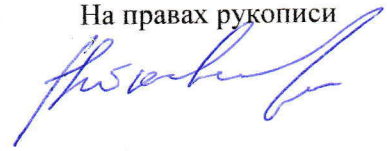
Федеральное государственное бюджетное учреждение науки
«Санкт-Петербургский Федеральный исследовательский центр
Российской академии наук»
(СПб ФИЦ РАН)

На правах рукописи



Виткова Лидия Андреевна

МОДЕЛИ, АЛГОРИТМЫ И МЕТОДИКА ПРОТИВОДЕЙСТВИЯ ВРЕДОНОСНОЙ ИНФОРМАЦИИ В СОЦИАЛЬНЫХ СЕТЯХ

Специальность 05.13.19 – «Методы и системы защиты информации, информационная безопасность»

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель:
кандидат технических наук,
Сахаров Дмитрий Владимирович

Санкт-Петербург – 2021

Оглавление

ВВЕДЕНИЕ

Актуальность темы диссертации. Глубина проникновения социальных сетей в повседневную жизнь значительна, и их преимуществом является возможность участников коммуникации оперативно высказывать свое мнение большой группе людей, публиковать медиа файлы. Сегодня социальные сети (СС) являются не только средством общения, но и инструментом распространения информации. Очевидной проблемой информационной безопасности современного общества стала вредоносная информация. Стоит отметить, что террористические и преступные группировки все чаще берут на вооружение средства информационного воздействия, пишут стратегии, направленные на расширение сферы влияния и вовлечение новых адептов через СС. Именно поэтому одной из составляющих обеспечения информационной безопасности государства представляется мониторинг, анализ и активное противодействие вредоносной информации в СС.

Само понятие «вредоносная информация» рассматривается экспертами разных наук, но консенсус пока не достигнут. Изучением вопросов противодействия распространению вредоносной информации стали заниматься еще в 1990г. Так, например, В.Н. Лопатин входил в состав парламентской комиссии Верховного Совета СССР и отвечал за вопросы информационной безопасности. В своих работах он определял «вредоносную информацию» как угрозу информационной безопасности. К этому он относил: распространение порнографии; клевету; недостоверную информацию, скрытую рекламу. В 21 веке к вредоносной информации все же чаще относят «фейковые новости». Особенно остро необходимость противодействия распространению таким новостям в СС, порождающим волны паники, возникла во время пандемии.

Однако, в настоящее время проблема противодействия имеет крайне малое количество научно-технических решений. Известные средства

обнаружения и противодействия вредоносной информации в СС не отвечают требованиям к скорости, полноте, точности и адекватности принимаемых решений. Это обусловлено несколькими причинами. Во-первых, системы разделены два не связанных модуля: (1) мониторинг; (2) противодействие. Посередине между ними находится оператор. Во-вторых, СС имеют сложную структуру и состоят из множества разнородных сообщений, что недостаточно учитывается при выборе цели противодействия, например тип сообщения, источник и другие характеристики. В-третьих, в реальном масштабе времени необходимо обрабатывать сверхбольшие объемы сообщений и в сжатые сроки выбирать цель для контрмеры, в ручном режиме оператор системы противодействия не в состоянии остановить распространение вредоносной информации.

Таким образом, основная сложность противодействия вредоносной информации в СС напрямую следует из современных тенденций развития информационной сферы, а именно, увеличения: (1) объема сообщений, содержащих вредоносную информацию; (2) скорости распространения вредоносной информации; (3) скорости тиражирования сообщений; (4) скорости появления новых источников распространения информации в СС; (5) количества способов привлечения внимания аудитории; (6) уровня гетерогенности данных. Это обуславливает необходимость повышения эффективности противодействия вредоносной информации в социальных сетях, в том числе за счет оперативности и обоснованности.

Степень разработанности темы. Большое внимание вопросам противодействия вредоносной информации, анализу и оценке источников вредоносной информации в СС уделяется такими исследователями как Д.А. Губанов, И.В. Котенко, М.В. Литвиненко, Д.А. Новиков, И.Б. Саенко, А.Л. Тулупьев, Д.Ю. Турдаков, А.А. Чечулин, А.Г. Чхартишвили, A.L. Barabasi, X. Zheng и др. Множество работ посвящено информационному конфликту и противоборству. К этой группе можно отнести труды С.А. Будникова, Ю.Л. Козирацкого, В.А. Липатникова, С.И. Макаренко, С.П. Расторгуева, Д.В.

Сахарова. Вопросы информатизации процессов и оценивания эффективности информационных систем раскрываются в работах М.В. Буйневича, В.П. Заболотского, А.А. Мусаева, Р.М. Юсупова. Однако, несмотря на сделанный учеными существенный задел, проблема противодействия вредоносной информации в СС не может считаться разрешенной и требует проведения новых исследований

Цели и задачи. Основной целью диссертационной работы является повышение эффективности противодействия вредоносной информации в СС за счет анализа источников вредоносной информации и автоматизации выбора контрмер. Для достижения данной цели в диссертационной работе поставлены и решены следующие задачи:

1) анализ существующих моделей вредоносной информации и информационного обмена;

2) анализ существующих алгоритмов оценки источников в СС, существующих систем мониторинга и методик противодействия вредоносной информации в СС;

3) разработка комплекса моделей социальной сети, источника и вредоносной информации;

4) разработка комплекса алгоритмов анализа источников вредоносной информации и ранжирования контрмер;

5) разработка методики противодействия вредоносной информации в социальных сетях;

6) разработка архитектуры и программных прототипов компонентов системы противодействия (СПД) вредоносной информации, экспериментальная и теоретическая оценка эффективности предложенных моделей, алгоритмов, методики и архитектуры.

Научная задача. Научная задача заключается в разработке моделей, алгоритмов и методики противодействия вредоносной информации в социальной сети.

Объектом исследования являются СС, в которых возможно наличие сообщений с вредоносной информацией и их источников.

Предметом исследования являются модели, методики и алгоритмы противодействия вредоносной информации в СС.

Научная новизна результатов диссертационной работы состоит в следующем:

1. Комплекс моделей социальной сети, источника и вредоносной информации отличается от аналогов учетом структуры информационного обмена в СС, информационных объектов и вредоносной информации с использованием предложенной классификации объектов социальной сети. Разработанные модели социальной сети и источника содержат новые классы, атрибуты объектов и связи, а модель вредоносной информации, в отличие от аналогов, основана на теории множеств и состоит из взаимосвязанных объектов и признаков вредоносной информации, вместе формирующих вредоносно-информационные объекты. Также в комплекс входит авторская информационно-признаковая модель вредоносной информации.

2. Комплекс алгоритмов анализа источников вредоносной информации и ранжирования контрмер, отличается от аналогов учетом связей и зависимых атрибутов объектов в социальной сети, а также учетом таких атрибутов как потенциал источника, активность пользователей на странице источника, количество просмотров сообщения с вредоносной информацией, количество друзей и подписчиков источника. В качестве результата работы алгоритмы анализа источников формируют ранжированный список объектов воздействия. Алгоритм ранжирования контрмер отличается от аналогов учетом авторских коэффициентов и уровней сложности для каждой меры противодействия в системе и в качестве результата работы формирует ранжированный список контрмер.

3. Методика противодействия вредоносной информации в СС отличается от известных тем, что она ориентирована на автоматический и автоматизированный выбор объектов воздействия и мер противодействия

вредоносной информации из списка ранжированных контрмер. Кроме того, методика отличается от аналогов использованием предложенных моделей представления социальной сети, источника, вредоносной информации, а также предложенных алгоритмов анализа источников и ранжирования контрмер.

4. Архитектура и программные прототипы компонентов СПД вредоносной информации отличаются от известных тем, что ориентированы на ранжирование и выбор доступных контрмер в системе для заданных типов вредоносной информации. Архитектура содержит оригинальные компоненты анализа и оценки источника вредоносной информации, базу данных с информацией о мерах противодействия вредоносной информации в СС, информацию об агентах реализации, через которые контрмеры будут реализованы. В силу этого архитектура позволяет формировать наборы исходных данных для исследований и разработок в области противодействия вредоносной информации, а также для исследований и разработок решений для систем поддержки принятия решения.

Теоретическая и практическая значимость работы. Теоретическая значимость диссертационной работы определяется ее вкладом в развитие теории и методов информационной безопасности, что проявляется в следующих аспектах: введены новые классы, объекты, их атрибуты и связи для анализа и оценки информационных объектов и информационного обмена в СС, расширен класс алгоритмов сортировки и ранжирования для анализа источников в СС, расширен набор критериев для мер противодействия и выделены новые функциональные связи между компонентами архитектуры СПД. Предложенный подход позволяет формулировать научно-обоснованные требования к решению задач, связанных с анализом источника вредоносной информации в СС и с противодействием сообщению или его источнику. Кроме того, предложенные комплексы моделей и алгоритмов, методика и архитектура могут быть использованы как часть системы поддержки принятия решений оператором в интересах противодействия вредоносной информации.

Методология и методы исследования. Для решения поставленных в диссертации задач применялись как классические, так и относительно молодые методы исследования: 1) системный и сравнительный анализ применен в равной степени для получения практически всех основных научных результатов; 2) сбор, систематизация и анализ научной и технической информации о предметной области позволили создать комплекс моделей; 3) объектно-ориентированный подход и структурный синтез использовался для создания алгоритмов анализа и оценки источников; 4) с помощью методов ранжирования, экспертной оценки и методов проектирования архитектур и программных систем были созданы методика противодействия и архитектура СПД.

Положения, выносимые на защиту. Основными положениями, выносимыми на защиту, являются:

1) комплекс моделей социальной сети, источника и вредоносной информации;

2) комплекс алгоритмов анализа источников вредоносной информации и ранжирования контрмер;

3) методика противодействия вредоносной информации в социальной сети;

4) архитектура и программные компоненты системы противодействия вредоносной информации.

Обоснованность и достоверность представленных в диссертационной работе научных положений обеспечивается за счет тщательного анализа состояния исследований в данной области, подтверждается согласованностью результатов, полученных при экспериментальных исследованиях, успешной апробацией основных теоретических положений диссертации на ряде научных конференций всероссийского и международного уровня, а также публикацией основных положений, раскрывающих данные результаты, в ведущих рецензируемых научных изданиях.

Реализация результатов работы. Отраженные в диссертационной работе исследования проведены в рамках следующих научно-исследовательских работ: грант российского научного фонда (РНФ) № 18-71-10094 «Мониторинг и противодействие вредоносному влиянию в информационном пространстве социальных сетей»; грант РНФ № 18-11-00302 «Интеллектуальная обработка цифрового сетевого контента для эффективного обнаружения и противодействия нежелательной, сомнительной и вредоносной информации»; проект «Проблема управления информационно-психологическими аспектами безопасности государства» конкурс инноваций и инновационных проектов в номинации «А» «Конкурс концептуальных идей, методик, рекомендаций» от Международной академии связи и др.

Апробация результатов работы. Основные положения и результаты диссертационной работы были представлены на следующих научных конференциях: МНТ НТК «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (АПИНО 2018, 2019, 2020) (Санкт-Петербург, Россия); 10-я Конференция по социальной информатике (SocInfo 2018) (Санкт-Петербург, Россия); 3 МНК «Интеллектуальные информационные технологии для промышленности» (IITI 2018) (Сочи, Россия); 4 МНК «Интеллектуальные информационные технологии для промышленности» (IITI 19) (Отава, Чехия); 28-Я НТК «Методы и технические средства обеспечения безопасности информации» (МиТСОБИ 2019) (Санкт-Петербург, Россия); МНТК «Автоматизация» (RusAutoCon 2019) (Сочи, Россия); 13я МНК «Intelligent Distributed Computing (IDC 2019) (Санкт-Петербург, Россия); XI Санкт-Петербургская межрегиональная конференция «Информационная безопасность регионов России» (ИБРР 2019) (Санкт-Петербург, Россия) и др.

