ и $\psi \leftarrow \psi + 1$.

5) Если счётчик ψ превысил значение k, то алгоритм «отвергает» f.

в) В случае успешного прохождения цикла алгоритм «принимает» f как функцию, существенно зависящую не более чем от k переменных.

г) (Модификация 1) для «принятой» функции алгоритм возвращает множество $[n] \setminus S$.

Вторая модификация алгоритма А2 (обозначим её А2.2*) состоит в следующем.

В А2.2* в пункт (а) процедуры П2.1* добавлен возврат индекса из множества S. Пункт (в.4) алгоритма А2.1* дополнен сохранением результата процедуры во множество I (которое инициализируется пустым множеством). В пункте (г) добавлен возврат полученного множества I.

С применённой модификацией алгоритм А2.1* вернёт избыточное множество индексов существенных переменных функции f, а алгоритм А2.2* – множество индексов с недостатком.

В табл.1 приведены характеристики рассмотренных в данном разделе алгоритмов.

Колонка «*Выделение множества*» указывает, реализована ли алгоритмом выдача множества индексов существенных переменных. В колонке «Множество индексов» отмечена характеристика множества индексов. «Множество индексов» может быть указано только для тех алгоритмов, которые реализуют «Выделение множества».

Таблица 1

Характеристики рассмотренных алгоритмов

Алгоритм	Выделение множества	Множество индексов
А1	Нет	
А2	Нет	
А1*	Да	С избытком
А2.1*	Да	С избытком
А2.2*	Да	С недостатком

4. Анализ результатов работы модифицированных алгоритмов

Для проведения численного эксперимента модифицированные алгоритмы была реализованы программно. Установлены следующие параметры:

- количество аргументов булевой функции – $n = 14$;
- множество аргументов $[n]$ разбито на $s = n/2$ подмножеств;
- количество проверяемых существенных переменных – $k = 10$;
- булевых функций – 128;
- если в алгоритме есть цикл, то он выполняется 1000 итераций.

Эксперимент состоит из следующих шагов.

а) Произвольно создаётся 128 булевых функций от 14 переменных со случайным выбором фиктивных и существенных аргументов;

б) Для каждой функции каждый алгоритм испытывается 1000 раз;

в) Количество обращений к булевой функции и полученные результаты группируются по количеству существенных аргументов созданных в (а) функций и усредняются.

Пусть k – проверяемое количество существенных переменных.

Сложность алгоритма А1* совпадает со сложностью А1 и равна: $O(k^4 \cdot \log(k+1)/\varepsilon)$.

Таблица 2

Сравнительная характеристика сложностей модифицированных алгоритмов

Алгоритм	Количество существенных аргументов	Обращений к функции	Аналитическая Сложность
A1*	0	14000	$O\left(k^4 \cdot \log(k+1)/\varepsilon\right)$
	2	10299	
	4	7038	
	6	3839	
	8	2574	
	10	1043	
	12	404	
	14	38	
A2.1*	0	2000	$O\left(k \cdot \log k + k/\varepsilon\right)$
	2	2019	
	4	2031	
	6	2041	
	8	1910	
	10	1057	
	12	429	
	14	71	
A2.2*	0	2000	$O\left(k \cdot \log n + k/\varepsilon\right)$
	2	2030	
	4	2052	
	6	2071	
	8	1945	
	10	1096	
	12	470	
	14	112	

Сложность алгоритма A2.1*: $O\left(k \log k + k/\varepsilon\right)$.

Сложность алгоритма A2.2* равна

$$O\left(2r + r \cdot \log|S|\right) = O\left(\frac{k}{\varepsilon} + k \cdot \log n\right). \quad (10)$$

В результате проведённого анализа получена сравнительная характеристика сложностей модифицированных алгоритмов (табл.2) и результатов их работы (табл.3). Графическое представление зависимости перечисленных характеристик от количества существенных аргументов произвольной булевой функции показано на рис.5.

Выполненный анализ показывает, что алгоритм A1* уступает A2.1* и A2.2* по количеству обращений к булевой функции.

По количеству верно и ложно определённых индексов существенных переменных он расположен между алгоритмами A2.1* и A2.2*.

Алгоритм A2.1* имеет наибольший показатель ошибочно выделенных индексов. При этом он обладает самым низким, по сравнению с A1* и A2.2*, количеством пропущенных существенных аргументов.

Результат выполнения алгоритма A2.2* характеризуется отсутствием ошибочно выделенных аргументов. С другой стороны, показатель пропущенных существенных индексов аргументов превышает показатели других рассмотренных в рамках сравнительной характеристики алгоритмов.

На практике можно использовать аналитическую программу на основе линейной комбинации приведенных алгоритмов. Можно также построить адаптивную процедуру, которая подстраивается под свойства анализируемого устройства.

5. Заключение

В настоящей работе исследованы возможности повышения степени защиты аппаратных средств путём определения участков схемы, находящихся под угрозой нарушения конфиденциальности.

Таблица 3

Сравнительная характеристика работы модифицированных алгоритмов

Алгоритм	Количество существенных аргументов	Выделено верно	Выделено ошибочно	Не выделено
A1*	0	0.00	0.00	0.00
	2	1.91	0.86	0.09
	4	3.63	1.49	0.37
	6	5.38	2.09	0.62
	8	6.50	1.72	1.50
	10	7.68	1.50	2.32
	12	8.53	0.82	3.47
	14	9.32	0.00	4.68
A2.1*	0	0.00	0.00	0.00
	2	2.00	1.71	0.00
	4	3.80	3.00	0.20
	6	6.00	4.20	0.00
	8	8.00	3.47	0.00
	10	10.00	3.00	0.00
	12	12.00	1.64	0.00
	14	14.00	0.00	0.00
A2.2*	0	0.00	0.00	0.00
	2	1.86	0.00	0.14
	4	3.40	0.00	0.60
	6	5.10	0.00	0.90
	8	5.73	0.00	2.27
	10	6.50	0.00	3.50
	12	6.82	0.00	5.18
	14	7.00	0.00	7.00

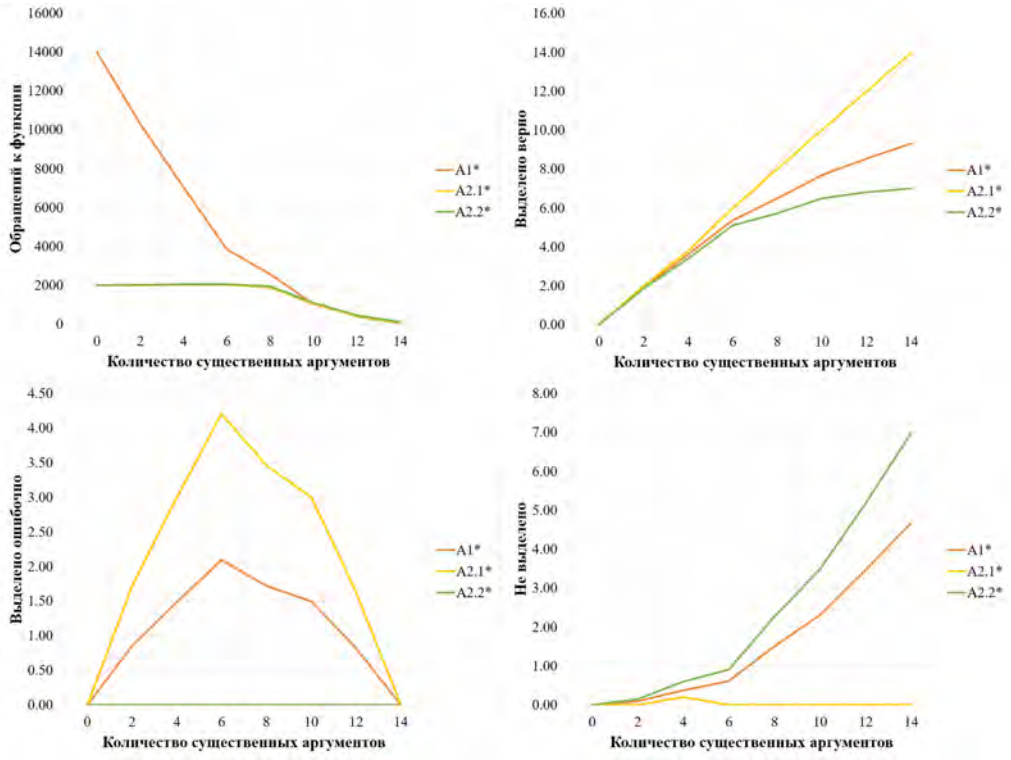

Рис. 5. Графическое представление зависимости характеристик модифицированных алгоритмов от количества существенных переменных

Построена конкретная модель конвейерной микроархитектуры на основе математической модели комбинационной и последовательностной логики.

Рассмотрена задача минимизации размерности для данных, которые аккумулируются в схеме. Для решения этой задачи проанализированы некоторые применяемые ранее вероятностные алгоритмы проверки свойств булевых функций, в частности проверки существенной зависимости функции от не более чем k переменных.

Предложены модификации этих алгоритмов и процедур для решения задачи 1 (см. раздел 3). Проведён анализ сложности и количества определяемых индексов при помощи изменённых алгоритмов.

Полученный в данной работе результат применим для обнаружения входов-выходов элементов комбинационной и последовательностной схемы, считывание данных с которых представляют угрозу нарушения конфиденциальности обрабатываемой информации в устройстве, использующем ППВМ.