Публикации. По материалам диссертационного исследования было опубликовано 20 статей, в том числе 6 в рецензируемых изданиях из перечня ВАК: «Защита информации. Инсайд» [1], «Вестник Санкт-Петербургского

государственного университета технологии и дизайна. Серия 1: Естественные и технические науки» [2, 4], «Вестник Воронежского института ФСИН России» [3], «Известия высших учебных заведений. Технология легкой промышленности» [5], «Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления» [6]. Одна статья опубликована в рецензируемом международном журнале [7] и 7 статей – в сборниках трудов международных конференций, индексируемых в базах WoS и/или SCOPUS [8-14]. Получено 3 свидетельства о государственной регистрации программ для ЭВМ и базы данных [15, 16, 17].

Личный вклад. Все результаты, представленные в диссертационной работе, получены лично автором в процессе выполнения научно-исследовательской деятельности.

Структура и объем диссертационной работы. Диссертационная работа включает введение, три главы, заключение, список использованных источников (162 наименования) и 6 приложений. Объем работы – 173 страницы машинописного текста; включая 35 рисунков и 15 таблиц.

Краткое содержание работы. Первая глава диссертации посвящена исследованию задачи противодействия вредоносной информации с учетом требования к обоснованности выбора объекта воздействия и контрмеры. Определены место и роль противодействия вредоносной информации в социальных сетях в информационной безопасности государства, общества и личности. Приведены основные определения и обзор релевантных моделей, алгоритмов и методик, описаны существующие системы противодействия вредоносной информации. Определены достоинства существующих подходов и решений, выделены их основные недостатки, затрудняющие противодействие вредоносной информации. Обоснована актуальность цели исследования. Предложено использование методики, основанной на анализе источников, сортировке объектов воздействия по приоритету, ранжировании контрмер для решения поставленной в исследовании цели. Итогом

проведенного анализа является формальная постановка задачи и определение критериев для оценки эффективности.

Во второй главе диссертации представлены разработанный комплекс моделей, состоящий из модели социальной сети, модели источника, теоретико-множественной модели вредоносной информации. В комплекс также входит информационно-признаковая модель. Модель социальной сети состоит из таких связанных между собой элементов, как сообщение и источник. Модель источника содержит атрибуты активности источника в социальной сети такие как: количество отметок «мне нравится», количество комментариев, количество репостов, количество просмотров сообщений, тип сообщения, количество сообщений на странице источника, индекс активности, индекс просматриваемости, индекс влиятельности и потенциал источника. Теоретико-множественная модель, состоит из взаимосвязанных объектов и признаков вредоносной информации, вместе формирующих вредоносно-информационные объекты. Также в данной главе представлен комплекс алгоритмов анализа источников и ранжирования контрмер. В комплекс входят следующие алгоритмы: 1) алгоритм ранжирования источников по потенциалу; 2) алгоритм оценки источников; 3) алгоритм сортировки объектов воздействия по приоритету и 4) алгоритм ранжирования контрмер с учетом сложности реализации. Особенностью разработанного комплекса алгоритмов является то, что он формирует расширенный набор параметров, учитываемых при выборе объекта воздействия и контрмеры в процессе противодействия вредоносной информации.

Третья глава диссертации содержит описание методики противодействия вредоносной информации в социальных сетях с учетом требований к обоснованности и определяет основные стадии использования разработанных моделей и алгоритмов. В методике выделяются две стадии выполнения: 1) стадия настройки, в которую входит формирование и сохранение исходных данных, таких как: информационные угрозы, агенты реализации, контрмеры и 2) стадия эксплуатации, включающая: запрос

и получение набора данных, анализ источников, сортировка объектов воздействия, формирование списков пар цель-контрмера и противодействие. Также в данной главе описана архитектура и программные прототипы компонентов системы противодействия. Архитектура состоит из компонентов, разделенных на уровни: 1) компонент менеджмента, 2) компонент визуализации; 3) компонент анализа и оценки источников, 4) сервер управления базами данных, 5) базу данных, 6) компонент выбора контрмер, 7) компонент реализации контрмер. В главе описаны следующие программные прототипы компонентов: 1) программный прототип компонента анализа и оценки источников в СС; 2) программный прототип компонента выбора контрмер; 3) программный прототип базы данных информационных угроз и контрмер. Проведена экспериментальная оценка предложенных алгоритмов и программных прототипов. Продемонстрированы результаты экспериментальной и теоретической оценки методики, по основным показателям эффективности (оперативность, обоснованной и ресурсопотребление). Результаты проведенных теоретических и экспериментальных исследований показали, что разработанная методика удовлетворяет предъявляемым требованиям. Также в третьей главе сформулированы предложения по использованию методики.

Выражаю глубокую и искреннюю благодарность научному руководителю и Чечулину А.А. за поддержку, внимание и интерес к работе. Отдельно выражаю благодарность моим родным, близким и друзьям за веру в меня и одобрение моего пути.

ГЛАВА 1. Современное состояние проблемы противодействия вредоносной информация в социальных сетях

Процессы и конфликты в информационном поле государства являются отражением активности различных субъектов деятельности, будь то индивидуальные, институциональные или групповые акторы. При этом мы наблюдаем обратную тенденцию, когда конфликты и процессы в информационном поле могут порождать события и конфликты, меняющие общество в целом, а также оказывать непосредственное влияние на социальную активность людей, их увлечения и жизненный путь. Важно то, что процессы, порождающие изменения состояния безопасности государства и общества протекают, как правило, в скрытой (латентной) форме и мы обнаруживаем результат воздействия на информационное поле государства, общества или личности, только в момент его кульминации, когда процесс или конфликт отражается в политическом дискурсе, экономике государства, жизни и здоровье общества или личности.

Отметим, что еще в 80 х годах двадцатого века известный американский учёный в областях исследования операций и теории систем Рассел Линкольн Акофф писал: «Само изменение постоянно меняется» [18]. Идеи Акоффа нашли свое отражение в экономике и в менеджменте, однако на методологии управления информационной безопасностью его достижения практически не отразились. Стоит отметить, что еще совсем недавно технические и социальные изменения в области информационной безопасности были достаточно медленными и государство могло к ним приспосабливаться путем обновлений требований регуляторов к игрокам рынка услуг телекоммуникаций и контента. Но сегодня скорость изменений в информационном поле государства настолько велика, что мы можем провести аналогию с «морским штормом», а неверная и замедленная реакция со стороны органов власти и безопасности государства может привести к катастрофе.

Адаптация к происходящим изменениям требует быстрых и значительных корректировок в области защиты информационного поля государства. Как отмечается в национальном стандарте ГОСТ Р 53647.9-2013 [19] необходимо быть: более осведомленными о потенциале и характере кризисных ситуаций; способным лучше противодействовать и способным лучше восстанавливаться после кризиса. Специалисты же по безопасности все еще ищут стабильности, их целью, можно сказать, является «гомеостаз». Однако информационное поле, в котором они добиваются этой цели, все более динамично и нестабильно, изменения в нем, сродни динамике событий в период кризисных ситуаций. Благодаря изменению среды коммуникации общества возрастает взаимосвязанность и взаимозависимость информационных систем физических лиц, социальных групп, организаций, институтов и государств. Наше окружение становится более широким, сложным и менее подконтрольным.

Сегодня остается открытым вопрос о том, каким образом распознать вредоносную информацию в информационном поле, каким образом государству противодействовать возможным цветным революциям, отдельным социальным вызовам (например, о детском и подростковом суициде), как защищать общество от панических настроений в период глобальных катастроф и изменений, или каким образом родителям защищать детей от рекламы наркотиков?

Только путь совершенствования систем противодействия распространению вредоносной информации со стороны государства, общества, организаций, семьи, личности может дать действенные результаты. В этом контексте тема диссертации представляется актуальной.

1.1 Место и роль проблемы противодействия вредоносной информации в социальных сетях

1.1.1 Сущность вредоносной информации в системе противодействия

Анализ состояния исследований показал, что известные работы в основном направлены на обеспечение нормативно-правовых или информационно-технических аспектов информационной безопасности в информационном пространстве, а также на мониторинг инцидентов информационной безопасности, на анализ качественных или количественных характеристик связей узлов в социальных сетях, кластеризацию полученных данных, систематизацию, хранение и пр.

Вредоносная информация, нежелательная информация, информационное воздействие, противоправная информация все эти понятия рассматриваются экспертами зачастую смешанно или как синонимы. С позиции специалистов военных наук – одно понимание, с точки зрения правоведов, политологов, психологов – другое, специалисты по анализу данных не имеют на этот счет ярко выраженной позиции. При этом ключевыми понятиями являются «информация» и «вредоносность».

Рассмотрим понятие «информация» (*I* – information). В общем представлении – это сведения об объектах и явлениях, их параметрах, свойствах и состоянии, которые подлежат сбору, накоплению, хранению, предобработке, обработке, преобразованию, непосредственному использованию и передаче [20].

Информация входит в понятия информационная сфера, информационное пространства.

Согласно Доктрине РФ, «информационная сфера» (*IR – information realm*) – это совокупность информации, объектов информатизации, информационных систем, сайтов в информационно-телекоммуникационной сети «Интернет» (далее – сеть «Интернет»), сетей связи, информационных

технологий, субъектов, деятельность которых связана с формированием и обработкой информации, развитием и использованием названных технологий, обеспечением информационной безопасности, а также совокупность механизмов регулирования соответствующих общественных отношений [21].

Понятие «информационного пространства» (*IA – information area*) трактуется как сфера человеческой деятельности, связанная с созданием, преобразованием и потреблением информации, включающая в себя всю совокупность информационных ресурсов данного общества [22, 23].

Между этими двумя понятиями справедливо соотношение $IR \subseteq IA$, т.е. *IR* является подмножеством множества *IA*.

Зачастую вредоносная информация воспринимается в современном научном сообществе как элемент информационной атаки (воздействия). Понятие «информационного воздействия» (*IE – information effect*) трактуется как основной поражающий фактор информационной войны, представляющий собой воздействие информационным потоком на объект атаки – информационную систему или ее компонент, с целью вызвать в нем в результате приема и обработки данного потока заданные структурные и/или функциональные изменения [24].

Формально это понятие можно определить следующим образом:

$$R = IE(IO),$$

где *IE* – некоторое информационное воздействие, *IO* – информационный объект, *R* – результат.

Информационный объект (*IO – information object*) – это логически цельный блок информации, представленный в определенной фиксированной форме, который создан и используется в ходе информационной составляющей деятельности человека [25].

Формально связь этого понятия с другими понятиями представляется следующим образом: $IO \in I$, т.е. информационный объект является элементом множества всей анализируемой информации *I*.

В связи с необходимостью классификации вредоносной информации, сформулируем определение, опираясь на классификацию типов информации в сети Интернет (*Int*) в общем.

Положим, что множество *Int* содержит «опасную» информацию *RI* (от английского risk – рискованный, опасный, рисковый, авантюрный) и «безопасную» *SI* (от английского safe – безопасный, надежный, безвредный, неопасный), т.е. $Int = RI + SI$.

Множество опасной информации содержит вредоносную (от английского harmful – вредный, пагубный, опасный, губительный), нежелательную (от английского unwanted – нежелательный, ненужный, неугодный), сомнительную (от английского doubtful– сомнительный, спорный) информацию. На рисунке 1.1 представлена модель опасной информации.

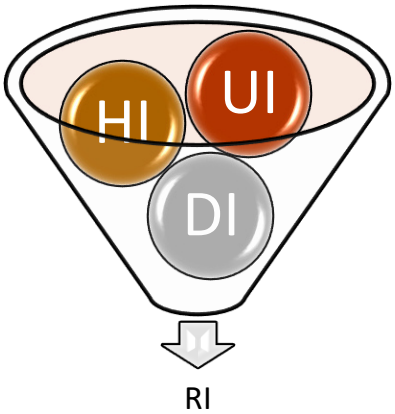


Рисунок 1.1 – Модель опасной информации в сети Интернет

Классификация типов информации в сети Интернет основана на понятии информационного объекта, который является логически цельным блоком информации, представленным в определенной фиксированной форме,

созданным и используемым в ходе информационной составляющей деятельности человека.

Для диссертации введем понятие **вредоносной информации** – это отдельный информационный объект и/или совокупность объектов в сети Интернет, содержащий запрещенную или ограниченную к распространению информацию. В дальнейшем предполагается, что понятие вредоносная информация включает в себя все множество опасной информации.

С точки зрения обеспечения государственной безопасности под категорию вредоносной информации попадают следующие виды информации:

1) информация, попадающая под критерии оценки материалов и (или) информации, позволяющие идентифицировать ее, как запрещенную к распространению в Российской Федерации [26];

2) информация, включенная в федеральный список экстремистских материалов [27];

3) информационный объект, включенный в реестр блокировок.

На примере организации (предприятия):

1) конфиденциальная информация;

2) персональные данные;

3) информация для служебного пользования.

На примере системы родительского контроля:

1) информация, имеющая возрастные ограничения;
2) информация, ресурсы, доступ к которым ограничен со стороны родителя.

В настоящей диссертации использованы следующие определения.

«Информация» – сведения (сообщения, данные) независимо от формы их представления [20]. Понятие «материалы», которое часто встречается в современном законодательстве, приравнивается по смыслу к понятию «информация».

В п. 6 ст. 10 ФЗ от 27.07.2006 N 149-ФЗ (ред. от 19.07.2018) «Об информации, информационных технологиях и о защите информации» говорится о том, что запрещается распространение информации, которая направлена на пропаганду войны, разжигание национальной, расовой или религиозной ненависти и вражды, а также иной информации, за распространение которой предусмотрена уголовная или административная ответственность [28].

«Распространение информации» – это все действия, направленные на получение информации неопределенным кругом лиц или передачу информации неопределенному кругу лиц.

«Источник» (Source) – это страница в социальной сети, на котором опубликована информация доступная неопределённому кругу лиц.

«Сообщение» – это информационный объект, содержащий текст, созданный и опубликованный в процессе информационного обмена в социальной сети.

1.1.2 Виды вредоносной информации в социальных сетях

Из множества классификаций видов вредоносной информации в социальных сетях можно выделить 2 основные концепции:

Классификация информационных объектов по содержанию [28-32]:

1) пропаганда либо оправдание войны;
2) пропаганда или оправдание терроризма и экстремизма;
3) пропаганда или оправдание правонарушений;
4) пропаганда расизма;
5) пропаганда национальной ненависти;
6) пропаганда религиозной ненависти;
7) осквернение, оскорбление исторической памяти, символов воинской славы
8) осквернение, оскорбление государственных символов;
9) оскорбление религиозных чувств верующих;

10) пропаганда деструктивных, нетрадиционных ценностей, установок;
11) оправдание насилия, жестокости;
12) оправдание девиантного поведения;
13) пропаганда, оправдание действий опасных для жизни человека;
14) заведомо ложная информация;
15) клевета;
16) информация, содержащая сведения о способах изготовления чего-то запрещенного;
17) рекламные объявления о покупке, продаже запрещенных товаров
18) сексуально откровенный контент;
19) сексуально откровенный контент с участием несовершеннолетних.

Классификация информационных объектов по дискретным признакам [33-36]:

1) дата регистрации в социальной сети;
2) время сообщений;
3) частота сообщений;
4) длина сообщений;
5) частота действий;
6) уникальность контента на странице профиля в социальной сети;
7) связь с другими участниками в социальной сети;
8) связь с сообществами;
9) география профиля;
10) география сообществ, с которыми связан профиль;
11) степень влияния;
12) количество просмотров;
13) интересы профиля;
14) история содержания профиля.

В зависимости от цели противодействия меняются или разрабатываются модели, алгоритмы, методики, используются разные архитектуры. В текущем исследовании автор преследует цель повысить эффективность противодействия вредоносной информации в социальных сетях.

1.2 Обзор релевантных моделей, алгоритмов, методик и архитектур систем противодействия вредоносной информации в социальных сетях

Платформы социальных сетей развивались постепенно, прообразы современных решений появились еще в XX веке. В 1995 году появился веб-сайт Classmates.com [37]. По идее создателя пользователь может найти своих одноклассников в базе данных учебных заведений, ресурс поддерживал добавление фотографий и отправку текстовых сообщений. В 1997 году начал функционировать веб-сайт SixDegrees.com. Проект закрылся в 2001 году. Сеть поддерживала возможность оформления своего профиля, поиска друзей по интересам, обмена сообщениями. В 1999 году появился Livejournal.com [38] в 2003 «MySpace» [39], в 2004 году была открыта платформа социальной сети Facebook [40]. В 2006 году Джек Дорси запустил новый проект «Twitter» [41].

Первые исследования проводятся учеными вслед за открытием новых платформ, с 1995-2000гг. в «Google Академии» [42] опубликовано 15 работб ссылающихся на ресурс Classmates.com, 28 – SixDegrees.com. С появлением новых сетей количество исследований в области анализа социальных сетей растет в геометрической прогрессии. Еще в 1990 «Social network analysis» был прерогативой таких наук как социология, политология. Напримерб в сборнике [43] собраны статьи, посвященные анализу поведения человека в социуме. В [44] обсуждается взаимопроникновение теории обмена и науки «анализ социальных сетей».

Уже через 15 лет, к 2005 ситуация начала кардинально меняться и к 2021 году «Social network analysis» (SNA) – это процесс исследования различных социальных структур с помощью сетей и теории графов [45]. Объектами исследований являются сетевые структуры с точки зрения узлов (отдельных

акторов, людей или вещей в сети) и связей, ребер или связей (отношений или взаимодействий). Множество работ посвящено анализу распространения мемов [46], информационному обмену [47], сетям связи между друзьями, коллегами, клиентами [48] и многому другому. То есть, современный раздел SNA содержит массивную теоретическую и практическую базу исследований релевантных теме диссертации. Однако, автор не ограничивается только SNA.

Обзор релевантных работ структурирован таким образом, что в начале рассматриваются модели, в том числе предложенные учеными-исследователями в рамках других теоретических школ, таких как теория коммуникации, социология. Затем в работе представлены обзоры алгоритмов анализа источника и выбора контрмер, методик и архитектур.

1.2.1 Обзор релевантных моделей противодействия вредоносной информации в социальных сетях

Так исторически сложилось, что в России теория коммуникаций рассматривается чаще в гуманитарных исследованиях, чем в информационной безопасности. Однако в Оксфордском словаре теория коммуникации трактуется следующим образом: «Это изучение и изложение принципов и методов, с помощью которых передается информация» [49]. Коммуникация в трансмиссионной/кибернетической традиции рассматривается как процесс обработки информации [50]. Такой подход позволяет анализировать процессы прохождения информации в сложно организованных социальных системах, например в социальных сетях, различать источники и получателей, выявлять потери информации и минимизировать их.

Коммуникация в теории трактуется как намеренное действие источника, выполняемое с целью достижения определенного результата. Для нахождения места и роли системы противодействия вредоносной информации в социальных сетях необходимо выяснить что собой представляет источник в контексте информационного обмена в социальных сетях, как происходит информационный обмен.

К релевантным моделям можно отнести общие модели теории коммуникации [50], такие как Модель SMCRE Гарольда Дуайта Лассуэлл [50], модель коммуникации Шеннона и Уивера [51] или интегральная (обобщенная) модель Б. Вестли и М. Маклина [50, 52]. Также к релевантным моделям относятся модели распространения информации в социальных сетях [53] (эпидемические модели: SI [54], SIS [54, 55], SIR [56, 57] SIRS [58], модели независимых каскадов [58-62]), предложенные исследователями в рамках SNA. Рассмотрим модели подробнее.

Линейная модель коммуникации SMCRE

В учебном пособии по теории коммуникации Д.П. Гавра [50] говорится, что в 1948 году американский исследователь пропаганды Харальд Лассуэлл предложил модель коммуникации, при построении которой он опирался на 5 вопросов: (1) Кто сообщает? (2) Что? (3) По какому каналу? (4) Кому? (5) С каким эффектом? Модель Лассвелла в сокращенном виде имеет общее название SMCRE и включает в себя следующие компоненты: (1) Source (источник) – отвечает на вопрос: «Кто сообщает?»; (2) Message (сообщение) – отвечает на вопрос: «Что сообщает?» (3) Channel (канал) – отвечает на вопрос: «По какому каналу?»; (4) Recipient (получатель) – отвечает на вопрос: «Кому?»; (5) Effect (эффект) – отвечает на вопрос: «С каким эффектом?»

Данная модель является линейной и может быть представлена графически следующим образом (рисунок 1.2):

Рисунок 1.2 – Модель коммуникации SMCRE

Каждый элемент модели SMSRE представляет собой сложную конструкцию и может быть представлен одноименным компонентом системы противодействия вредоносной информации в социальных сетях: (1)

компонент противодействия источникам; (2) компонент противодействия сообщениям; (3) компонент противодействия каналам распространения (связанным ресурсам); (4) компонент противодействия вредоносной информации на стороне пользователя (получателя); (5) компонент оценки эффективности системы противодействия.

В статье [63] авторы демонстрируют результаты анализа коммуникационной политики и сообщений террористов на их медиа ресурсах в социальных сетях. В [64] предложен алгоритм анализа коммуникации пользователей, разработанный на основе модели SMCRE. В [65] предложена методика поиска сообщений и источников в Twitter, основанная на модели Лассуэлла.

Линейная математическая модель Шеннона и Уивера

Практически в это же время в 1949 году математик Клод Е. Шеннон и инженер-электронщик Уоррен Уивер опубликовали «Математическую модель коммуникации» [51]. Авторы предложили модель точного количественного анализа процесса передачи информации. В своих исследованиях они выделяли три уровня проблем:

1) уровень А (технические проблемы);
2) уровень В (семантические проблемы);
3) уровень С (проблемы эффективности).