Литература

1. Антонов А. А., Барабанов А. В., др. Цифровой синтез: практический курс / под общ. ред. А.Ю. Романова, Ю.В. Панчула. – М.: ДМК Пресс, 2020. – 556 с.
2. Xue M., Gu C., Liu W., et al. Ten years of hardware Trojans: a survey from the attacker's perspective. // IET Computers & Digital Techniques. 14. pp. 231-246. 2020. DOI: 10.1049/iet-cdt.2020.0041
3. Wan Z., Yu B., et al. A Survey of FPGA-Based Robotic Computing. // IEEE Circuits and Systems Magazine, 21, pp. 48-74. 2020. DOI: 10.1109/MCAS.2021.3071609.
4. Quraishi M., Tavakoli E., Ren F. A Survey of System Architectures and Techniques for FPGA Virtualization. // IEEE Transactions on Parallel & Distributed Systems, vol. 32, no. 09, pp. 2216-2230. 2021. DOI: 10.1109/TPDS.2021.3063670
5. Туринцев К. А., Поплавский Д. А., Калинкина А. А. Аппаратное средство для ускорения решения задач дизассемблирования // МОЛОДЫЕ УЧЁНЫЕ РОССИИ: сборник статей XVII Всероссийской научно-практической конференции. – Пенза: Наука и Просвещение, 2023. С. 32-37.
6. Zhang J., Qu G. Recent Attacks and Defenses on FPGA-based Systems. ACM Transactions on Reconfigurable Technology Syst. 12, 3, Article 14 (2019), p. 24. 2019. DOI: 10.1145/3340557
7. Деменкова Т. А., Певцов Е. Ф. Аппаратно-программные ресурсы защиты интегральных схем и интеллектуальных систем // Научно-технический вестник Поволжья. 2018. № 12. С. 213-218.
8. Huang H., Shen H., Li S., et al. A Hardware Trojan Trigger Localization Method in RTL based on Control Flow Features. 2022 IEEE 31st Asian Test Symposium (ATS), Taichung City, Taiwan, pp. 138-143. 2022. DOI: 10.1109/ATS56056.2022.00036
9. Palumbo A., Cassano L., Luzzi B., Hernandez J., et al. Is your FPGA bitstream Hardware Trojan-free? Machine learning can provide an answer. // Journal of Systems Architecture. Volume 128. 2022. DOI: 10.1016/j.sysarc.2022.102543
10. Chithra C., Kokila J., Ramasubramanian N. Detection of Hardware Trojans using Machine Learning in SoC FPGAs // 2020 IEEE International Conference on Electronics, Computing and Communication Technologies, Bangalore, India. pp. 1-7. 2020. DOI: 10.1109/CONECCT50063.2020.9198475
11. Zhang L., Dong Y., Wang J., et al. A hardware Trojan detection method based on the electromagnetic leakage. China Communications, vol. 16, no. 12, pp. 100-110. 2019. DOI: 10.23919/JCC.2019.12.007
12. Yu S., Gu C., Liu W., et al. A Novel Feature Extraction Strategy for Hardware Trojan Detection 2020 IEEE International Symposium on Circuits and Systems, Seville, Spain, 2020, pp. 1-5. DOI: 10.1109/ISCAS45731.2020.9180479
13. Cruz J., Posada C., Masna N., et al. A Framework for Automated Exploration of Trojan Attack Space in FPGA Netlists. IEEE Transactions on Computers. pp. 1-12. 2023. DOI: 10.1109/TC.2023.3266592
14. Ahmed Q., Platzner M. On the Detection and Circumvention of Bitstream-level Trojans in FPGAs // 2022 IEEE Computer Society Annual Symposium on VLSI , Nicosia, Cyprus. pp. 434-439. 2022. DOI: 10.1109/ISVLSI54635.2022.00097
15. Yang J., Zhang Y., Hua Y., et al. Hardware Trojans Detection Through RTL Features Extraction and Machine Learning. 2021 Asian Hardware Oriented Security and Trust Symposium (AsianHOST), Shanghai, China, pp. 1-4, 2021. DOI: 10.1109/AsianHOST53231.2021.9699658
16. Zhang Q., Liu L., Yuan Y., et al. A Gate-Level Information Leakage Detection Framework of Sequential Circuit Using Z3. // Electronics, 11, 4216. 2022.
17. Matrosova A., Provkin V. Applying Incompletely Specified Boolean Functions for Patch Circuit Generation // 2021 IEEE East-West Design & Test Symposium. Batumi, Georgia. pp. 1-4 2021. DOI: 10.1109/EWDTS52692.2021.9581029
18. Sabri M., Shabani A., Alizadeh B. SAT-Based Integrated Hardware Trojan Detection and Localization Approach Through Path-Delay Analysis // IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 68, no. 8. pp. 2850-2854. 2021. DOI: 10.1109/TCSII.2021.3074549
19. Xie Z., Daowen Q., Guangya C., et al. Testing Boolean Functions Properties. FundamFundamenta Informaticae 182(4). pp. 321-344. 2021. DOI: 10.3233/FI-2021-2076
20. De A., Mossel E., Neeman J. Junta Correlation is Testable // 2019 IEEE 60th Annual Symposium on Foundations of Computer Science, Baltimore, MD, USA. pp. 1549-1563. 2019. DOI: 10.1109/FOCS.2019.00090
21. Харрис Д. М., Харрис С. Л. Цифровая схемотехника и архитектура компьютера. / пер. с англ. Imagination Technologies. – М.: ДМК Пресс, 2018. – 792 с.: цв. ил.
22. Liu Z., Chen X., Servedio R., et al. Distribution-free Junta Testing. ACM Trans. Algorithms 15, 1, Article 1, 23 p. 2018. DOI: 10.1145/3264434
23. Bshouty N. Almost optimal distribution-free junta testing // 34th Computational Complexity Conference, CCC 2019, NJ, USA. pp. 2:1–2:13, 2019. DOI: 10.4230/LIPIcs.CCC.2019.2

ON THE ONE ALGORITHMS CLASS APPLICABILITY FOR THE COMPONENTS BEHAVIOR ANALYSIS OF DEVICES WITH FIELD-PROGRAMMABLE GATE ARRAYS

Titov A.S.[7], Gordeev E.N.[8]

The research purpose: *research of opportunities to increase the hardware security using detection of the register-transfer level schema parts that are at confidentiality violation risk.*

Methods: *the use of register-transfer level scheme mathematical modeling and the application of classical probabilistic algorithms for Boolean functions property testing to the model in order to locate potentially vulnerable areas in the microcircuit internal logic.*

Results: *based on the combinational and sequential circuits mathematical model, which defines the internal logic via Boolean function sets, a concrete pipeline microarchitecture model with split states and data transfer has been developed, which allows the one to research using chosen mathematical apparatus further.*

Singled out the processed by special computing devices data confidentiality intruder model. For the specific pipeline microarchitecture model and the confidentiality violator, that can be placed at any production process stage, the accumulated from the schema data dimension minimization problem has been considered.

In the context of the researching model, the complexity analysis of the one algorithms class has been performed. Based on the results of the analysis, modifications of some of the algorithms are proposed.

The constructed algorithms make it possible to find the location of input and output pins that are potentially vulnerable to sequential and combinational circuits elements confidentiality violations, by determining the indices of arguments of Boolean functions.

The scientific novelty: *consists in the one probabilistic algorithms class applicability analysis to the detection vulnerable schema logic device areas problem and in based on the algorithm's modification implementation in the purpose of vulnerable elements input pins detection accuracy increasing.*

Keywords: *complex programmable logic device, register-transfer level, pipeline microarchitecture, intruder model, Boolean functions, Boolean circuits, actual arguments.*

References

1. Antonov A. A., Barabanov A. V., dr. Cifrovoj sintez: prakticheskij kurs / pod obshh. red. A.Ju. Romanova, Ju.V. Panchula. – M.: DMK Press, 2020. – 556 s.
2. Xue M., Gu C., Liu W., et al. Ten years of hardware Trojans: a survey from the attacker's perspective. // IET Computers & Digital Techniques. 14. pp. 231-246. 2020. DOI: 10.1049/iet-cdt.2020.0041
3. Wan Z., Yu B., et al. A Survey of FPGA-Based Robotic Computing. // IEEE Circuits and Systems Magazine, 21, pp. 48-74. 2020. DOI: 10.1109/MCAS.2021.3071609.
4. Quraishi M., Tavakoli E., Ren F. A Survey of System Architectures and Techniques for FPGA Virtualization. // IEEE Transactions on Parallel & Distributed Systems, vol. 32, no. 09, pp. 2216-2230. 2021. DOI: 10.1109/TPDS.2021.3063670
5. Turincev K. A., Poplavskij D. A., Kalinkina A. A. Apparatnoe sredstvo dlja uskorenija reshenija zadach dizassemblirovanija // MOLODYE UChJoNYE ROSSII: sbornik statej XVII Vserossijskoj nauchno-prakticheskoj konferencii. – Penza: Nauka i Prosveshhenie, 2023. S. 32-37.
6. Zhang J., Qu G. Recent Attacks and Defenses on FPGA-based Systems. ACM Transactions on Reconfigurable Technology Syst. 12, 3, Article 14 (2019), p. 24. 2019. DOI: 10.1145/3340557
7. Demenkova T. A., Pevcov E. F. Apparatno-programmnye resursy zashhity integral'nyh shem i intellektual'nyh sistem // Nauchno-tehnicheskij vestnik Povolzh'ja. 2018. № 12. S. 213-218.
8. Huang H., Shen H., Li S., et al. A Hardware Trojan Trigger Localization Method in RTL based on Control Flow Features. 2022 IEEE 31st Asian Test Symposium (ATS), Taichung City, Taiwan, pp. 138-143. 2022. DOI: 10.1109/ATS56056.2022.00036
9. Palumbo A., Cassano L., Luzzi B., Hernandez J., et al. Is your FPGA bitstream Hardware Trojan-free? Machine learning can provide an answer. // Journal of Systems Architecture. Volume 128. 2022. DOI: 10.1016/j.sysarc.2022.102543

7 Anatoliy S. Titov, student of the «Information Security» department, Bauman Moscow State Technical University, Moscow, Russia. E-mail: toliakpurple@gmail.com

8 Eduard N. Gordeev, Dr.Sc. (Math.), Professor of the «Information Security» department, Bauman Moscow State Technical University, Moscow, Russia. E-mail: werhorn@yandex.ru

10. Chithra C., Kokila J., Ramasubramanian N. Detection of Hardware Trojans using Machine Learning in SoC FPGAs // 2020 IEEE International Conference on Electronics, Computing and Communication Technologies, Bangalore, India. pp. 1-7. 2020. DOI: 10.1109/CONECCT50063.2020.9198475
11. Zhang L., Dong Y., Wang J., et al. A hardware Trojan detection method based on the electromagnetic leakage. China Communications, vol. 16, no. 12, pp. 100-110. 2019. DOI: 10.23919/JCC.2019.12.007
12. Yu S., Gu C., Liu W., et al. A Novel Feature Extraction Strategy for Hardware Trojan Detection 2020 IEEE International Symposium on Circuits and Systems, Seville, Spain, 2020, pp. 1-5. DOI: 10.1109/ISCAS45731.2020.9180479
13. Cruz J., Posada C., Masna N., et al. A Framework for Automated Exploration of Trojan Attack Space in FPGA Netlists. IEEE Transactions on Computers. pp. 1-12. 2023. DOI: 10.1109/TC.2023.3266592
14. Ahmed Q., Platzner M. On the Detection and Circumvention of Bitstream-level Trojans in FPGAs // 2022 IEEE Computer Society Annual Symposium on VLSI , Nicosia, Cyprus. pp. 434-439. 2022. DOI: 10.1109/ISVLSI54635.2022.00097
15. Yang J., Zhang Y., Hua Y., et al. Hardware Trojans Detection Through RTL Features Extraction and Machine Learning. 2021 Asian Hardware Oriented Security and Trust Symposium (AsianHOST), Shanghai, China, pp. 1-4, 2021. DOI: 10.1109/AsianHOST53231.2021.9699658
16. Zhang Q., Liu L., Yuan Y., et al. A Gate-Level Information Leakage Detection Framework of Sequential Circuit Using Z3. // Electronics, 11, 4216. 2022.
17. Matrosova A., Provkin V. Applying Incompletely Specified Boolean Functions for Patch Circuit Generation // 2021 IEEE East-West Design & Test Symposium. Batumi, Georgia. pp. 1-4 2021. DOI: 10.1109/EWDTS52692.2021.9581029
18. Sabri M., Shabani A., Alizadeh B. SAT-Based Integrated Hardware Trojan Detection and Localization Approach Through Path-Delay Analysis // IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 68, no. 8. pp. 2850-2854. 2021. DOI: 10.1109/TCSII.2021.3074549
19. Xie Z., Daowen Q., Guangya C., et al. Testing Boolean Functions Properties. FundamFundamenta Informaticae 182(4). pp. 321-344. 2021. DOI: 10.3233/FI-2021-2076
20. De A., Mossel E., Neeman J. Junta Correlation is Testable // 2019 IEEE 60th Annual Symposium on Foundations of Computer Science, Baltimore, MD, USA. pp. 1549-1563. 2019. DOI: 10.1109/FOCS.2019.00090
21. Harris D. M., Harris S. L. Cifrovaja shemotehnika i arhitektura komp'jutera. / per. s angl. Imagination Technologies. – M.: DMK Press, 2018. – 792 s.: cv. il.
22. Liu Z., Chen X., Servedio R., et al. Distribution-free Junta Testing. ACM Trans. Algorithms 15, 1, Article 1, 23 p. 2018. DOI: 10.1145/3264434
23. Bshouty N. Almost optimal distribution-free junta testing. // 34th Computational Complexity Conference, CCC 2019, NJ, USA. pp. 2:1–2:13, 2019. DOI: 10.4230/LIPIcs.CCC.2019.2

ПРИМЕНЕНИЕ ЛОГИКО-ВЕРОЯТНОСТНОГО МЕТОДА В ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ (ЧАСТЬ 2)

Калашников А.О.[1], Бугайский К.А.[2], Аникина Е.В.[3], Перескоков И.С.[4], Петров Ан.О.[5], Петров Ал.О.[6], Храмченкова Е.С.[7], Молотов А.А.[8]

Цель исследования: *адаптация логико-вероятностного метода оценивания сложных систем к задачам построения систем защиты информации в многоагентной системе.*

Метод исследования: *при проведении исследования использовались основные положения методологии структурного анализа, системного анализа, теории принятия решений, методов оценивания событий при условии неполной информации, логико-вероятностных методов.*