Графически модель Шеннона-Уивера может быть представлена следующим образом (рисунок 1.3)

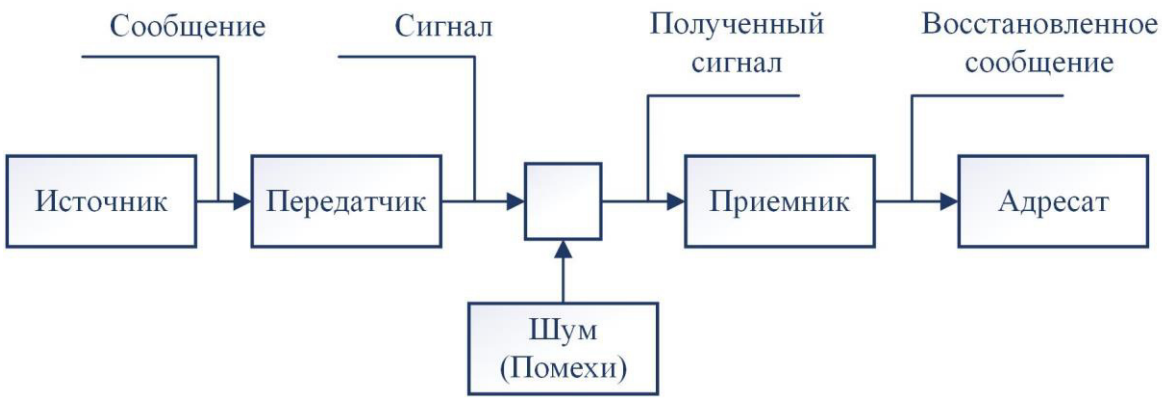


Рисунок 1.3 – Модель коммуникации Шеннона – Уивера

Один из очень важных элементов, который был введен авторами Шеноном и Уивером в их модели – это шум. Они же предложили типологию коммуникативных шумов: (1) Механические шумы – возникают за счет технических параметров канала, то есть среды, через которую проходит сигнал. (2) Семантические шумы – шумы, имеющие нетехническую природу. Они связаны со сложностью передачи контекста сообщения через текст. Семантические шумы в свою очередь делятся на две группы: шум на стороне источника, шум на стороне получателя.

Помимо «шума» в модели Шеннона-Уивера выделены отдельно такие элементы как передатчик и приемник. В дальнейшем Д. Файск в своей работе обосновал необходимость введения такого понятия как «посредник» (medium) [66]. Посредник, по Файску – это технические или физические средства преобразования сообщения в сигнал, который можно передавать по каналу.

Модель Шенона-Уинвера позволяет уточнить компоненты и задачи системы противодействия, например компонент противодействия каналам может быть расширен или дополнен таким объектом воздействия как посредник. Понятие посредник более широкое и включает в себя как связанные ресурсы в сети интернет, так и новостные агрегаторы, информационные видео каналы на видео хостинге, публичные страницы в социальных сетях и др.

В [67] предложен метод прогнозирования связей, основанный на теории Демпстера–Шейфера [68] и новый метод измерения предсказуемости связей с помощью локальной информации и энтропии Шеннона. В [69, 70] авторы предлагают модель, в основе которой лежит линейная модель Шенона-Уинвера.

A-B-X модель Теодора Ньюкомба

A-B-X модель Теодора Ньюкомба [50, 71] больше связана с такими науками, как социология, журналистика, лингвистика, психология

коммуникаций и рассматривает отношения между участниками коммуникаций и объектом и описывает влияние этих отношений на характер и результат коммуникативного взаимодействия. Однако тот подход, который предложили авторы позволяет расширить спектр задач и функционал системы за счет механизмов анализа обратной связи. Например, в работе [72] авторы предлагают модель для обнаружения источников и сообщений в СС, и одна из стратегий строится на базе A-B-X модели Т. Ньюкомба. В [73] авторы проводят исследование того, каким образом в социальных сетях студенты выбирают друзей и опираются также на рассматриваемую модель. В [74] авторы предлагают архитектуру системы обнаружения деструктивного воздействия в социальных сетях, при этом для понимания динамики социальных групп ими используется эта же модель.

Рассмотрим A-B-X модель подробнее, для этого также обратимся к [50].

A-B-X модель Теодора Ньюкомба отвечает на ряд вопросов, как и модель Лассвелла: (1) что побуждает субъектов к вступлению в коммуникацию; (2) каким образом влияют на коммуникацию отношения между субъектами; (3) какими будут возможные психологические, социологические эффекты для участников коммуникации.

В качестве базовой модели Ньюкомб рассматривал ситуацию элементарного коммуникативного взаимодействия, то есть диалога, в котором субъекты A и B вступают в коммуникацию по поводу некоторого внешнего по отношению к ним объекта X. При этом, X – это индивид, событие, сообщение, любая информация, любое сообщество. То есть X может быть некоторым индивидом, событием или сообщением, связанным с вредоносной информацией. Тогда в качестве A и B могут выступать также любые социальные субъекты – индивиды, социальные группы, социальные организации.

Согласно Ньюкомбу A и X объединяет некоторая тема, названная автором в его работе «ориентация». Ориентация, согласно Ньюкомбу может быть описана в виде позитивных (+) или негативных (-) аттитюдов

(attitude «отношение»). Понятие аттитюда в психологии и социологии связано с социальными установками, и под ними понимаются наборы убеждений, интересов субъекта [75]. В настоящем диссертационном исследовании аттитюд может быть выражен через положительное отношение к теме пользователем социальных сетей, либо отрицательное. Такая модель позволяет сегментировать источники и получателей на тех, кто солидарен и поддерживает тему X, связанную с вредоносной информацией и на тех, кто осуждает тему X.

A-B-X модель может быть представлена в виде следующей схемы (рисунок 1.4)

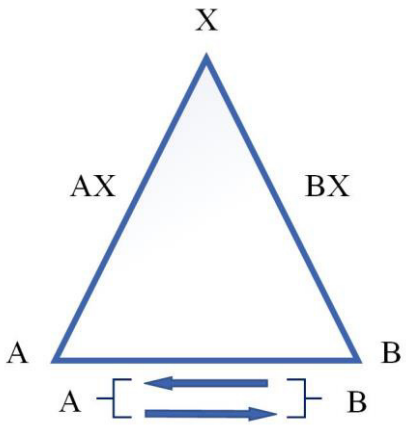


Рисунок 1.4 – A-B-X модель Теодора Ньюкомба

Условные обозначения A-B-X модели Теодора Ньюкомба:

A – субъект коммуникации A;

B – субъект коммуникации B;

X – объект коммуникации X;

AB – межсубъектный аттитюд коммуникации между AB;

BA – межсубъектный аттитюд коммуникации между BA;

AX – субъектно-объектный аттитюд AX;

BX – субъектно-объектный аттитюд BX.

Важно то, что согласно A-B-X модели в любой момент времени ориентация в коммуникации может быть симметричной и ассиметричной.

Вопросы симметрии и асимметрии коммуникации и ее эффектов активно развиваются в рамках исследований, направленных на изучение поведения пользователей социальных сетей (behavior analysis) [76].

Анализ A-B-X модели Теодора Ньюкомба позволяет уточнить и расширить набор признаков для алгоритмов анализа и оценки источников и методики противодействия вредоносной информации в социальных сетях: (1) связь между источниками; (2) связи между источником и получателями; (3) связи между получателями; (4) общие интересы источников и получателей; (5) общая информация между источниками; (6) общие информация между источниками и получателями; (7) общая информация между получателями; (8) связи между посредниками; (9) связи между посредниками и источниками; (10) связи между посредниками и получателями; (11) общая информация между посредниками; (12) общая информация между посредниками и источниками; (13) общая информация между посредниками и получателями.

Интегральная (обобщенная) модель Б. Вестли и М. Маклина

Как говориться в [50, 52], модель Вестли и Маклина является развитием идей Ньюкомба. Ученые Исследователи Б. Вестли и М. Маклин добавили в модель элемент, позволяющий учитывать потребность субъекта в информации при условии доступности разных источников (рисунок 1.5). В сегодняшних условиях – это тот набор источников, который доступен пользователю для получения информации, то есть веб-сайты, социальные сети, новостные агрегаторы, видео-хостинги и другое.

Авторы предполагают, что субъекты формируют свое информационное пространство с целью удовлетворения потребностей или решения проблем. Очевидно, что круг интересов субъекта ограничен набором ($Int_1, Int_2, ..., Int_n$).

Коммуникация инициируется тогда, когда некий субъект B осознает интерес, потребность в получении информации из информационного пространства. В модели Вестли-Маклина информационное пространство вокруг субъекта называют пространством выборов *Int1, Int2, ..., Intn.*

Предположим для выбора ориентации (аттитюда) выбран интерес *Int3,* тогда в зависимости от способов получения информации субъект может реализовать свою потребность через разные источники в современном цифровом пространстве.

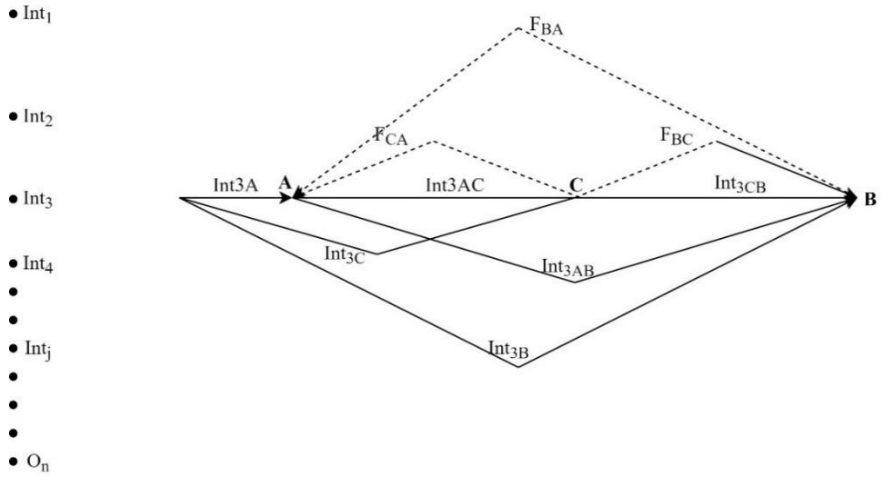


Рисунок 1.5 – Интегральная модель коммуникации Вестли – Маклина

В простейшем примере субъект имеет возможность просто получить информацию за счет прямого взаимодействия с событием, текстом (дуга *Int3B*), но при использовании современных веб-ресурсов, виде-хостингов, социальных сетей всегда есть источник доступа к информации – *A*. Возможно источник А непосредственно наблюдал, создавал объект с информацией на интересующую субъекта тему – *Int3*, тогда он формирует аттитюд для *B*, появляется дуга *Int3AB*. Еще одним вариантом, характерным для социальных сетей является наличие ретранслятора или репитера, которым является любой субъект, повторившим информацию из источника *A* об информации *Int3*. Авторы модели называют его «информационным посредником», или «информационным привратником» (gatekeeper). По сути, этот субъект и есть «media» – «между».

В отличии от A-B-X модели Теодора Ньюкомба авторы Б. Вестли и М. Маклин в своей модели учитывают наличие обратной связи. Для этого авторы делят коммуникацию на непосредственную и опосредованную. На рис. 1.5 это дуги *B* к *A* (дуга *FBA*), от *B* к *C* (дуга *FBC*) и от *C* к *A* (дуга *FCA*). В социальных

сетях примером обратной связи служат комментарии, ответы на комментарии, отметки «мне нравится» (лайк), отметки «мне не нравится» (диз-лайк), подписки и другие действия участников информационного обмена.

Модель Б. Вестли и М. Маклина (Westley and MacLean's Model) ложится в основу многих работ, в том числе посвященных исследованию вопросов обнаружения и противодействия распространению вредоносной информации [77].

Эпидемические модели SI, SIS, SIR, SIRS

В работе «Обнаружение источника слухов в социальных сетях» [53] авторы представили свой подход к систематизации направлений исследований в области анализа социальных сетей. В том числе исследователи утверждают, что задача обнаружения источника в социальной сети состоит в том, чтобы найти человека или узел, откуда изначально возникли такие сущности, как вирус или дезинформация. Также они предлагают таксономию, которая содержит различные аспекты (факторы): 1) сетевая структура; 2) модели распространения; 3) меры центральности; 4) оценочные метрики.

В работе [54] авторы решают задачу задача оценки источника инфекции для восприимчиво-инфицированной модели (Susceptible-Infected model, SI), в которой не все узлы инфицированные. Авторы [54] показывают, что для социальных сетей, чья структура больше напоминает дерево (Twitter, YouTube) оценка исходного узла, связанного с наиболее вероятным путем заражения, задается центром Жордана, то есть узлом с минимальным расстоянием до множества наблюдаемых зараженных узлов.

В работе [55] рассматривается модель распространения вредоносной информации SIS (от англ. Susceptible-Infected-Susceptible – восприимчивые-инфицированные-восприимчивые), согласно которой любой узел в социальной сети может быть заражен некоторой вредоносной информацией в процессе ее распространения и тогда он передает ее соседям, однако этот узел остается восприимчивым к схожей вредоносной информации от своих соседей.

В отличие от SI и SIS, модель распространения вредоносной информации SIR (от английского Susceptible-Infected-Removed) предполагает, что узел в сети может находиться в трех состояниях: восприимчив, инфицирован и удален. В [57] авторы рассматривают биоинспирированные подходы анализа социальных сетей, в том числе модель SIR. Предлагают таксономию для классификации вредоносного информационного контента на различных стадиях и распространенных технологий для решения этой проблемы на стадиях происхождения, распространения, обнаружения и локализации. В исследовании [78] ученые демонстрируют механизм распространения настроений на веб-форумах. Для этого авторы исследуют возможность применения эпидемической модели SIR к распространению настроений. В частности, в работе ими предложена модифицированная модель компетентности, которая имеет два отсека для выражения положительных и отрицательных настроений и отражает процесс распространения мнений на веб-форуме.

Согласно моделям SIRS восстановленный узел может снова стать восприимчивым узлом с некоторой вероятностью. В работе [79] авторы в рамках исследований, направленных на изучение эволюционного механизма и процессов расхождений во мнениях участников дискуссий, предлагают модель групповой поляризации, интегрированную в эпидемическую модель SIRS. В начале они вводят эпидемическую модель и определяют факторы силы отношений для усиления передачи информации и взаимодействия между индивидами, основанные на модели J-A, предложенной Ягером и Амблардом [80]. Кроме того, в работе используется Модель Барабаши-Альберта [81] для формирования случайных безмасшабных сетей.

Модели независимых каскадов и модели влияния в социальных сетях

Влияние (на английском языке Influence) – это процесс и результат изменения субъектом влияния поведения другого субъекта (индивидуального или коллективного объекта влияния), его мнения, намерений, представлений

и оценок и основывающихся на них действий в ходе взаимодействия с ним [82]. Влияние может рассматриваться со стороны воздействующей стороны и тогда это целенаправленное или ненаправленное влияние. Или со стороны реципиента: 1) требующее принятие решения (например, политические выборы), 2) не требующее принятия решений (например распространение вредоносной информации, тиражирование ее).

Исследования, направленные на обнаружение целенаправленного воздействия, делятся на те, в которых авторы ищут сообщества, основные узлы в сообществах и др. При этом само воздействие может содержать такие признаки как добавление или удаление агентов, добавление или удаление отношений, изменение характеристик агентов влияния, через которых происходит воздействие или изменениях характеристик отношений. Создание информационных объектов в сети также может являться признаком целенаправленного воздействия. Рассмотрим некоторые модели влияния подробнее.

Модель проникновения нововведений Эвертта Роджерса (innovation diffusion)

В 1962 году Э. Роджерс предложил модель проникновения инноваций, в 2003 году им была опубликована работа, в которой была представлена 5-я версия, усовершенствованная автором [83]. К 2003 году на базе его модели было проведено и опубликовано более 5000 исследований. В своей работе Э. Роджерс исследовал уровни принятия инноваций и обнаружил, что большая часть графиков напоминает кривую нормального распределения, разделенную на 5 частей. Эта модель до сих пор используется в исследованиях, направленных на обнаружение управления мнением или манипуляцией пользователя в социальных сетях, а также в исследованиях, посвященных распространению вредоносной информации [84].

Множественная модель проникновения нововведений Френка Басса

Согласно модели Ф. Басса [85] агенты в сети могут находиться, как в активном, так и не в активном состоянии, то есть у агентов бинарное состояние. При этом Басс выделяет 2 вида агентов:

1) новаторы;
2) имитаторы.

Математическая модель диффузии инноваций Френка Басса имеет вид (1.1):

$$n_t = (p + q \times \frac{N_t}{M}) \times (M - N_t), \quad (1.1)$$

где n_t – количество принявших инновацию в момент времени t; M – потенциал рынка; N_t – суммарное число принявших инновация в момент времени t; p – коэффициент внешнего влияния (реклама); q – коэффициент внутреннего влияния (межличностные коммуникации, рекомендации).

В работе [86] авторы предлагают новую модель, усовершенствованный вариант модели Ф. Басса, который позволяет выделять влиятельные каналы на YouTube в процессе распространения информации.

Пороговая модель Грановеттера (threshould model)

Пороговая модель часто применяется исследователями в разных областях. В 1978 году в статье [87] предложил модель, которая формально может быть описана следующим образом:

Пусть доля действующих агентов в момент времени t: $r(t)$

Зададим правила принятия решения агентом:

1) если $r(t) < 0$, то агент «бездействует»;
2) если $r(t) \geq 0$, то агент «действует».

Функция распределения порогов (1.2):

$$F_0(\cdot): [0; 1] \to [0; 1]. \quad (1.2)$$

Доля агентов с порогами, не превышающими $r(t)$, равно $F_0(r(t))$.

Следовательно, в момент времени $t + 1$ будут действовать (1.3):

$$r(t + 1) = F_0(r(t)), \quad (1.3)$$

это поможет спрогнозировать количество готовых присоединиться агентов, когда уже присоединилась часть.

Таким образом, положение равновесия характеризуется $r(t+1) = r(t)$, то есть $r = F_0(r)$.

В работе [88] авторы показывают насколько более точно линейная пороговая модель GLT (развитие модели Грановеттера) позволяет смоделировать процесс распространения информации в социальных сетях. В начале они сравнивают GLT с биоинспирированными подходами (SIS, SIRS), а затем предлагают объединить их, чтобы смоделировать гибридный процесс распространения, в котором простая инфекция накапливает критическую массу для сложной инфекции, которая приводит к глобальным каскадам.

Модель Грановеттера позволяет спрогнозировать развитие революционных ситуаций. В социологии модель получила сленговое название «Модель массовых беспорядков Грановеттера».

Модели независимых каскадов

Модели независимых каскадов относятся к моделям влияния, управления и противоборства и описывают информационный каскад, то есть распространение информации через соседей. Для блокировки распространения каскада между сообществами высчитывается максимальное значение доли связей агента вне сообщества, то есть если оно меньше порога, тогда каскад не пройдет. В моделях независимых каскадов выделяются «хабы» – вершины с большой степенью, такие вершины переключаются только, если большое количество их связей уже переключено в каскаде. Однако при переключении «хаба» идет высокий уровень влияния на большое количество агентов. Согласно моделям независимых каскадов одиночное изменение состояния больше влияет на небольшое множество [60].

Общую модель каскадов можно задать через модель байесовского обучения.

Пусть существуют изначальные «a priori» вероятности гипотез $p(h_1)$, $p(h_2)$, проводится эксперимент с получением подтверждения $p(e)$, производится переоценка вероятностей по формуле (1.4):

$$p(h_1|e) = \frac{p(e|h_1)p(h_1)}{p(e)}, p(h_2|e) = \frac{p(e|h_2)p(h_3)}{p(e)}, \quad (1.4)$$

где $p(e) = p(e|h_1)p(h_1) + p(e|h_2)p(h_2)$.

При получении подсказки предыдущего агента влияния в информационном каскаде о том, что принятие информации является хорошим решением, вероятность принятия информации новым агентом в каскаде повышается (1.5):

$$p(a|c) = \frac{p(c|a)p(a)}{p(c)} = \frac{pq}{pq+(1-p)(1-q)}, \quad (1.5)$$

где $ap + b(1-p) = 0$ –ожидание выгоды, при a – выгода правильного решения, b – потеря от неправильного решения. Подсказка к действию: c –«принять – это хорошее решение» с вероятностью (достоверностью) q, и d –«принять –это плохое решение». И $p(c|a) = q, p(d|b) = q, p(c|b) = 1 - 1, p(d|b) = 1 - q$.

При множестве подтверждений (1.6): m подтверждений c и n подтверждений d

$$p(a|c^m, d^n) = \frac{pq^m(1-q)^n}{pq^m(1-q)^n+(1-p)(1-q)^m q^n}. \quad (1.6)$$

В работе [59] авторы опираются на известные модели влияния, в том числе используют модели независимых информационных каскадов для определения наиболее влиятельных узлов в сети «хабов».

В работе [89] авторы предлагают модели, описывающие распространение информационной угрозы, информационного противоборства. В основу модели распространения информационной угрозы легли модели распространения инноваций, однако авторы делят информационные каналы на внешний и внутренний (по отношению к объекту влияния). Также авторы показывают, что скорость распространения информационной угрозы в сети ограничена только ресурсами противника [90].

К моделям информационного противоборства авторы относят модели, в которых сталкиваются два информационных потока и в конечном счете опираются на подходы, связанные с теорией игр [91].

Обзор релевантных моделей информационного обмена и противодействия вредоносной информации в социальных сетях показывает, что несмотря на то, что моделей великое множество, каждая из них разрабатывалась в рамках научных школ и концепций, которые не преследовали цели противодействия распространению вредоносной информации. Однако модели применимы для обнаружения источника, информационного канала в социальной сети и для моделирования встречных информационных потоков в рамках противодействия.

1.2.2 Обзор релевантных алгоритмов противодействия вредоносной информации в социальных сетях

Большинство исследователей в своих работах рассматривают как модели, так и алгоритмы, причем алгоритмы можно условно разделить по типам моделей, которые взяты учеными в основу:

1) биоинспирированные алгоритмы и подходы к обнаружению вредоносной информации;
2) алгоритмы, основанные на моделях независимых каскадов;
3) алгоритмы, основанные на моделях влияния;
4) алгоритмы, основанные на моделях информационного обмена.

Отдельно алгоритмы могут быть сегментированы по месту и роли в задачах мониторинга и противодействия вредоносной информации.

В работе [92] авторы предлагают структуру молодой науки анализа данных социальных сетей, в которой авторы выделяют основные блоки: 1) сбор данных из социальных сетей; 2) предобработка данных на основе BigData; 3) выбор метрик оценивания; 4) измерение социального влияния; 5) разработка алгоритмов максимизации влияния; 6) анализ производительности по соответствующему алгоритму или модели.

Алгоритмы противодействия вредоносной информации в социальных сетях могут быть также разделены по предложенной структуре. Однако блоки 1 и 2 относятся исключительно к системам мониторинга.

Рассмотрим алгоритмы противодействия вредоносной информации более подробно.

В исследовании [93] ученые отталкиваются от того, что эффективность противодействия вредоносной информации обнаружения в социальных сетях теоретически может быть повышена искусственными пользователями, называемых «псевдозащитниками», которые распространяют «антиинформацию». Роль этих «псевдозащитников» заключается в распространении клеветы, информации против вакцинации и др. Предполагается, что изначально существует единый источник информации и «псевдозащитник» действует, как противовес ему. Для обнаружения источника и «псевдозащитника» авторы предлагают следующий алгоритм обучения: а) анализ параметров распределения расстояний MLE (метод максимального правдоподобия) для обнаружения «псевозащитника» и б) обнаружение источника информации при MAPE (оценка апостериорного максимума) на основе изученных параметров.