Полученный результат: *данная статья продолжает рассмотрение вопросов информационной безопасности на основе анализа отношений между субъектами и объектом защиты. Обосновано представление субъекта и объекта защиты в виде интеллектуального агента с учетом требований по защите информации. Даны формальные определения агента информационной безопасности и его основных характеристик: информационный ресурс, информационный поток и права доступа субъекта. Показано, что понятие агента информационной безопасности представляет собой основу для описания структур в информационной системе. Разработана аксиоматика отношений субъекта и объекта как агентов информационной безопасности, а также отношений между информационными ресурсами и информационными потоками внутри агента. Показана возможность определения состояния агента на основе формируемых в процессе его функционирования событий и сообщений.*

Научная новизна: *рассмотрение вопросов защиты информации с использованием аппарата математических и логических отношений. Разработка формальных определений агента информационной безопасности и составляющих его информационных ресурсов и информационных потоков, являющихся базовыми универсальными компонентами описания структур в информационной системе. Определение понятия агента информационной безопасности за счет рассмотрения отображения субъекта и его целеполагания на объект.*

Вклад авторов: Калашников А. О. *выполнил постановку задачи и общую разработку модели применения логико-вероятностного метода в информационной безопасности;* ***Бугайский К.А., Аникина Е.В.*** *разработали модель описания проблем информационной безопасности через отображение субъекта и его целеполагания на объект, а также типы и аксиоматику отношений агентов;* ***Перескоков И.С*** *и* ***Петров Андрей О.*** *разработали модель отображения субъекта на объект;* ***Петров Александр О.*** *и* ***Храмченкова Е.С.*** *разработали модель отображения целеполагания субъекта на объект;* ***Молотов А.А.*** *разработал модель формирования событий и сообщений агента.*

Ключевые слова: *модель информационной безопасности, оценка сложных систем, теория отношений, системный анализ, многоагентная система.*

DOI:10.21681/2311-3456-2023-5-113-127

1 Калашников Андрей Олегович, доктор технических наук, главный научный сотрудник лаборатории «Сложных сетей» ФГБУН Институт управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: aokalash@ipu.ru

2 Бугайский Константин Алексеевич, младший научный сотрудник Института проблем управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: kabuga@ipu.ru

3 Аникина Евгения Владимировна, научный сотрудник Института проблем управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: ajanet@ipu.ru

4 Перескоков Илья Сергеевич, младший научный сотрудник Института проблем управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: pereskokov@phystech.edu

5 Петров Андрей Олегович, младший научный сотрудник Института проблем управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: petrovaojob@gmail.com

6 Петров Александр Олегович, младший научный сотрудник Института проблем управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: petrovalexandr@ipu.ru

7 Храмченкова Екатерина Сергеевна, младший научный сотрудник Института проблем управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: hramchenkovaes@yandex.ru

8 Молотов Александр Анатольевич, инженер-программист Института проблем управления имени В.А. Трапезникова РАН, г. Москва, Россия. E-mail: alpha.sphere@ya.ru

Введение

Данная статья является второй из серии публикаций, посвященных исследованию вопроса применения логико-вероятностного метода при изучении вопросов защиты информации. Метод был разработан Рябининым И.А. [1, 21]. Метод получил высокую популярность при проведении исследований, связанных с анализом и оценкой сложных систем. Прежде всего для решения вопросов надежности работы систем и причин возникновения аварийных ситуаций. Логико-вероятностный метод предполагает решение следующих задач.

1. Построение структурно-логической модели системы за счет выделения и использования событий с несовместными исходами.

2. Проведение преобразований полученных логических уравнений на основе функций булевой алгебры с целью получения системы уравнений с конечным числом переменных.

3. Теоретически обоснованный переход от уравнений булевой алгебры к уравнениям с вероятностными переменными.

К несомненным достоинствам логико-вероятностного метода следует отнести его способность обеспечить прозрачность процедур анализа и оценки сложных систем, а также хорошие адаптационные способности к новым задачам. Результатом применения логико-вероятностного метода являются количественные оценки риска как вероятности нарушения работоспособности системы. Интерес к логико-вероятностному методу – помимо типичных вопросов надежности систем, – в настоящее время подкрепляется исследованием задач машинного обучения и связанных с ними проблем оптимизации расчетов [см., например, 2-5]. В частности, логико-вероятностный метод обеспечивает хорошую точность и стабильность результатов в задачах распознавания объектов. Логико-вероятностный метод также находит свое применение при решении задач защиты информации [см., например, 6-11].

Тем не менее, представляется, что логико-вероятностный метод обладает значительно большим, пока не раскрытым, потенциалом в случае его дальнейшего развития и адаптации к решению задач в области информационной безопасности (далее – ИБ).

Постановка задачи

Современные информационные системы (далее – ИС) [12, 13] отличаются большим разнообразием обрабатываемой информации, сложными типами связей между аппаратными и программными компонентами, распределенным характером обработки и управления информацией и компонентами ИС. Что с большой вероятностью влечет за собой проблему экспоненциального взрыва при непосредственном использования для описания структурно-логических схем ИС функций алгебры логики в рамках логико-вероятностного метода. Вместе с тем, логико-вероятностный метод содержит теоретические положения, позволяющие заместить систему логических равенств описывающих структурно-логическую схему одним равенством.

В рамках достижения общей цели исследования (адаптации логико-вероятностного метода для решения задач ИБ) в настоящей статье разработаны формально-логические основы для определения и последующего выделения фрактальных структур, присущих ИС как сложной системе. Для решения этой задачи выделяется макроуровень ИС, состоящий из агентов информационной безопасности и проводится рассмотрение отношений между ними.

Определение агента ИБ

Объектом в отношениях «субъект-объект» для Защитника и Нарушителя, определенных в первой части настоящей работы, является ИС, представленная в виде графа $G(V,E)$. На основании [14, 15] можно сказать, что выполнение действий субъектом невозможно без функционала, обеспечивающего отображение как целей субъекта на объект, так и собственно субъекта на объект. Исходя из того, что в каждый конкретный момент времени субъект взаимодействует с определенной компонентой ИС, то в качестве объекта в отображениях следует принять узел ИС. Сам факт существования указанных отображений позволяет при рассмотрении вопросов защиты информации предположить условное наличие «ограниченной субъектности» узлов ИС.

Рассмотрим эти отображения на основе категориального подхода и с учетом положений и выводов, изложенных в [13, 16, 17, 18]

Отображение субъекта на объект. Существующая парадигма ИБ предполагает рассмотрение этого отношения с точки зрения возможностей (прав) субъекта *AS* по использованию ресурсов объекта – ИС, представленной графом $G(V,E)$. Каждый узел графа описывается ресурсами $V[\![Res]\!]$, которые могут быть представлены набором, состоящим из перечня данных, обрабатывающих их программ и кон-

фигураций, обеспечивающих правила обработки $Res = \{Data, Prog, Conf\}$. В рамках существующих архитектурных решений операционных систем компонент ИС субъект может быть представлен только в виде аккаунта *AC* пользователя операционной системы данного узла, который манипулирует данными $Data$ с помощью программ $Prog$. Поскольку ресурсы заданы изначально, это означает, что для функции распределения ресурсы узла являются областью определения, а аккаунт – областью назначения. При этом данные могут рассматриваться как область определения для программ. Если ввести отношение «больше» в описание узла $V[\![Data, Prog, AC, >]\!]$, то получаем выражение для отображения субъекта на объект:

$$AS \xrightarrow{Res} V : Data \to Prog \to AC \to AS \qquad (1).$$

Деятельность субъекта в рамках аккаунта *AC* и функционирование программ *Prog* формируют события и сообщения *ME* на узле. Для ИС и ее компонент целесообразно рассматривать сообщения и ошибки как часть данных узла: $ME \subset Data$. Что позволяет определить отношения $ME \to AC \wedge ME \to Prog$. Коммутационная диаграмма отображения субъекта на объект приведена на рисунке 1.

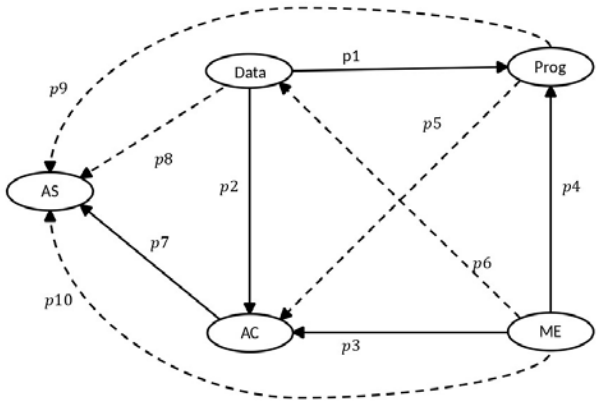

Рис. 1. Диаграмма отображения субъекта на объект

На этом и последующих рисунках сплошными стрелками обозначены отношения явно сформулированные при описании отображения. Пунктирные стрелки обозначают отношения, являющиеся следствием коммутативности диаграммы.

Отображение целей субъекта на объект. В общем случае при нормальном функционировании узел ИС выполняет правила, которые определяются субъектом и по сути являются отображением его целей на графе $G(V, E)$. Для Защитника целеполагание может быть представлено как набор отдельных целей достижимых на различных узлах $\overline{GS} = \bigcup_V \overline{g_v}$, $v \in V$. В случае успешных действий Нарушителя узел может также реализовывать и задаваемые им цели – $GS = \bigcup_V g_v$, $v \in V$. Обозначим итоговое целеполагание узла $GG = (\overline{g_v} \wedge g_v)$. Положим, что реализация целей g_v и $\overline{g_v}$ требует выполнения определенных правил функционирования ресурсов узла – *Conf*. Целесообразно считать, что целеполагание является первоочередным в деятельности субъекта и у него может быть несколько целей. Кроме того, положим, что для достижения конкретной цели субъект может использовать различные комбинации правил функционирования программ и аккаунта узла – конфигурации. Следовательно, можно записать отображение как:

$$AS \xrightarrow{GG} V : GG \to AS \wedge Conf \to GG \qquad (2).$$

С точки зрения архитектуры современных ИС и их компонент, узел графа $G(V, E)$ должен предоставлять пути доступа (*AP*) к своим ресурсам. Под путями доступа будем понимать различные комбинации программных интерфейсов (API), портов и протоколов, применяемых ресурсами узла как для доступа субъектов к аккаунтам, так и для обмена информацией с другими узлами в процессе функционирования.

Доступ субъекта к ресурсам узла через аккаунт должен осуществляться посредством пути доступа. Современные операционные системы обладают свойством предоставлять несколько путей доступа к ресурсам узла. Соответственно, субъект для доступа к ресурсам узла может выбрать один из существующих путей доступа, то есть речь идет об отображении: $AP \to AS$.

Коммутационная диаграмма отображения целей субъекта на объект приведена на рисунке 2.

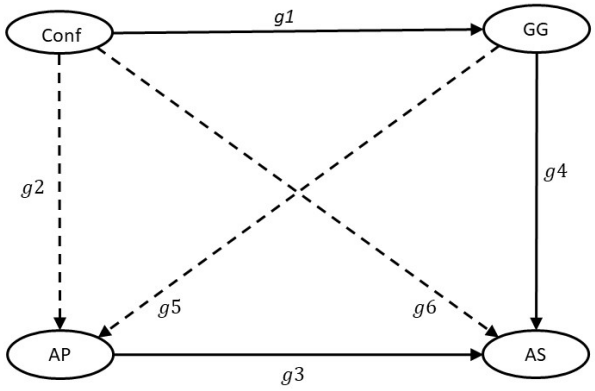

Рис. 2. Диаграмма отображения целей субъекта

На диаграмме выделим морфизм $g6 = g4 \circ g1$, представляющий собой выражение (2).

Полное отображение субъекта на объект. Построение коммутационной диаграммы полного отображения субъекта на объект необходимо и возможно на основании того, что пути доступа и конфигурации должны быть соотнесены с объектом – узлом ИС. На основании анализа деятельности Нарушителя, по разделу тактик, техник и процедур базы знаний организации MITRE (mitre.org) можно говорить, что получение доступа к узлу по любому из путей дает возможность оперировать с множеством ресурсов узла. Таким образом, речь можно вести речь о признании аккаунтов и программ узла областью определения для пути доступа: $AC \to AP \wedge Prog \to AP$.