В работе [94] авторы в начале строят временной агрегированный граф (от анг. – time aggregated graph, TAG [95]) и используют модель SIR для характеристики динамики диффузии каждого узла. Далее, если в сети обнаружен один источник заражения (влияния), авторы предлагают к использованию алгоритм обратного заражения на основе тегов (RI-TAG). Также авторы предлагают алгоритм для одновременной оценки источника информации и времени диффузии. Тогда, когда в сети обнаруживается несколько источников заражения (распространения вредоносной информации), авторы делят набор зараженных узлов на различные разделы, а затем запускают алгоритм их оценки с одним источником в каждом разделе.

В работе [96] авторы предлагают модель, алгоритмы и методику контроля слухов за счет создания сети консультантов, которым доверяют

пользователи. Для выбора доверенного консультанта авторы предлагают использовать модель распространения эпидемий SI, для противодействия вредоносной информации – теорию игр.

Авторы [97] предлагают алгоритм динамического периода блокировки в качестве меры противодействия вредоносной информации в социальных сетях. Предлагаемый метод позволяет выбирать и блокировать узлы, которые с наибольшей вероятностью распространяют большое количество слухов и поддерживают их. В отличие от существующих методов, предлагаемое решение не блокирует узлы на неограниченный срок и этот срок оценивается по высокой активности узла в СС.

В работе [98] авторы используют алгоритмы, основанные на моделях информационных каскадов для обнаружения кластера ботов в социальной сети Twitter, которые распространяли информацию о референдуме Brexit в Великобритании. В работе [99] также демонстрируется результат исследования политической повестки дня, то есть распространения слухов в Twitter во время предвыборной компании в США в 2012 году. Авторы опираются на модели независимых информационных каскадов.

В работе [100] предлагается модель машинного обучения и алгоритмы, в которых за основу берется пост в социальной сети и формируются наборы данных для обучения модели, функцией которой является различение искусственной компании от живой активности в социальных сетях. Исследователи проводят контент анализ каждого сообщения в наборе данных.

В работе [101] предлагается алгоритм противодействия распространению вредоносной информации, основанный на разбиении подграфов, в котором авторы ищут некоторые важные узлы для блокирования. Основной алгоритм состоит из двух основных фаз. Во-первых, дается метод разбиения подграфов на основе структуры сообщества для быстрого извлечения структуры сообщества сети распространения информации. Во-вторых, в полученных подграфах предлагается к использованию авторский алгоритм блокировки и уточнения узлов, основанный на Жордановом центре.

Широкий корпус российских исследований в области противодействия вредоносной информации в социальных сетях представлен такими работами как [24, 102-106] и др. В обобщенном виде обзор моделей и алгоритмов может быть представлен в виде сводной таблицы (таблица 1.1).

Таблица 1.1 − Модели и алгоритмы противодействия вредоносной информации в социальных сетях

Релевантные модели и алгоритмы	**Релевантные работы**
Модели информационного обмена: **(1)** Линейная модель коммуникации SMCRE **(2)** Линейная математическая модель Шеннона и Уивера **(3)** A-B-X модель Теодора Ньюкомба **(4)** Интегральная (обобщенная) модель Б. Вестли и М. Маклина	**(1)** Wojtasik K 2018 [63], Shao X. 2019 [64], Lipschiltz J.H. 2017 [65]; **(2)** Yin L 2017 [67], Koohikamali M 2017 [69], Mirsarraf M. 2017 [70]; **(3)** Guo Y.H. 2018 [72], Small M.L. 2020 [73], Охапкин В.П. 2020 [74]; **(4)** Haug, M 2020 [77].
Эпидемические модели: **(1)** SI **(2)** SIS **(3)** SIR **(4)** SIRS	**(1)** Shelke S. 2019 [53], Luo W. 2014 [54], George B., 2016 [95]; **(2)** Zh Wang 2015 [55]; **(3)** Meel P., 2021 [57], Choi Minnseok 2017 [78], Chai Y. 2021 [94]; **(4)** Chen T. 2020 [79], Jager, W. 2004 [80], Albert R. 2002 [81].
Диффузионные модели влияния и независимых каскадов: **(1)** Модели влияния; **(2)** Модель проникновения нововведений Э. Роджерса (innovation diffusion); **(3)** Модель проникновения проникновения нововведений Френка Басса; **(4)** Пороговая модель Грановеттера (threshould model); **(5)** Модель информационного каскада.	**(1)** Губанов Д.А. 2009 [82], Choi J. 2019 [93], Губанов Д.А. 2020 [103]; **(2)** Forest J. 2009 [84]; **(3)** Susarla A. 2012 [86]; **(4)** Ran Y. 2020 [88]; **(5)** Kumari A. 2017 [59], Cheng J. 2014 [60], Bastos M.T. 2019 [98], Shin J. 2018 [99].
Другие модели и алгоритмы **(1)** Теория игр; **(2)** Модели машинного обучения; **(3)** Модели противоборства.	**(1)** Марцева Н.А. 2010 [91], Расторгуев С.П. 2014 [102]; **(2)** Varol O. 2017 [35], Ferrara E. [33], Alizadeh M. 2020 [100], Kotenko I. 2017 [104], Kotenko I. 2018 [106]; **(3)** Макаренко С.И. 2014 [24], Михайлов А.П. 2005 [90], Askarizadeh M. 2021 [96], Hosni A.I.E. 2020 [97], Yuan D. 2020 [101].

На основании проведенного исследования моделей информационного обмена и противодействия вредоносной информации рассмотрим концептуальную модель противодействия вредоносной информации и ее связь с системой мониторинга (рисунок 1.6)

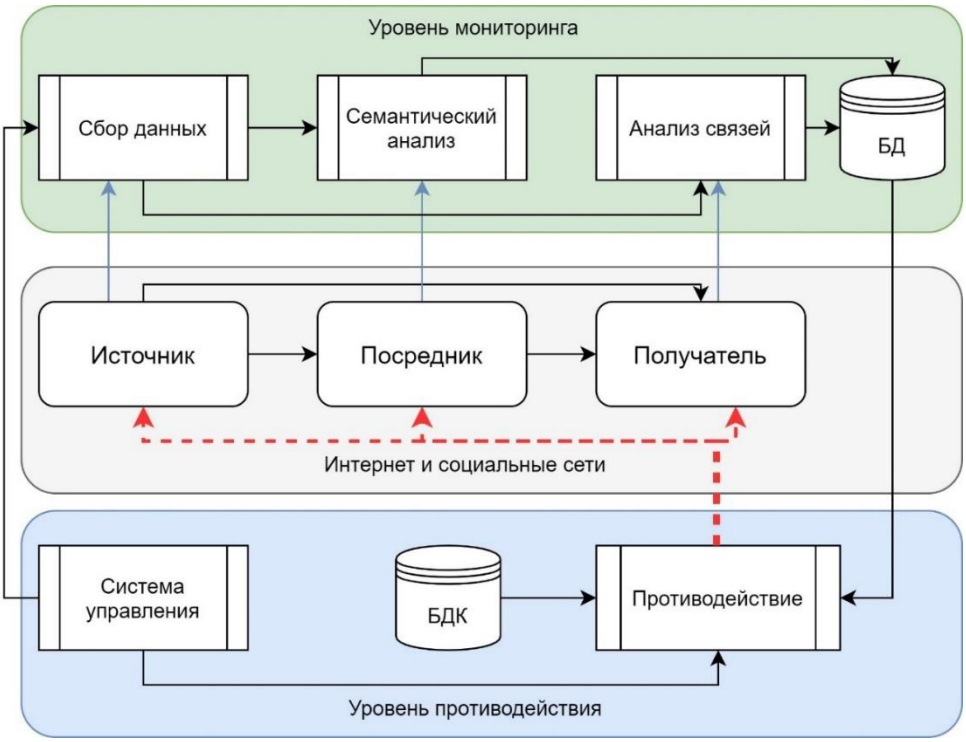


Рисунок 1.6. – Концептуальная модель системы противодействия вредоносной информации в социальных сетях

Модель включает 2 уровня: 1) Уровень мониторинга; 2) Уровень противодействия. Между ними находится информационное пространство, в котором происходит информационный обмен.

Формирование списка информационных угроз, содержащих вредоносную информацию, происходит на уровне противодействия. Сбор данных, семантический анализ текстов, анализ связей происходит на уровне мониторинга. Рассмотрим существующие подходы и решения подробнее.

1.2.3 Обзор релевантных систем противодействия вредоносной информации в социальных сетях

В параграфах 1.2.1, 1.2.2 диссертации автор рассмотрел модели алгоритмы и предложил концептуальную модель системы противодействия вредоносной информации в СС, в которой выделяются 3 уровня (рис. 1.6). Далее в работе рассматриваются релевантные методики и системы противодействия ВИ.

Прежде чем говорить о существующих решениях в области противоборства, разделим процесс обработки цифрового сетевого контента в социальных сетях на 2 этапа:

Этап 1. Мониторинг информации в социальных сетях.

Этап 2. Противодействие вредоносной информации в социальных сетях.

Каждый этап процесса может быть представлен в виде линейного алгоритма (рис. 1.7).

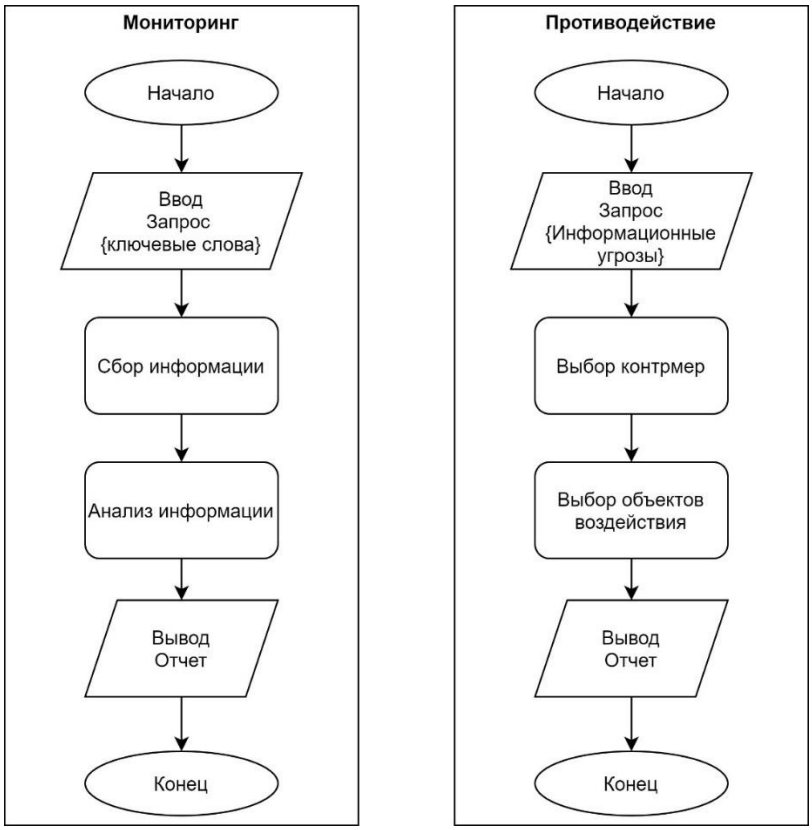


Рисунок 1.7. – Алгоритмы мониторинга и противодействия в процессе обработки цифрового сетевого контента в социальных сетях

1.2.3.1 Обзор систем мониторинга вредоносной информации.

Согласно документу «Рекомендации по стандартизации Р 50.1.053-2005. Информационные технологии. Основные термины и определения в области технической защиты информации (утвержден Приказом Ростехрегулирования от 06.04.2005 № 77-ст)» [107] мониторинг безопасности информации (при применении информационных технологий) – это процедуры регулярного наблюдения за процессом обеспечения безопасности информации при применении информационных технологий.

Согласно стандарту «ГОСТ Р 50922-2006 Защита информации. Основные термины и определения» мониторинг безопасности информации (при применении информационных технологий) [108] – это постоянное наблюдение за процессом обеспечения безопасности информации в информационной системе с целью установить его соответствие требованиям безопасности информации.

К таким системам относятся системы мониторинга состояния ИТ-инфраструктуры предприятия. К ним можно отнести и системы обнаружения вторжений, ситуационные центры, системы обнаружения аномалий и др.

Однако существует также такие решения как системы мониторинга сети Интернет и социальных сетей. Эти системы более релевантные для данного исследования, так как они работают со схожей средой и по схожему алгоритму мониторинга. К таким системам могут быть отнесены следующие известные Российские и зарубежные решения: IQBase [109], YouScan [110], Avalanche Online («Лавина Пульс») [111], Brand Analytics [112], ПрессИндекс [113], Hootsuite, Google Alerts, Hootsuite Insights Powered by Brandwatch (HIPB), Talkwalker, Reddit search, Reputology, Synthesio, Mentionlystics, Hootsute RSS Syndicator, RewiewTrcakers, Reviewlnc, NetBase, Nexalogy [114].

Также к системам мониторинга можно отнести и информационные системы поисковых систем: Yandex, Google, которые предоставляют свои сервисы для сбора и анализа информации по запросу пользователя.

К специальный системам мониторинга можно отнести решения, разработанные компанией БалтИнфоКом, такие как Зверобой, Октопус [115].

С позиции маркетинга мониторинг сети интернет и социальных сетей – это процесс мониторинга на предмет получения информации, относящейся к бизнесу (бренду):

1. Упоминания о бренде (с прямым тегированием или без него, он же @mention). Для исследования аналогией может быть упоминание вредоносной информации. Например, упоминание конфиденциальной информации, запрещенной к распространению и таким образом являющейся вредоносной.

2. Соответствующие хэштеги. Для исследования аналогией могут использоваться хештеги, содержащие прямое указание на вредоносную информацию. Например, самым популярным хештегом для привлечения внимания к «суицидальным» группам в социальной сети Вконтакте был #Синийкит.

Таким образом, мониторинг сети интернет и социальных сетей – это сбор информации, то есть сбор всех данных и деталей, которые можно собрать. Современные мониторинговые системы позволяют фиксировать то, что люди пишут по любому поводу или событию. Можно измерять вовлеченность в информационный повод, информационный поток или событие, создавать отчеты о том, что происходит в социальном пространстве.

В сущности, все системы собирают тексты сообщений, иногда – связи между сообщениями, дополнительно могут анализировать социальные сети – связи между пользователями, практически всегда такие системы собирают из сообщения – ссылки на сторонние ресурсы. Развитие современной коммуникации позволяет пользователю реагировать на сообщение, оставлять фидбек, поэтому зачастую системы могут видеть также статистику по обратной связи пользователей.

Использование всех рассмотренных программных решений и их аналогов возможно только на коммерческой основе, что затрудняет их применение в процессе мониторинга вредоносной информации. При этом необходимо отметить тот важный факт, что собрать и анализировать тысячи и сотни тысяч сообщений с вредоносной информацией без соответствующего программного обеспечения очень сложно. Все рассмотренные решения анализируют только ограниченное количество сообщений, например 5000 шт. [109] или 25000 шт. [116].

В рамках диссертации вопросы мониторинга информационного пространства и анализа текстов не рассматриваются и отделены от проблемы противодействия вредоносной информации.

1.2.3.2 Обзор систем противодействия вредоносной информации в социальных сетях.

В век развития цифровых технологий информация является важнейшим элементом в жизни людей, организаций и государств. Проблематикой ее защиты заняты различные научные коллективы и коммерческие компании, рядовые пользователи и официальные лица. Каждый выстраивает свою защиту в сети на основе тех инструментов, которые ему известны и доступны. Помимо законодательных мер, применяемых государством, такими инструментами являются операционные системы, антивирусы, браузеры и различные приложения, а также специальные сервисы и расширения.

Прежде всего условно разделим все существующие решения для противодействия вредоносной информации в социальных сетях на 2 группы:

1) группа 1 «Противодействие вредоносной информации» на стороне пользователя;

2) группа 2 «Противодействие вредоносной информации» на стороне информационной системы, компании или государства.

Первая группа решений

К первой группе можно отнести такие инструменты как расширения в браузерах, например «Adblock Plus» [117]. Или расширения в браузере Яндекс: «Антишок» [118], «Блокировка мешающей рекламы», Adguard. Большая часть критериев вредоносной рекламы разрабатывается специальными ассоциациями, например Яндекс.Браузер фильтрует рекламные информационные объекты, следуя рекомендациям IAB Russia [119]. Корпорация Google также на территории России предоставляет собственный браузер с возможностью дополнительной настройки расширений для защиты от вредоносной информации. В русской версии Google Chrome может быть настроена защита от информационных объектов на сайтах, которые показывают навязчивую или вводящую в заблуждение рекламу (список должен пополняться вручную пользователем браузера). Существует целый ряд дополнительных сервисов, отдельных расширений для Google Chrome: родительский контроль – блокировка порно сайтов [120]; антикремлебот – подсветка ботов [121]; MetaBot – подсветка ботов в YouTube [122]; Site blocker – ограничение доступа к сайтам по списку (вводится вручную пользователем) [123].

Рассмотрим таблицу с основными встроенными возможностями противодействия вредоносной информации в самых популярных браузерах с учетом только персональных компьютеров (таблица 1.2) [124].

Таблица 1.2 − Встроенные возможности защиты от вредоносной информации через браузеры

Браузер	**Основные встроенные возможности защиты от нежелательной, сомнительной и вредоносной информации**
Google Chrome	Система черных списков для фишинговых сайтов и ресурсов с вирусными угрозами. Блокировщик рекламы и баннеров. Большое количество дополнений под различные задачи, в том числе и антивирусы.
Яндекс браузер	Встроенный блокировщик рекламы, способный закрывать всплывающие баннеры. Возможность установки расширений от сторонних производителей (родительский контроль, антивирус, антиспам).

Opera	Встроенный сканер фишинговых модулей, блокировщик рекламных баннеров. Блокировка подозрительных скриптов на сайтах. Возможность установки расширений от сторонних производителей (родительский контроль, антивирус, антиспам).
Mozilla Firefox	Возможность установки расширений от сторонних производителей (родительский контроль, антивирус, антиспам).

Можно сделать вывод, что защитой от рекламы и баннеров обладают все современные браузеры, кроме того возможна настройка защиты функций родительского контроля, но только для защиты от просмотра веб-ресурсов, содержащих контент откровенного сексуального характера. Ни один браузер не содержит собственных функций анализа контента, выявления такой информации и противодействия ей.

Большинство браузеров имеют встроенное расширение антивируса или могут взаимодействовать с установленным в системе. В своей работе [125] Э.И. Балау с соавторами пишет, что основная задача антивируса – не допустить проникновения в систему вредоносных программ, предотвратить заражение системы вредоносным кодом. Ключевыми методами работы антивирусных программ являются сканирование сигнатур вирусов, проверка целостности и сканирование подозрительных команд, а также протоколировании всех событий, угрожающих безопасности системы [125].

К примеру, антивирус Kaspersky Internet Security – обладает функцией «Родительский контроль», что позволяет для каждой учетной записи на ПК установить ограничения. Он позволяет указать страницы, на которые вход будет запрещен, или выбрать определенные тематики, заблокировав тем самым доступ к сайтам с таким содержимым. В патенте RU № 2651252 [126], принадлежащем лаборатории Касперского описан метод защиты пользователя социальной сети от запрещенных объектов (текст, фотография, медиафайл), сохраненных ранее в базе данных. Настоящий технический результат достигается с помощью построения социального графа, кластера из аккаунтов в социальной сети, базой запрещенных объектов, которые связаны с социальными графами и аккаунтами. Изобретение относится к системам

родительского контроля, ограничивающим доступ пользователя к ресурсам социальной сети в соответствии с правилами, которые задал родитель.

Подобные инструменты противодействия вредоносной информации предлагает и продукт «Доктор Веб». Как сказано в [127] Его «Модуль родительского контроля» от этой компании не только ограничивает доступ к сайтам, где размещен контент 18+, но и вычисляет ресурсы локальной сети и папки на ПК, в названии или описании которых есть слова, связанные с индустрией развлечений для взрослых.

Современные операционные системы (ОС) позволяют также настраивать функции родительского контроля для противодействия вредоносной информации. В самую популярную для компьютеров систему − Windows, интегрирована программа Defender, являющаяся усовершенствованным вариантом антивируса. Операционная система IOS для мобильных устройств поддерживает функции защиты нецензурной лексики, ограничения доступа к контенту 18+.

Вторая группа решений

Ко второй группе систем противодействий относятся сами платформы социальных сетей, как информационные системы. Популярность социальных сетей в России в 2020 году претерпевает изменения, так по данным актуальным на июль 2020 года лидирует Youtube [128], однако в разрезе более длительного периода статистика следующая (рис. 1.8).

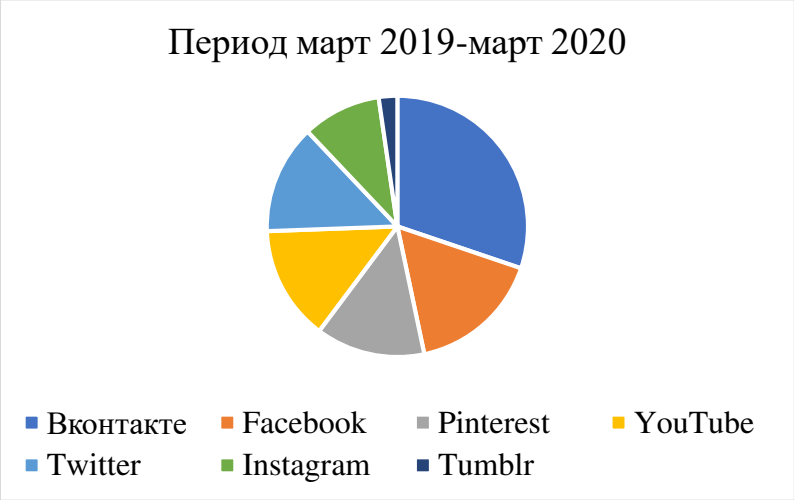


Рисунок 1.8. – Популярность социальных сетей в России период март 2019-март 2020

По своей архитектуре социальные сети, популярные в России представляют собой многокомпонентные решения, в архитектуре которых находится несколько компонентов, которые осуществляют обработку и отдельные компоненты, которые обеспечивают функции: администрирование, маркетинг, разработка, хранение данных. Такие платформы не содержат отдельного компонента противодействия вредоносной информации. Однако платформы социальных сетей со своей стороны контролируют соблюдение закона об авторском праве (Facebook, Instagram, YouTube), защиту от вредоносной информации в зависимости от возрастных ограничений, если пользователь верно указал возраст.