Архитектура современных операционных систем обеспечивает доступ с одного аккаунта не только к нескольким программам, но и к нескольким конфигурациям. Что дает основания считать конфигурации областью определения аккаунта: $Conf \to AC$.

Полное отображение субъекта на объект в виде коммутационной диаграммы представлено на рисунке 3.

Отношение $Conf \to Prog$ (s1) вытекает из морфизмов $s3 = p4 \circ s1$ и $g2 = s2 \circ s1$.

На диаграмме отметим морфизм $s6 = g3 \circ s4$, подтверждающий выражение (1).

В первой части статьи было показано, что функционирование узла может быть представлено импликацией: $G(V,E) \to Res \to SA \to ME$. В свою очередь, события и сообщения оказывают решающее воздействие на выбор дальнейших действий субъекта: $ME \to SA \to Res$. Диаграммы отображения субъекта на объект (рисунок 1) и полного отображения (рисунок 3) позволяют дать следующее определение

Def. 1. Допустимые действия субъекта SA. Это такие действия субъекта, которые позволяют проводить необходимые ему манипуляции с данными, конфигурациями, событиями и сообщениями в рамках действующего аккаунта. Что можно формализовать в виде:

$$SA = (p9 \vee s6 \circ p4) \to (p1 \vee p6 \vee s1) \quad (3)$$

В результате построения коммутационных диаграмм полное описание узла принимает вид:

$$V[\![\, Prog, Conf, Data, ME, AP, AC, GG, AS, \cup_{i=[1,10]} p_i, \cup_{j=[1,6]} g_j, \cup_{k=[1,6]} s_k]\!] \quad (4).$$

Данное описание позволяет ввести следующие определения, которые будут формализованы как функционалы.

Def. 2. Права доступа. Права доступа субъекта к ресурсам определяются коммутационной диаграммой, приведенной на рисунке 4. Данная диаграмма построена на основании следующих морфизмов, определяющих права доступа:

$p2 = p5 \circ p1$ – соответствующий соотношению между данными, программами и аккаунтом (см. рисунок 1);

$p9 = p7 \circ p5$ – соответствующий использованию субъектом программ (см. рисунок1);

$s3 = p4 \circ s1$ – соответствующий соотношению между конфигурациями, программами и аккаунтом (см. рисунок 3);

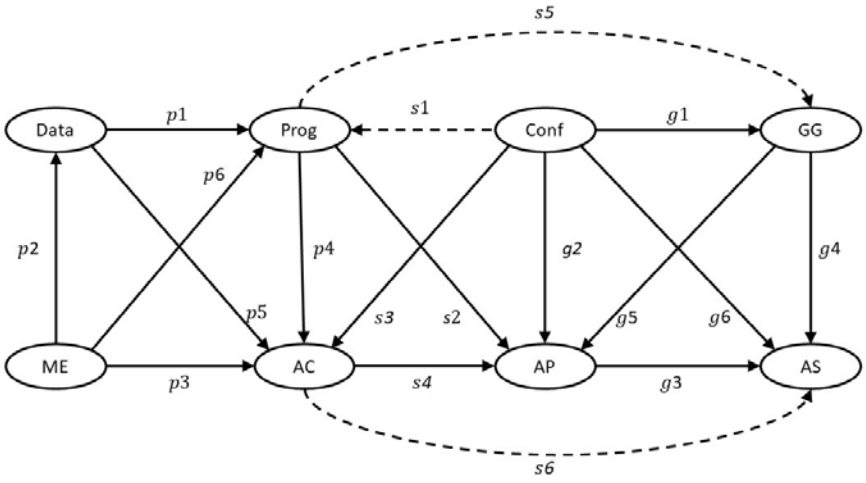

Рис. 3. Диаграмма полного отображения субъекта на объект

$s6 = g3 \circ s4$ – подтверждающий выражение (1), а также необходимость и достаточность представления аккаунта как отображения субъекта на узле (см. рисунок 4).

Указанные морфизмы дают функцию субъекта:

$$AR(AS) = (p1 \circ p9 \vee p2 \circ s6) \wedge \wedge (p9 \circ s1 \vee s6 \circ s3) \quad (5)$$

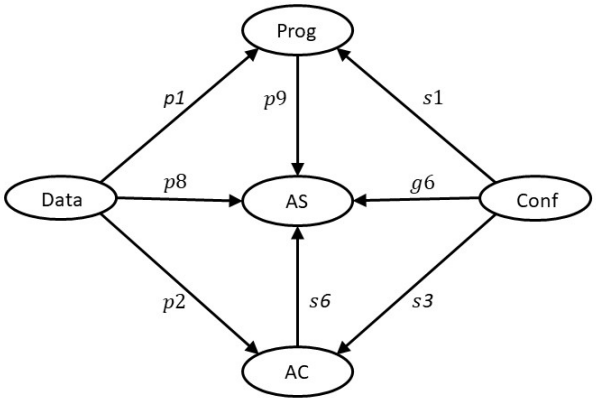

Рис. 4. Диаграмма прав доступа

Необходимо обратить внимание на наличие оператора ИЛИ в каждой из скобок выражения (5). Его наличие может свидетельствовать о *неполноте классических моделей*, рассматривающих проблему доступа «субъект-объект» в терминах «человек» и «данные». Данный вопрос заслуживает отдельного исследования.

Def. 3. Информационный ресурс. Под информационным ресурсом (далее – ИР) будем понимать набор данных и средств их обработки (программы). Диаграмма ИР представлена на рисунке 1 и как левая часть на рисунке 3.

$$IR = \{\bigcup_{i=[1,10]} p_i, AR(AS)\} \quad (6)$$

Def. 4. Информационный поток. Под информационным потоком (далее – ИП) будем понимать набор путей доступа к ресурсам узла со стороны субъекта или другого узла. Диаграмма ИП представлена центральной и правой частями на рисунке 3.

$$IS = \{\bigcup_{k=[1,6]} s_k, \bigcup_{j=[1,3]} g_j, \bigcup_{j=[5,6]} g_j, AR(AS)\} \quad (7)$$

Отметим, что в выражении (7) отсутствует морфизм: $g_4: GG \rightarrow AS$ в виду его имманентности.

Выражения (1 – 7) позволяют установить соотношения между понятиями «субъект», «аккаунт», «информационный ресурс» и «информационный поток» в виде орграфа как показано на рисунке 5.

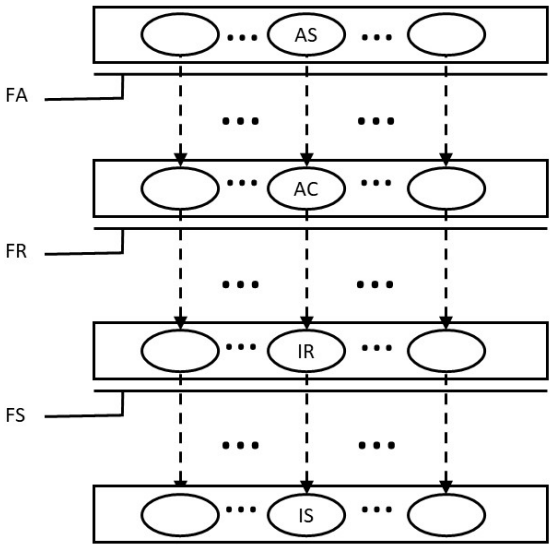

Рис. 5. Соотношения понятий

На рисунке связи орграфа показаны условно для сохранения наглядности и отображают только направление взаимодействия между понятиями. В реальности, например соотношение между аккаунтом и ИР как правило будет «один ко многим». В силу архитектуры ИС и компонент ИС соотношения между понятиями на каждом из трех уровней, обозначенные на рисунке как *FA*, *FR*, *FS* образуют иерархию. Это дает возможность рассматривать *FA*, *FR*, *FS* как функции связей для двудольных графов, образуемых *AS, AC, IR, IS*.

Обозначим:

e_1 – вершина верхнего уровня двудольного графа;

e_2 – вершина нижнего уровня двудольного графа;

$x(e_1, e_2)$ – связь между вершинами двудольного графа.

$$y = f(x) = \begin{cases} 1, \exists x(e_1, e_2) \\ 0, \nexists x(e_1, e_2) \end{cases}$$

Тогда можем записать:

$$FA(AC) = \langle y_1, \ldots, y_n \rangle, e_1 \in AS, e_2 \in AC, n = |AC|$$
$$FR(IR) = \langle y_1, \ldots, y_n \rangle, e_1 \in AC, e_2 \in IR, n = |IR|$$
$$FS(IS) = \langle y_1, \ldots, y_n \rangle, e_1 \in IR, e_2 \in IS, n = |IS|$$

При условии фиксации каждой из вершин верхнего уровня функции будут представлять собой бинарные вектора. В итоге узел ИС для конкретного субъекта может быть представлен как списки ИР доступных данному аккаунту и соответствующих ресурсу ИП:

$$V[\![AR(AS), FA(AC), FR(IR), FS(IS)]\!] \quad (8)$$

Одним из основополагающих принципов построения архитектуры узлов ИС является наличие для всех субъектов нестрогого порядка аккаунтов определяемого правами доступа $AC = \{ac_1, \ldots, ac_n, \leq_{AR}\}$,

$n = |AC|$. С этой точки зрения *FA*, *FR*, *FS* формируют граф подчиненности аккаунтам, который является графом со слабыми связями, то есть такой, в котором нижележащий узел связан с более чем одним узлом верхнего уровня.

Одним из основополагающих требований ИБ является установление категорий данных на основе определения их ценности для субъекта. При этом категории данных:

- представляют собой меру «абсолютной ценности» данных для субъектов (в настоящем исследовании не рассматривается);
- едины для всех субъектов и их аккаунтов в пределах конкретного объекта (например, узла ИС);
- отображаются на аккаунты в отношении «многие к одному», то есть один аккаунт может работать с несколькими категориями данных.

Обозначим множество категорий данных как Ξ и введем отношение строгого упорядочивания прав доступа по категориям данных $AR = \{ar_1, \dots, ar_n, <_\Xi\}$, $n = |\Xi|$.

Поскольку целью деятельности субъекта является обработка данных, а аккаунты, ИР и ИП являются средствами для этого, то можно считать, что эти средства также могут быть упорядочены по категориям данных. Но с учетом нестрогого порядка аккаунтов и подчиненности функций *FA*, *FR*, *FS* аккаунту, множества *AC*, *IR*, *IS* следует считать частично упорядоченными. То есть, каждый из элементов этих множеств может использоваться для обработки различных категорий данных. Тогда имеем:

$$AC = \{ac_1, \dots, ac_n, \leq_\Xi\}, n = |\mathrm{AC}|,$$
$$IR = \{ir_1, \dots, ir_n, \leq_\Xi\}, n = |\mathrm{IR}|,$$
$$IS = \{is_1, \dots, is_n, \leq_\Xi\}, n = |\mathrm{IS}|.$$

Морфизмы *s6* и *s5* позволяют сделать утверждение, что аккаунт, используемый субъектом для обработки данных определенной категории, может быть отождествлен с достижением конкретной цели. Это утверждение может быть представлено как морфизм: $AC \to GG$. Истинность этого морфизма подтверждается следующими выражениями (см. рис. 3) $Prog \to GG = (AC \to GG) \circ (Prog \to AC)$ и $AC \to AP = (AC \to GG) \circ (GG \to AP)$. При этом, каждое из выражений содержит морфизмы из определения ИП и ИР. Таким образом можно говорить об иерархии целей и средств их достижения в рамках аккаунта.

Обозначим:

$AC^\Xi, AC^\Xi \in AC$ – подмножество аккаунтов, используемых при обработке определенной категории данных для достижения заданной цели;

$IR^\Xi, IR^\Xi \in IR$ – подмножество ИР, используемых при обработке определенной категории данных для достижения заданной цели;

$IS^\Xi, IS^\Xi \in IS$ – подмножество ИП, используемых при обработке определенной категории данных для достижения заданной цели.

Тогда выражение (8) можно записать как

$$V = \bigcup_i (AC_i^\Xi, FR(IR_i^\Xi), FS(IS_i^\Xi)), i = [1, |\Xi|] \quad (9)$$

Иерархия целей обработки основана на категории данных и представляет собой решетку. Тогда, следуя [19, см. литературу там же], на основе категории данных можно сформировать подмножества множества путей, которые представляют собой цепочки связей от верхнего уровня иерархии к нижнему в графе подчиненности. То есть, фиксация (обозначаемая как $\hbar$) определенных аккаунта и категории данных формирует подмножество множества путей графа подчиненности.

$$ISA = ac_i^\Xi, FR(IR_i^\Xi), FS(IS_i^\Xi)$$
$$\hbar\, i \in \Xi, \hbar\, ac^\Xi \in AC^\Xi \quad (10)$$

Все современные компоненты ИС реализуют многопользовательский и многозадачный режим работы узлов обеспечивающий:

- взаимодействие между компонентами за счет обмена данными и сигналами, в том числе без участия субъекта;
- целенаправленную обработку данных с возможностью вариации имеющихся алгоритмов для разных типов данных;
- совместное использование своих ресурсов в зависимости от взаимодействия с субъектами и другими компонентами;
- оптимальность использования своих ресурсов зависимости от взаимодействия с субъектами и другими компонентами.

Опираясь на положения [20] совместно с выражениями (6 – 10) дадим следующее определение:

Def. 5. Агентом ИБ (information security agent – *ISA*) называется представление субъекта как аккаунт узла ИС, обеспечивающего ограниченно-рациональное и ограниченно-интеллектуальное использование доступных информационных ресурсов и информационных потоков узла для обработки определенной категории данных в интересах субъекта. Ограниченность

свойств рациональности и интеллектуальности определяется встроенными алгоритмами и конфигурациями узла.

С целью упрощения дальнейшего изложения введем следующие обозначения для ИП, входящих в состав агента – *QR* и для ИР из состава агента – *QS*.

Генерация событий и сообщений

Как было показано на коммутационной диаграмме отношений «субъект-объект» (рисунок 1), ИС является единственным источником доступных субъектам событий и сообщений *ME*. Собственно события и сообщения для субъектов формируются непосредственно узлами ИС, что выражается функцией *Gen(ME)*. Фактически речь идет о функции регистрации параметров работы каждого из агентов ИС. С учетом диаграммы полного отображения субъекта на объект через отношения: $Prog \to ME$, $AC \to ME$ и $Conf \to ME$ и соответствующие морфизмы можно записать:

$$Gen(ME) = p6 \vee p3 \vee \big((s3 \circ p3) \vee (s1 \circ p3)\big) = p6 \vee p3 \vee (s3 \vee s1) \quad (11)$$

Морфизм $AC \to ME$ соответствует регистрации действий субъекта в рамках аккаунта. В процессе реализации отношения формируются сообщения и события о фактах использования субъектом программ, данных и конфигураций, а также об ошибках при доступе и выполнении действий. Можно положить, что все ошибки для данного отношения так или иначе связаны с отказом в доступе конкретному субъекту $as \in AS$ к объектам *Prog*, *Conf*, *Data* через аккаунт $ac \in AC$ данного агента.

Морфизм $Conf \to ME$ соответствует регистрации фактов обращения к конфигурациям. Отметим, что все операции с конфигурациями – чтение, модификация, создание и удаление – совершаются с помощью программ. Кроме того, формирование ошибочных правил в конфигурациях ведут к ошибкам в работе программ.

Морфизм $Prog \to ME$ соответствует регистрации фактов использования программ. К которым относятся не только их запуск и останов, но и регистрация режимов обработки данных (возможно, и типов обрабатываемых данных). Сюда же целесообразно отнести сообщения о текущем статусе программ, которые могут включать описание программы в виде версии, подключенных модулях и т.п.

В общем виде функционирование агента в виде шагов приводящих к формированию событий и сообщений представлено на рисунке 6. Связь между «ME» и «старт обработки» показывает, что формирование сигналов и сообщений не прерывает обработку данных. Прерыванию обработки соответствует «сбой работы».

События и сообщения формируются внутри программ, – на основе их собственных параметров χ_i описывающих как обработку данных, так и состояние аппаратных средств. В общем виде формирование события или сообщения можно рассматривать как результат решения SMT-формулы (*satisfiability modulo theories*, SMT): $f\{\chi_1,..,\chi_i\} = true$. Отличительной особенностью SMT-формулы является наличие в ее составе количественных неравенств, результат вычисления которых дает значения «истина» или «ложь». Задача разрешения SMT-формул относится к NP-полным задачам, что может сказываться на достоверности и полноте формируемых событий и сообщений.

Таким образом, формирование событий и сообщений может быть представлено в общем виде как: $f\{\chi_1,..,\chi_i\} \to me, me \in ME$. Отметим, что решение

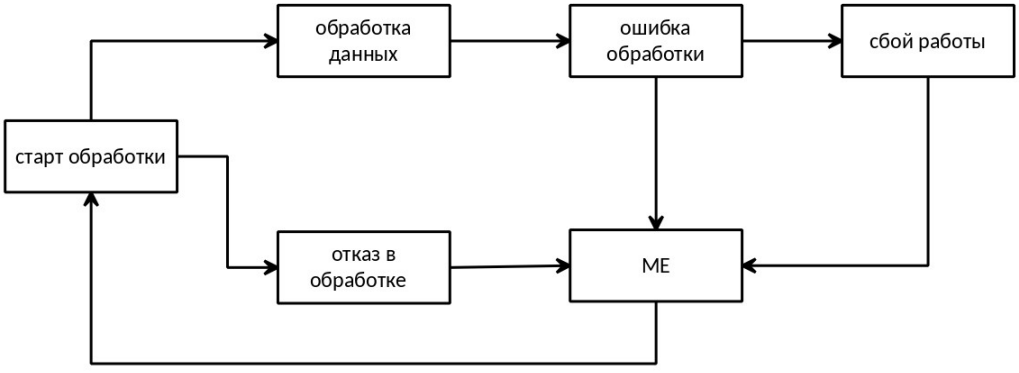

Рис. 6. Формирование ME

SMT-формулы дает только истинностное значение, а собственно событие или сообщение представляют из себя некоторую семантическую конструкцию пригодную для дальнейшей обработки. Это означает, что:

а) перечень событий и сообщений формируется разработчиками ИП и ИР и является конечным предопределенным множеством;

б) для каждого ИП и ИР определено собственное (уникальное) множество событий и сообщений;

в) решение SMT-формулы сопровождается выбором определенного события или сообщения из соответствующего множества $me = s(ME)$;

г) процедура выбора также определена разработчиками ИП и ИР.

Отметим, что механизмы формирования собственных параметров χ_i, наборов событий и сообщений, а также их соотнесение друг с другом для любого ИП и ИР не являются целью настоящего исследования.

Также отметим, в качестве промежуточного вывода, что:

1) функция генерации событий и сообщений может быть представлена в виде $Gen(ME): f\{\chi_{1,\dots,}\chi_i\} \to s(ME)$;

2) функция генерации событий и сообщений уникальна для каждого ИП или ИР $Gen(ME_i) \neq Gen(ME_j), i,j \in (QS \vee QR)$.

Введем функцию генерации событий и сообщений для конкретного ИП или ИР как $FE(x) : f\{\chi_{1,\dots,}\chi_i\} \to s(ME_x), \quad x \in (QS \vee QR)$. Результатом работы функции $FE(x)$ является подмножество выявленных в процессе работы событий ИП или ИР $M = FE(x)$ множества всех предопределенных событий и сообщений этого ИП или ИР ME_x, $M \subset ME$.

События и сообщения являются необходимыми исходными данными для субъекта при проведении им оценки агента, что позволяет определить в соответствии с выражением (10) новое множество, описывающее состояние агента на основании работы ИП и ИР из состава данного агента

$$QM = (\bigcup_{i \in QS} FE(QS_i)) \cup (\bigcup_{j \in QR} FE(QR_j)) \quad (12)$$

Рассуждения и выводы, позволившие сформулировать выражение (12), позволяют представить описание состояния агента посредством фиксированных наборов событий и сообщений – паттернов, которые представляют из себя подмножества множества *QM*, то есть: $QP = \{m_i, \dots, m_n\}, \ n = |QP|, \ m \in QM, \ QP_i \cap QP_j \neq \emptyset, QP_i \cup QP_j = QM$.

Соответственно можно определить функцию, определяющую вхождение события или сообщения в состав паттерна $FP(m): (m \in QM) \to (m \in QP)$. Результатом работы функции будет вектор, каждый элемент которого будет представлять из себя двоичную величину $BP = FP(QM), \ BP = \{b_i, \dots, b_n\}$, $b_i = \{0,1\}$. Значение «1» для элемента будет означать вхождение события или сообщения в паттерн. Отметим, что паттерн и соответствующий ему вектор будут формироваться каждым ИП и ИР из состава агента. Это дает основание говорить, что соответствующие вектора и множества описывают ИП и ИР: $QS[\![QP, BP]\!]$, $QR[\![QP, BP]\!]$. Представляется целесообразным положить, что в общем виде каждому состоянию агента будет соответствовать определенный паттерн *QP* и вектор *BP* для каждого ИП и ИР. То есть для определения состояния агента необходимо вычислить логическую функцию выполнения паттерна: $\psi = QP \wedge BP$, а затем уже определять состояние агента на основании функции

$$Ref: (\bigcup_{i \in QS} \psi_i) \cup (\bigcup_{j \in QR} \psi_j) \quad (13).$$

Таким образом, состояние агента будет считаться определенным, если $Ref = true$.

Следует отметить, что в процессе работы агента результаты вычислений функции ψ для отдельных паттернов того или иного ИП или ИР могут изменяться в силу того, что регистрируемые события и сообщения $m \in QM$ будут нести один из следующих видов информации:

- новое *m* подтверждает изменение параметров работы;
- новое *m* содержит ошибочные параметры;
- новое *m* является повторением имевшего место ранее;
- новое *m* не изменяет результатов.

То есть, следует вести речь о накоплении определенного количества изменений в текущем состоянии для перехода агента в иное состояние. Таким образом, целесообразно рассматривать функции определения состояния агента как функции времени $\psi(t)$ и $Ref(t)$. Детальное исследование механизма оценки состояния агента – функции *Ref* – будет проведено в следующей статье данного цикла статей. Выражения (9 – 10) показывают, что обработка данных в ИС может быть представлена через взаимодействие агентов. Для дальнейшего рассмотрения вопросов взаимодействия агентов необходимо определить типы их состояний

Состояния агента

В самом общем виде состояние агента можно рассматривать с двух точек зрения: что он знает о себе

(внутренняя оценка) и что агент знает о своем окружении (внешняя оценка). Подчеркнем еще раз, что единственным источником, позволяющем оценить состояние агента и его окружения, являются события и сообщения формируемые в результате обработки информации внутри данного агента в процессе его взаимодействия с этим окружением.

Архитектурные решения современных ИС основаны на взаимодействии их компонент в процессе обработки информации. В ИБ принято, как правило, рассматривать взаимодействие с точки зрения нарушения конфиденциальности, целостности и доступности данных. Без потери общности нарушение конфиденциальности, целостности и доступности можно отождествить с нарушениями работоспособности и/или прав доступа. Что может быть вызвано как внешним воздействием (атакой), так и внутренними причинами. На основании выражений (9 – 13) можно говорить о нарушениях конфиденциальности, целостности и доступности для ИП и ИР агентов, образующих ИС. Таким образом, внутренняя и внешние оценки состояний агента следует рассматривать с точки зрения его отношения к своему окружению – взаимодействующим агентам. Причем такие оценки состояния необходимо проводить для каждого из агентов окружения.

Прежде всего выделим состояние, когда внутри агента не фиксируется нарушений при взаимодействии с окружением. Это можно назвать состоянием *Лояльности* (*Lr*) – когда взаимодействующий агент осуществляет корректное, благожелательное сотрудничество, подразумевающее отсутствие намерений по нанесению ущерба оценивающему агенту.

Фиксация агентом нарушений своей работоспособности на основе собственных событий и сообщений позволяет определить состояние *Нелояльности* (*Dr*) – когда действия другого агента влекут за собой или могут расцениваться как подразумевающие нанесение ущерба агенту, производящему такую оценку. Здесь принципиально отметить следующую особенность. Если в процессе деструктивного воздействия агент оказывает противодействие, то это следует расценивать как стремление нанести ущерб атакующему. Следовательно, для атакующего защищающийся агент также будет иметь состояние *Нелояльности*.

Наконец, выделим отказы в отдельное состояние. Представляется очевидным, что конечные паттерны описания отказа в самом агенте не будут зависеть от внешних или внутренних причин. Также можно положить, что паттерны описания отказов среди окружающих агентов будут независимы от внешних или внутренних причин этих отказов. Обозначим это состояние как *Безразличное* (*Ur*) – когда агент по тем или иным причинам не может участвовать во взаимодействии.

Приведенное в предыдущем разделе описание функции *Gen(ME)* позволяет считать, что паттерны описания состояний агента могут так или иначе пересекаться, вплоть до полного совпадения для отдельных паттернов. Это связано как с возможностью ошибок при оценке состояния, так и с тем, что все паттерны агента формируются на основе одного и того же множества событий и сообщений *QM*. Соответственно, необходимо ввести еще одно состояние агента *Индифферентное* (*Ir*). Под этим будем понимать ситуацию, когда агент не может различить состояние окружающих его агентов и с целью обеспечения функционирования ИС готов принимать любые действия окружающих агентов, даже связанные с возможным нанесением ему ущерба.

Таким образом, каждый агент на основании доступного ему множества событий и сообщений *QM* принимает определенное состояние для взаимодействия с каждым из окружающих агентов из множества возможных состояний: $R = \{Lr, Dr, Ir, Ur\}$. Учтем, что оценка состояний агентом имеет определенную долю уверенности *P* в силу ошибок первого и второго рода и тогда формальное описание состояния данного агента для отношений с любым из агентов его окружения будет иметь вид

$$r = P(Ref(\bigcup_{i \in QS} \psi_i) \cup (\bigcup_{j \in QR} \psi_j)), r \in R \quad (14)$$

Приведенные рассуждения позволяют положить, что обобщенное состояние агента представляет собой вектор его состояний, определяемых для каждого из агентов окружения. Обозначим множество агентов как *AG*, а подмножество агентов, взаимодействующих с данным – как *AV*. Соответственно, общее состояние агента *Q* с учетом (14) описывается как:

$$Q_a = [r_1, \dots, r_n], n = |AV|, AV \subset AG, \quad a \in AG \wedge a \notin AV \quad (15)$$

Отношения агентов

Выражения (12) и (13) показывают, что состояние агента и его оценки окружения определяются отношениями между ИП и ИР из его состава. В ИБ принято оценивать такие отношения с точки зрения соблюдения конфиденциальности, целостности и доступности. Для агента конфиденциальность полностью определяется аккаунтом, в рамках которого функционируют ИП

и ИР агента. Из требования обеспечения целостности и доступности вытекает необходимость рассматривать отношения между ИП и ИР агента только как Лояльные (Lr), а также симметричные и рефлексивные. Это позволяет сформулировать аксиому лояльности (здесь и далее нумерация аксиом является продолжением нумерации первой части статьи).

Аксиома 7. Лояльность ИП и ИР внутри агента

$$x[Lr]x =: x \in (QR \vee QS),$$
$$x[Lr]y = y[Lr]x : x \in QR, y \in QS,$$
$$a[Lr]b : a, b \in QR, \quad (16)$$
$$a[Lr]b : a, b \in QS,$$

где *QR* – ИП и *QS* – ИР из состава агента.

Полагаем, что отношение Безразличия (Ur) между ИП и ИР соответствует неработоспособности агента. Характеристика Нелояльности или Индифферентности для отношений ИП и ИР агента невозможно в силу того, что в структуре агента отсутствует возможность разделять деятельность субъектов в пределах аккаунта. То есть, для субъекта (Нарушителя или Защитника) целеполагание может быть сведено к получению и использованию прав доступа к ИП и ИР агента. Эти же рассуждения позволяют положить наличие транзитивности в отношениях ИП и ИР агента. Таким образом, при применении логико-вероятностного метода в ИБ можно исключить из рассмотрения отношения ИП и ИР внутри агента в силу их эквивалентности, независимо от их состава (мощности множеств *QR* и *QS*), определяемого аккаунтом агента.

На диаграмме отображения целей субъекта на объект (рисунок 2) отметим следующие морфизмы, сочетание которых показывает, что для достижения своих целей субъект должен манипулировать с конфигурациями узла:

$g6 = g4 \circ g1$ – представляющий выражение (2);

$g4 = g3 \circ g5$ – показывающий, что для достижения целей субъекта необходимо наличие пути доступа к объекту;

$g6 = g3 \circ g2$ – показывающий, что для манипулирования правилами субъекту необходимо наличие пути доступа к объекту.

Таким образом, можем сформулировать следующую аксиому

Аксиома 8. Любой субъект – пользователь ИС представлен в ИС как агент

$$\forall s \in AS \; \exists a \in AG \quad (17)$$

где *AS* – множество субъектов – пользователей и *AG* – множество агентов ИС.

На диаграмме, приведенной на рисунке 3, отметим следующие морфизмы:

$s6 = g3 \circ s4$ – подтверждающий необходимость и достаточность представления аккаунта как отображения субъекта на узле;

$g1 = s5 \circ s1$ – показывающий, что достижение целей субъекта полностью определяются множествами конфигураций и программ узла;

$s3 = p4 \circ s1$ – соответствующий соотношению между конфигурациями, программами и аккаунтом, то есть определяющий права доступа.

Эти морфизмы позволяют рассматривать выражение (10) как описание одного из возможных аккаунтов узла ИС. Но если посмотреть на выражение (10) с точки зрения системного администрирования, то оно будет соответствовать аккаунту системного администратора узла (root, system, superuser и т.п.), которому доступны для выполнения действий все ИП и ИР узла. Эту аналогию можно распространить и на другие уровни современных ИС – виртуализация ИР и ИП позволяет рассматривать их как часть ресурсов соответствующей платформы виртуализации [12]. С учетом рисунка 5, рассуждений при выводе выражений (8) и (9), можно предположить наличие аккаунтов для различных подсистем и типов данных в ИС. То есть, все уровни ИС с точки зрения ИБ могут быть описаны единым образом – с помощью выражения (10) или с помощью понятия агента ИБ (*ISA*). Универсальность понятия агента базируется на едином подходе к описанию различных уровней – от отдельного сервиса (один сервис-один аккаунт-один поток) до ИС целиком – за счет определения подмножеств ИП и ИР включенных в отдельный аккаунт. То есть можем сформулировать аксиому эквивалентности агентов.

Аксиома 9. Для иерархии аккаунтов ИС существует эквивалентная иерархия агентов данной ИС.

$$\forall a \in AG \; \exists c \in AC \mid AG\{a_1, \ldots, a_i, \leq\}, \quad (18)$$
$$\leftrightarrow AC\{c_1, \ldots, c_j, \leq\}$$

где *AG* – множество агентов и *AC* – множество аккаунтов ИС.

На диаграмме отображения субъекта на объект (рисунок 1) отметим следующие морфизмы:

$p9 = p7 \circ p5$ и $p8 = p9 \circ p1$ – соответствующие использованию субъектом программ и данных, то есть определяющие его возможные действия *SA*;

$p2 = p5 \circ p1$ – соответствующий соотношению между данными, программами и аккаунтом, то есть определяющий права доступа;

$p10 = p7 \circ p3 = p9 \circ p4$ – определяющий возможности субъекта по оценке состояния узла в процессе его функционирования.

Эти морфизмы, а также рисунок 6 и выражения (3), (11), (14) показывают, что функционирование агента, выражающее целеполагание субъекта, может быть представлено как: $Gen(ME) \to Ref(ME) \to Sel(SA)$, что соответствует положениям и выводам первой части статьи. Обозначим отношение между агентами как *RA*. Предложенные ранее описание функций *Gen, Ref* как программно реализуемых автоматов позволяют считать, что каждое отношение каждого агента представляет собой выражение: $RA = \{Ref(ME) \circ Gen(ME) \circ Sel(SA)\}$. То есть, с учетом предыдущих аксиом, определим аксиому эквивалентности отношений

Аксиома 10. Отношения «субъект-субъект»(*RS*) и «субъект-объект» (*RO*) в ИС эквивалентны отношениям агентов из состава ИС.

$$(s_i RS s_j \vee sROv) \mid \forall s \in AS\ v \in V \leftrightarrow \exists (a_k RA a_l) \mid \forall a \in AG, \quad (19)$$

где *AS* – множество субъектов – пользователей, *V* – множество узлов (компонент) и *AG* – множество агентов ИС.

На основании выражения (10) агент описывается ИП, ИР и аккаунтом $a[\![QR, QS, ac]\!]$, $a \in AG$, $ac \in AC$. Соответственно, можем определить отношение агентов следующим образом: $xRAy \mid x[\![QR_x, QS_x. ac_x]\!], y[\![QR_y, QS_y, ac_y]\!]$. Прежде всего отметим, что взаимодействующие агенты в большинстве случаев будут иметь разные аккаунты $ac_x \neq ac_v$. Аналогично можно сказать и относительно ИР $QR_x \neq QR_y$. Выражение (11) показывает, что каждый агент, участвующий во взаимодействии, определяет состояние другого участника отношения только на основании событий и сообщений собственных ИР. Тогда по аналогии с первой частью отношения между двумя агентами $\forall (x, y) \in AG$:

$$xRAy = \{Ref^x(ME^x) \circ Gen^x(ME^x) \circ Sel^x(SA^x)\}$$
$$yRAx = \{Ref^y(ME^y) \circ Gen^y(ME^y) \circ Sel^y(SA^y)\}$$

В силу независимости генерации событий и сообщений каждым из агентов – участников отношения действуют аксиомы 1 и 7, что позволяет положить асимметричность отношений агентов $xRAy \neq yRAx$ или $x[R] \neq y[R]$. Для формирования отношения, то есть определения агентом состояния респондента отношения, этот респондент должен воздействовать на агента так, чтобы у агента сформировались события и сообщения. То есть, взаимодействие агентов определяется связями графа ИС $G(V, E)$, что позволяет отождествить ИП агентов со связями графа $E\left[\!\left[x \overset{R}{\to} y \neq y \overset{R}{\to} x\right]\!\right]$. Или иначе – рассматривать ИП как носители отношений (что соответствует положениям теории системного анализа) $xRy \mid (x[\![QR_x, QS_x]\!], y[\![QR_y, QS_y]\!]) \Rightarrow R[\![QS_x, QS_y]\!]$.

Это позволяет ввести аксиому двойственности. Следуя соглашениям, принятым в первой части статьи, формальная запись будет иметь вид:

Аксиома 11. Любая связь в многоагентной ИС или отношение агентов этой ИС описывается (должна быть помечена) двумя значениями состояний отношения

$$G(V, E): E[\![x[R] \wedge y[R]]\!], \quad \forall (x, y) \in AG, x \neq y \quad (20)$$

Поскольку для обеспечения взаимодействия ИП должны быть согласованы, то набор возможных состояний агента в процессе его функционирования и взаимодействия с окружением могут быть представлены единым множеством $R = \{Lr, Dr, Ir, Ur\}$, независимо от типа аккаунта и числа ИП и ИР в составе агента.

В качестве итога на основании (15) определим состояние агента в виде вектора

$$Q_a = [x[R]_1, \dots, x[R]_n], n = |AV|, \quad AV \subset AG, x \in AV, a \in AG, a \neq x \quad (21)$$

Приведенные аксиомы позволяют указать на следующие свойства подобия, позволяющие описывать различные уровни ИС в терминах агентов.

Prop 1, На любом уровне ИС (представляющая собой множество ИП, ИР и аккаунтов) может быть представлена как подмножество ИП и ИР включенных в отдельный аккаунт, то есть как агент ИБ.

Prop 2. Отношения между агентами формируются единым способом – на основе имеющихся в его распоряжении события и сообщений.

Prop 3. Отношения агентов описываются единообразным набором состояний не зависимо от того, какой уровень описывает агент ИБ.

Prop 4. Состояние агента определяется вектором состояний его окружения не зависимо от того, какой уровень описывает агент ИБ.

Таким образом, агент ИБ обладает свойством самоподобия с точки зрения ИБ и любая ИС может быть представлена в виде вложенных структур, состоящих из ИР, ИП и аккаунтов, то есть агент является универсальной структурой.

Заключение

Для достижения общей цели исследования (адаптации логико-вероятностного метода для решения задач ИБ) в статье разработаны формально-логические основы для определения и последующего выделения фрактальных структур, присущих ИС как сложной системе. Предложено выделить в ИС макроуровень в виде агентов информационной безопасности, состоящих из информационных ресурсов, информационных потоков и прав доступа аккаунта субъекта. Проведен анализ отношений между агентами с использованием аппарата математических и логических отношений. В качестве основных результатов настоящего исследования отметим следующее.

1. Целесообразно в дальнейшем рассматривать с точки зрения ИБ любую ИС, обычно представляемую в виде графа *G(V,E),* как многоагентную систему.

2. Определение агента как обладающего ограниченными свойствами рациональности и интеллектуальности, а также свойства подобия позволяют перенести рассмотрение отношений «субъект-субъект» и «субъект-объект» в ИБ на уровень отношений между агентами.

3. Агент может определять действия субъектов (других агентов) только за счет ИП со своим окружением и/или внешней информации из других источников.

4. Отношения агентов является местом формирования и разрешения конфликта, вызванного отношениями субъектов.

5. Рассмотрение отношений ИП и ИР внутри агента можно исключить из дальнейшего анализа.

6. Объективно в структуре агента отсутствует возможность разделять деятельность субъектов в пределах одного аккаунта, то есть каждый агент и ИС в целом индифферентны к целеполаганию субъектов.

7. С точки зрения ИБ отношение агентов может быть представлено как попарное взаимодействие агентов, определяемое соответствующими состояниями этих агентов.

Литература

1. Рябинин И. А. Решение одной задачи оценки надежности структурно-сложной системы разными логико-вероятностными методами / И.А. Рябинин, А.В. Струков // Моделирование и анализ безопасности и риска в сложных системах, Санкт-Петербург, 19–21 июня 2019 года. – Санкт-Петербург: Санкт-Петербургский государственный университет аэрокосмического приборостроения, 2019. – С. 159-172.
2. Демин А. В. Глубокое обучение адаптивных систем управления на основе логико-вероятностного подхода / А.В. Демин // Известия Иркутского государственного университета. Серия: Математика. – 2021. – Т. 38. – С. 65-83. – DOI 10.26516/1997-7670.2021.38.65
3. Викторова В.С. Вычисление показателей надежности в немонотонных логико-вероятностных моделях многоуровневых систем / В.С. Викторова, А.С. Степанянц // Автоматика и телемеханика. – 2021. – № 5. – С. 106-123. – DOI 10.31857/S000523102105007X.
4. Леонтьев А.С. Математические модели оценки показателей надежности для исследования вероятностно-временных характеристик многомашинных комплексов с учетом отказов / А.С. Леонтьев, М.С. Тимошкин // Международный научно-исследовательский журнал. – 2023. – № 1(127). С. 1 – 13. – DOI 10.23670/IRJ.2023.127.27.
5. Пучкова Ф.Ю. Логико-вероятностный метод и его практическое использование / Ф.Ю. Пучкова // Информационные технологии в процессе подготовки современного специалиста: Межвузовский сборник научных трудов / Министерство просвещения Российской Федерации; Федеральное государственное бюджетное образовательное учреждение высшего образования «Липецкий государственный педагогический университет имени П.П. СЕМЕНОВА-ТЯН-ШАНСКОГО». Том Выпуск 25. – Липецк: Липецкий государственный педагогический университет имени П.П. Семенова-Тян-Шанского, 2021. – С. 187-193.
6. Россихина Л.В. О применении логико-вероятностного метода И.А. Рябинина для анализа рисков информационной безопасности / Л.В. Россихина, О.О. Губенко, М.А. Черноситова // Актуальные проблемы деятельности подразделений УИС: Сборник материалов Всероссийской научно-практической конференции, Воронеж, 20 октября 2022 года. – Воронеж: Издательско-полиграфический центр «Научная книга», 2022. – С. 108-109.
7. Карпов А.В. Модель канала утечки информации на объекте информатизации / А.В. Карпов // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2018): VII Международная научно-техническая и научно-методическая конференция. Сборник научных статей. В 4-х томах, Санкт-Петербург, 28 февраля – 01 марта 2018 года / Под редакцией С.В. Бачевского. Том 2. – Санкт-Петербург: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2018. – С. 378-382.
8. Методика кибернетической устойчивости в условиях воздействия таргетированных кибернетических атак / Д.А. Иванов, М.А. Коцыняк, О.С. Лаута, И.Р. Муртазин // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2018): VII Международная научно-техническая и научно-методическая конференция. Сборник научных статей. В 4-х томах, Санкт-Петербург, 28 февраля – 01 марта 2018 года / Под редакцией С.В. Бачевского. Том 2. – Санкт-Петербург: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2018. – С. 343-346.

9. Елисеев Н. И. Оценка уровня защищенности автоматизированных информационных систем юридически значимого электронного документооборота на основе логико-вероятностного метода / Н.И. Елисеев, Д.И. Тали, А.А. Обланенко // Вопросы кибербезопасности. – 2019. – № 6(34). – С. 7-16. – DOI 10.21681/2311-3456-2019-6-07-16.
10. Коцыняк М.А. Математическая модель таргетированной компьютерной атаки / М.А. Коцыняк, О.С. Лаута, Д.А. Иванов // Наукоемкие технологии в космических исследованиях Земли. – 2019. – Т. 11, № 2. – С. 73-81. – DOI 10.24411/2409-5419-2018-10261.
11. Белякова Т.В. Функциональная модель процесса воздействия целевой компьютерной атаки / Т.В. Белякова, Н.В. Сидоров, М.А. Гудков // Радиолокация, навигация, связь: Сборник трудов XXV Международной научно-технической конференции, посвященной 160-летию со дня рождения А.С. Попова. В 6-ти томах, Воронеж, 16–18 апреля 2019 года. Том 2. – Воронеж: Воронежский государственный университет, 2019. – С. 108-111.
12. Калашников А. О. Инфраструктура как код: формируется новая реальность информационной безопасности / А.О. Калашников, К.А. Бугайский // Информация и безопасность. – 2019. – Т. 22, № 4. – С. 495-506.
13. Бугайский К.А. Расширенная модель открытых систем (Часть 1) / К. А. Бугайский, Д. С. Бирин, Б. О. Дерябин, С. О. Цепенда // Информация и безопасность. – 2022. – Т. 25, № 2. – С. 169-178. – DOI 10.36622/VSTU.2022.25.2.001.
14. Нестеров А. Ю. Проблема субъекта в искусственной природе / А. Ю. Нестеров // Гуманитарный вектор. – 2021. – Т. 16, № 2. – С. 22-28. – DOI 10.21209/1996-7853-2021-16-2-22-28.
15. Дыдров А. А. Построение дискурса о цифровом как феномене информационной современности / А. А. Дыдров, Р. В. Пеннер // Социум и власть. – 2022. – № 3(93). – С. 114-126. – DOI 10.22394/1996-0522-2022-3-114-126. .
16. Бугайский К. А. Расширенная модель открытых систем (Часть 2) / К.А. Бугайский, И.С. Перескоков, А.О. Петров, А.О. Петров // Информация и безопасность. – 2022. – Т. 25, № 3. – С. 321-330. – DOI 10.36622/VSTU.2022.25.3.001.
17. Бугайский К. А. Расширенная модель открытых систем (Часть 3) / К.А. Бугайский, Б.О. Дерябин, К.В. Табаков, Е.С. Храмченкова, С.О. Цепенда // Информация и безопасность. – 2022. – Т. 25, № 4. – С. 501-512.
18. Калашников А. О. Модель количественного оценивания агента сложной сети в условиях неполной информированности / А. О. Калашников, К. А. Бугайский // Вопросы кибербезопасности. – 2021. – № 6(46). – С. 26-35. – DOI 10.21681/2311-3456-2021-6-26-35.
19. Максимов Д. Ю. Формирование оптимального маршрута в конфигурационном пространстве больших групп интеллектуальных агентов с помощью линейной логики / Д. Ю. Максимов // Управление развитием крупномасштабных систем (MLSD'2018) : Материалы одиннадцатой международной конференции. В 2-х томах, Москва, 01–03 октября 2018 года / Под общей редакцией С.Н. Васильева, А.Д. Цвиркуна. Том I. – Москва: Институт проблем управления им. В.А. Трапезникова РАН, 2018. – С. 309-311.
20. Левкина И. Н. Общая структура многоагентной системы поддержки принятия решений shape * MERGEFORMAT / И. Н. Левкина, Т. М. Леденева // Евразийский союз ученых. – 2020. – № 5-5(74). – С. 43–46.
21. Применение логико-вероятностного метода в информационной безопасности (Часть 2) / Калашников А.О., Бугайский К.А., Бирин Д.С., Дерябин Б.О., Цепенда С.О., Табаков К.В. // Вопросы кибербезопасности. – 2023. – № 4(56). – С. 23–32. – DOI 10.21681/2311-3456-2023-4-23-32.

APPLICATION OF THE LOGICAL-PROBABILISTIC METHOD IN INFORMATION SECURITY (PART 1)

Kalashnikov A.O.[9], Bugajskij K.A.[10], Anikina E.V.[11], Pereskokov I.S.[12], Petrov Andrej O.[13], Petrov AleksandrO.[14], Hramchenkova E.S.[15], Molotov A.A.[16]

The purpose of the article: *adaptation of the logical-probabilistic method of evaluating complex systems to the tasks of building information security systems in a multi-agent system.*

Research method: *during the research, the main provisions of the methodology of structural analysis, system analysis, decision theory, methods of evaluating events under the condition of incomplete information were used.*

The result: *this article continues the consideration of information security issues based on the analysis of the relationship between the subjects and the object of protection. The presentation of the subject and object*

[9] Andrey Kalashnikov, Dr.Sc., Chief Scientist of the Laboratory «Complex networks» Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia. E-mail: aokalash@ipu.ru

[10] Konstantin Bugajskij, Junior Researcher of the Laboratory «Complex networks» Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia. E mail: kabuga@ipu.ru

[11] Eugenia Anikina – research fellow, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, e-mail: ajanet@ipu.ru

[12] Ilia Pereskokov – junior researcher, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, e-mail: pereskokov @phystech.edu

[13] Andrei Petrov – junior researcher, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, e-mail: petrovaojob@gmail.com

[14] Aleksandr Petrov – junior researcher, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, e-mail: petrovalexandr@ipu.ru

[15] Ekaterina Hramchenkova– junior researcher, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, e-mail: hramchenkovaes@yandex.ru

[16] Aleksandr Molotov, software engineer Institute of Control Sciences of Russian Academy of Sciences. E mail: alpha.sphere@ya.ru

of protection in the form of an intelligent agent is justified, taking into account the requirements for information protection. Formal definitions of the information security agent and its main characteristics are given: information resource, information flow and access rights of the subject. It is shown that the concept of an information security agent is the basis for identifying structures in an information system. The axiomatics of the relations of the subject and the object as agents of information security, as well as the relations between information resources and information flows within the agent, has been developed. The possibility of determining the state of an agent based on events and messages generated during its operation is shown.

Scientific novelty: *consideration of information security issues using the apparatus of mathematical and logical relations. Development of formal definitions of the information security agent and its constituent information resources and information flows, which are the basic universal components of the description of structures in the information system. Definition of the concept of an information security agent by considering the mapping of the subject and its goal-setting on the object..*

Keywords: *information security model, assessment of complex systems, logical-probabilistic method, theory of relations, system analysis, multi-agent system.*

References

1. Rjabinin I. A. Reshenie odnoj zadachi ocenki nadezhnosti strukturno-slozhnoj sistemy raznymi logiko-verojatnostnymi metodami / I.A. Rjabinin, A.V. Strukov // Modelirovanie i analiz bezopasnosti i riska v slozhnyh sistemah, Sankt-Peterburg, 19–21 ijunja 2019 goda. – Sankt-Peterburg: Sankt-Peterburgskij gosudarstvennyj universitet ajerokosmicheskogo priborostroenija, 2019. – S. 159-172.
2. Demin A. V. Glubokoe obuchenie adaptivnyh sistem upravlenija na osnove logiko-verojatnostnogo podhoda / A.V. Demin // Izvestija Irkutskogo gosudarstvennogo universiteta. Serija: Matematika. – 2021. – T. 38. – S. 65-83. – DOI 10.26516/1997-7670.2021.38.65
3. Viktorova V.S. Vychislenie pokazatelej nadezhnosti v nemonotonnyh logiko-verojatnostnyh modeljah mnogourovnevyh sistem / V.S. Viktorova, A.S. Stepanjanc // Avtomatika i telemehanika. – 2021. – № 5. – S. 106-123. – DOI 10.31857/S000523102105007X.
4. Leont'ev A.S. Matematicheskie modeli ocenki pokazatelej nadezhnosti dlja issledovanija verojatnostno-vremennyh harakteristik mnogomashinnyh kompleksov s uchetom otkazov / A.S. Leont'ev, M.S. Timoshkin // Mezhdunarodnyj nauchno-issledovatel'skij zhurnal. – 2023. – № 1(127). S. 1 – 13. – DOI 10.23670/IRJ.2023.127.27.
5. Puchkova F.Ju. Logiko-verojatnostnyj metod i ego prakticheskoe ispol'zovanie / F.Ju. Puchkova // Informacionnye tehnologii v processe podgotovki sovremennogo specialista: Mezhvuzovskij sbornik nauchnyh trudov / Ministerstvo prosveshhenija Rossijskoj Federacii; Federal'noe gosudarstvennoe bjudzhetnoe obrazovatel'noe uchrezhdenie vysshego obrazovanija «Lipeckij gosudarstvennyj pedagogicheskij universitet imeni P.P. SEMENOVA-TJaN-ShANSKOGO». Tom Vypusk 25. – Lipeck: Lipeckij gosudarstvennyj pedagogicheskij universitet imeni P.P. Semenova-Tjan-Shanskogo, 2021. – S. 187-193.
6. Rossihina L.V. O primenenii logiko-verojatnostnogo metoda I.A. Rjabinina dlja analiza riskov informacionnoj bezopasnosti / L.V. Rossihina, O.O. Gubenko, M.A. Chernositova // Aktual'nye problemy dejatel'nosti podrazdelenij UIS: Sbornik materialov Vserossijskoj nauchno-prakticheskoj konferencii, Voronezh, 20 oktjabrja 2022 goda. – Voronezh: Izdatel'sko-poligraficheskij centr "Nauchnaja kniga", 2022. – S. 108-109.
7. Karpov A.V. Model' kanala utechki informacii na ob#ekte informatizacii / A.V. Karpov // Aktual'nye problemy infotelekommunikacij v nauke i obrazovanii (APINO 2018): VII Mezhdunarodnaja nauchno-tehnicheskaja i nauchno-metodicheskaja konferencija. Sbornik nauchnyh statej. V 4-h tomah, Sankt-Peterburg, 28 fevralja – 01 marta 2018 goda / Pod redakciej S.V. Bachevskogo. Tom 2. – Sankt-Peterburg: Sankt-Peterburgskij gosudarstvennyj universitet telekommunikacij im. prof. M.A. Bonch-Bruevicha, 2018. – S. 378-382.
8. Metodika kiberneticheskoj ustojchivosti v uslovijah vozdejstvija targetirovannyh kiberneticheskih atak / D.A. Ivanov, M.A. Kocynjak, O.S. Lauta, I.R. Murtazin // Aktual'nye problemy infotelekommunikacij v nauke i obrazovanii (APINO 2018): VII Mezhdunarodnaja nauchno-tehnicheskaja i nauchno-metodicheskaja konferencija. Sbornik nauchnyh statej. V 4-h tomah, Sankt-Peterburg, 28 fevralja – 01 marta 2018 goda / Pod redakciej S.V. Bachevskogo. Tom 2. – Sankt-Peterburg: Sankt-Peterburgskij gosudarstvennyj universitet telekommunikacij im. prof. M.A. Bonch-Bruevicha, 2018. – S. 343-346.
9. Eliseev N. I. Ocenka urovnja zashhishhennosti avtomatizirovannyh informacionnyh sistem juridicheski znachimogo jelektronnogo dokumentooborota na osnove logiko-verojatnostnogo metoda / N.I. Eliseev, D.I. Tali, A.A. Oblanenko // Voprosy kiberbezopasnosti. – 2019. – № 6(34). – S. 7-16. – DOI: 10.21681/2311-3456-2019-6-07-16.
10. Kocynjak M.A. Matematicheskaja model' targetirovannoj komp'juternoj ataki / M.A. Kocynjak, O.S. Lauta, D.A. Ivanov // Naukoemkie tehnologii v kosmicheskih issledovanijah Zemli. – 2019. – T. 11, № 2. – S. 73-81. – DOI 10.24411/2409-5419-2018-10261.
11. Beljakova T.V. Funkcional'naja model' processa vozdejstvija celevoj komp'juternoj ataki / T.V. Beljakova, N.V. Sidorov, M.A. Gudkov // Radiolokacija, navigacija, svjaz': Sbornik trudov XXV Mezhdunarodnoj nauchno-tehnicheskoj konferencii, posvjashhennoj 160-letiju so dnja rozhdenija A.S. Popova. V 6-ti tomah, Voronezh, 16–18 aprelja 2019 goda. Tom 2. – Voronezh: Voronezhskij gosudarstvennyj universitet, 2019. – S. 108-111.
12. Kalashnikov A. O. Infrastruktura kak kod: formiruetsja novaja real'nost' informacionnoj bezopasnosti / A.O. Kalashnikov, K.A. Bugajskij // Informacija i bezopasnost'. – 2019. – T. 22, № 4. – S. 495-506.
13. Bugajskij K.A. Rasshirennaja model' otkrytyh sistem (Chast' 1) / K. A. Bugajskij, D. S. Birin, B. O. Derjabin, S. O. Cependa // Informacija i bezopasnost'. – 2022. – T. 25, № 2. – S. 169-178. – DOI 10.36622/VSTU.2022.25.2.001.
14. Nesterov A. Ju. Problema sub#ekta v iskusstvennoj prirode / A. Ju. Nesterov // Gumanitarnyj vektor. – 2021. – T. 16, № 2. – S. 22-28. – DOI 10.21209/1996-7853-2021-16-2-22-28.

15. Dydrov A. A. Postroenie diskursa o cifrovom kak fenomene informacionnoj sovremennosti / A. A. Dydrov, R. V. Penner // Socium i vlast'. – 2022. – № 3(93). – S. 114-126. – DOI 10.22394/1996-0522-2022-3-114-126. .
16. Bugajskij K. A. Rasshirennaja model' otkrytyh sistem (Chast' 2) / K.A. Bugajskij, I.S. Pereskokov, A.O. Petrov, A.O. Petrov // Informacija i bezopasnost'. – 2022. – T. 25, № 3. – S. 321-330. – DOI 10.36622/VSTU.2022.25.3.001.
17. Bugajskij K. A. Rasshirennaja model' otkrytyh sistem (Chast' 3) / K.A. Bugajskij, B.O. Derjabin, K.V. Tabakov, E.S. Hramchenkova, S.O. Cependa // Informacija i bezopasnost'. – 2022. – T. 25, № 4. – S. 501-512.
18. Kalashnikov A. O. Model' kolichestvennogo ocenivanija agenta slozhnoj seti v uslovijah nepolnoj informirovannosti / A. O. Kalashnikov, K. A. Bugajskij // Voprosy kiberbezopasnosti. – 2021. – № 6(46). – S. 26-35. – DOI 10.21681/2311-3456-2021-6-26-35.
19. Maksimov D. Ju. Formirovanie optimal'nogo marshruta v konfiguracionnom prostranstve bol'shih grupp intellektual'nyh agentov s pomoshh'ju linejnoj logiki / D. Ju. Maksimov // Upravlenie razvitiem krupnomasshtabnyh sistem (MLSD'2018) : Materialy odinnadcatoj mezhdunarodnoj konferencii. V 2-h tomah, Moskva, 01–03 oktjabrja 2018 goda / Pod obshhej redakciej S.N. Vasil'eva, A.D. Cvirkuna. Tom I. – Moskva: Institut problem upravlenija im. V.A. Trapeznikova RAN, 2018. – S. 309-311.
20. Levkina I. N. Obshhaja struktura mnogoagentnoj sistemy podderzhki prinjatija reshenij shape * MERGEFORMAT / I. N. Levkina, T. M. Ledeneva // Evrazijskij sojuz uchenyh. – 2020. – № 5-5(74). – S. 43–46.
21. Primenenie logiko-verojatnostnogo metoda v informacionnoj bezopasnosti (Chast' 2) / Kalashnikov A.O., Bugajskij K.A., Birin D.S., Derjabin B.O., Cependa S.O., Tabakov K.V. // Voprosy kiberbezopasnosti. – 2023. – № 4(56). – S. 23–32. – DOI 10.21681/2311-3456-2023-4-23-32.

CYBERSECURITY ISSUES

SCIENTIFIC PEER-REVIEWED JOURNAL

2023, № 5(57)

Cybersecurity Issues is a research periodical scientific and practical publication specializing in information security. Published six times a year

WWW.CYBERRUS.COM

The journal is being published from 2013 (Registration Certificate PI No. FS 77-75239). CrossRef number (DOI): 10.21681/2311-3456

The journal is included in the Russian list of peer-reviewed academic publications of the Higher Attestation Commission (VAK), it is registered in the Russian Science Citation Index (RSCI/RINTs) on the Web of Science (WoS) platform and holds the 1st place in its cyber security rating. The journal's articles are available in full text

Founder and publisher
JSC "NPO "Echelon"

Postal address: Elektrozavodskaya str., 24, bld. 1, 107023, Moscow, Russia

E-mail: editor@cyberrus.info

CONTENTS

УДК: 004.056

Международная безопасность в среде информационно-коммуникационных технологий. Коллективная монография по проблеме применения норм ответственного поведения государств в ИКТ-среде

Стрельцов А.А. и др.; Предисловие В.П.Шерстюка; под ред. А.А.Стрельцова, А.Я.Капустина, Т.А.Поляковой, А.С.Маркова, Б.Н.Мирошникова. - М.: НАМИБ, 2023, 132 с.

В монографии представлены результаты исследования проблемы применения норм ответственного поведения государств в среде информационно-коммуникационных технологий (ИКТ) и даны научно обоснованные рекомендации по вопросам международного сотрудничества в области безопасного использования информационно-коммуникационных технологий в формирующемся глобальном информационном обществе. Задача изысканий состояла в содействии выработке новых принципов и норм международного права, регулирующих деятельность государств в глобальном информационном пространстве, в целях установления международно-правового режима обеспечения безопасности в сфере использования информационно-коммуникационных технологий.

Фактическую основу исследования составили «Нормы, правила и принципы ответственного поведения государств в ИКТ-среде», применение которых может «снизить риск нарушения международного мира, безопасности и стабильности», сформулированные с учетом предложений многих стран. Методологическую основу работы составили теоретические положения, обоснованные в результате выполнения исследовательского проекта «Методологические вопросы применения норм, правил и принципов ответственного поведения государств, призванных способствовать обеспечению открытой, безопасной, стабильной, доступной и мирной ИКТ-среды» (2018–2020 г.г.). Материалы монографии могут быть использованы представителями Национальной Ассоциации международной информационной безопасности и заинтересованных федеральных органов государственной власти в процессе участия в деятельности Рабочей группы открытого состава ООН (РГОС) и на других международных площадках по проблематике международной информационной безопасности.

CYBERSECURITY ISSUES
VOPROSY KIBERBEZOPASNOSTI

№5 2023

DOI: 10.21681/2311-3456

- Secure Artificial Intelligence
- Security of Cyber-Physical Systems
- Application Security

www.cyberrus.com
editor@cyberrus.com