Facebook, в свою очередь, разрабатывает и активно внедряет свои собственные встроенные системы мониторинга информационного потока для защиты своих пользователей. Так, например, в работе [129] предлагается метод оценки эффективности экспериментальных итераций для систем верификации социальных сетей. Этот метод позволяет сокращать объем размеченных данных человеком, используемых в процессе распознавания искусственного поведения со стороны автоматизированных аккаунтов (ботов). В 2020 году компания Facebook объявила конкурс научных проектов,

направленных на распознавание заведомо ложной информации, фейковых новостей. Из 1000 заявок из 77 стран мира были поддержаны 25 проектов [129].

Facebook и Вконтакте имеют схожие внешне страницы, общую логику графового представления связей, однако первым был получен патент у компании Facebook [130] (рис. 1.10). А архитектура социальной сети Вконтакте представляет из себя кластер серверов: 1) Front-сервер; 2) Backend; 3) Content Server, 4) pu/pp (photo upload, photo proxy); 5) Sun (распределение нагрузки); 6) Cache; 7) proxy; 8) engines (система управления); 9) Data Base (рис. 1.9) [131].

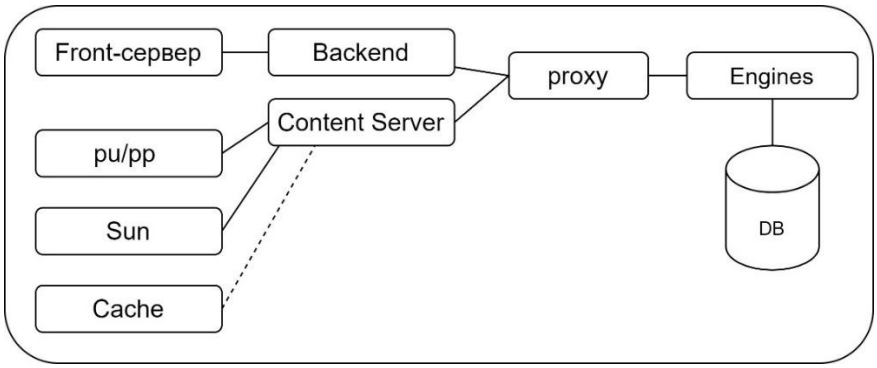


Рисунок 1.9 – Архитектура социальной сети Вконтакте

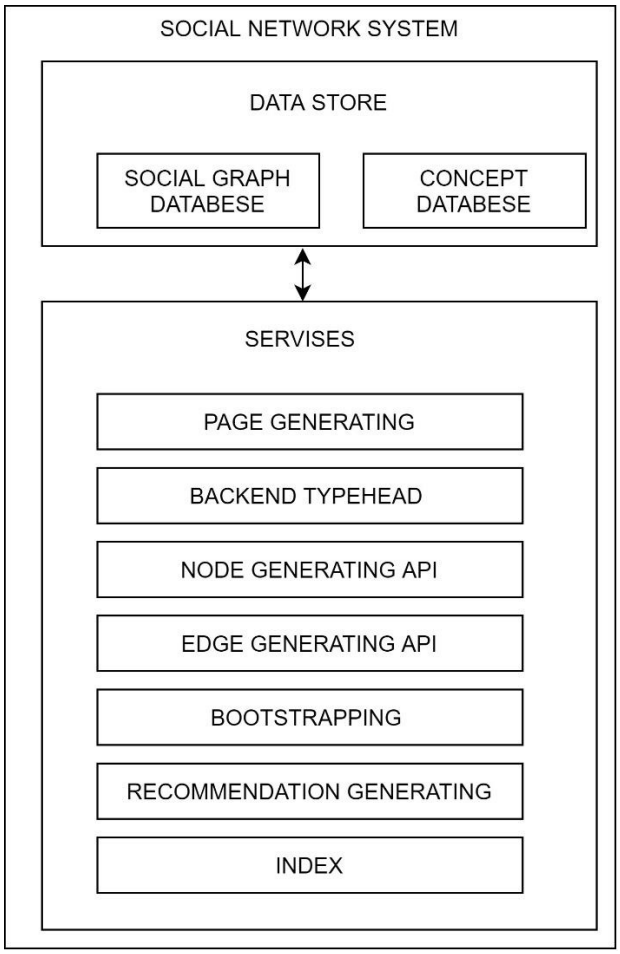


Рисунок 1.10 – Архитектура социальной сети Facebook

Анализ архитектуры платформ социальных систем показывает, что отдельного компонента противодействия у них нет.

Государство обеспечивает безопасность общества и индивида от вредоносной информации. Для этого реализованы такие механизмы, как ограничение к ресурсу через оператора связи путем внесения адреса URL –Единый реестр доменных имен, указателей страниц сайтов в сети «Интернет» и сетевых адресов, позволяющих идентифицировать сайты в сети «Интернет», содержащие информацию, распространение которой в Российской Федерации запрещено [132]. Решение о включении в реестр может быть обжаловано владельцем сайта в сети «Интернет», провайдером хостинга, оператором связи, оказывающим услуги по предоставлению доступа к информационно-телекоммуникационной сети «Интернет» в суде в течение трех месяцев со дня принятия такого решения. Однако система

противодействия государственная не является полностью автоматизированной и требует ручной обработки и анализа материалов экспертом до того, как они попадут в список запрещенных. Государственная система противодействия не предполагает в рамках реестра поиска источника распространения, копий, дубликатов, рерайтов и анализа обратной связи от пользователя сети интернет и социальных сетей.

Существуют научные работы, в которых прямо или косвенно описаны методики противодействия вредоносной информации. Так, например, в [133] предполагается методика создания списков подозрительных объектов и участников социальной сети, с которыми в дальнейшем сверяется каждый новые объект и на основе полученных данных ограничивается либо разрешается доступ пользователя СС к новому информационному объекту. В [134] предлагают комплекс из связанных устройств, системы и методики, которые позволяют родителям отслеживать действия своих детей с помощью смартфонов и анализировать их. Система позволяет блокировать ребенку видимость вредоносной информации.

В [135] предложен подход, при котором система обработки информации создает первый лингвистический профиль, соответствующий учетной записи пользователя в тот момент, когда было создано первое вредоносное сообщение в одной социальной сети. Далее система обработки информации вычисляет накопленный балл для пользователя на основе корреляции первого языкового профиля со вторым языковым профилем в другой сети. В [136] предложена методика обнаружения спамеров и поддельных профилей в социальных сетях, согласно которой из базы данных существующих профилей выбираются отрицательные примеры, положительные. Далее извлекаются признаки и обучаются модели машинного обучения (обучение с учителем).

В работе [137] предложена методика и метрики расчета вовлеченности пользователей исследуемого ресурса для системы мониторинга. В [138] авторы предлагают методику сдерживания распространения вредоносной

информации с использованием настраиваемых каналов разведки. Согласно их методики предусматриваются оценки достоверности контента путем создания онтологии и выбора темы по ключевым словам. Оценка достоверности осуществляется при помощи обученных моделей машинного обучения и группы экспертов.

В [139] предложена архитектура системы фильтрации спам сообщений в социальных сетях, основанная на rule base методе. В [139] описана архитектура системы мониторинга вредоносной информации в социальных сетях, согласно которой происходит контент анализ текстов и устанавливаются связи между постами автора вредоносного сообщения в разных сетях.

Подводя итог обзора релевантных моделей, алгоритмов, методик и архитектур можно сказать, что несмотря на множество разрозненных работ, сервисов, методов комплексного противодействия от вредоносной информации нет ни со стороны пользователя ни со стороны платформы социальной сети. Проблема противодействия имеет крайне малое количество научно-технических решений. Известные средства обнаружения и противодействия вредоносной информации в СС не отвечают требованиям к скорости, полноте, точности и адекватности принимаемых решений. Это обусловлено несколькими причинами. Во-первых, системы разделены два несвязанных модуля: (1) мониторинг; (2) противодействие. Посередине между ними находится оператор (рис. 1.11). Во-вторых, СС имеют сложную структуру и состоят из множества разнородных сообщений, что недостаточно учитывается при выборе цели противодействия, например тип сообщения, источник и другие характеристики. В-третьих, в реальном масштабе времени необходимо обрабатывать сверхбольшие объемы сообщений и в сжатые сроки выбирать цель для контрмеры, в ручном режиме оператор системы противодействия не в состоянии остановить распространение вредоносной информации.

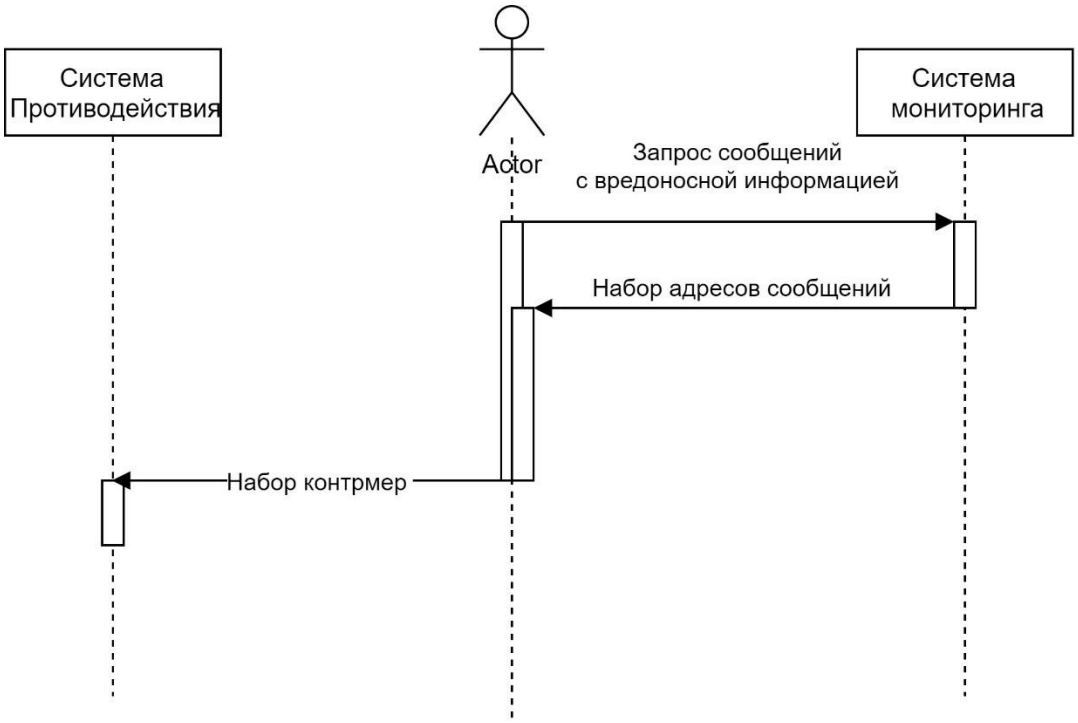


Рисунок 1.11 – Схема противодействия вредоносной информации в социальных сетях

На рисунке демонстрируется зависимость связи между оператором, экспертом, который всегда находится между двух раздельных систем, независимо от того каким образом реализуется противодействие (система родительского контроля или черные списки для операторов связи). Таким образом, основная сложность противодействия вредоносной информации в СС напрямую следует из современных тенденций развития информационной сферы. Это обуславливает необходимость разработки новых моделей, алгоритмов, методики и архитектуры для повышения эффективности противодействия вредоносной информации в социальных сетях.

1.3 Требования к системе противодействия вредоносной информации в социальных сетях.

Сравнительный анализ исследовательских работ в области противодействия вредоносной информации в социальных сетях позволил определить требования к системе противодействия, в основу реализации

которой должен быть положен модельно-методический аппарат, разрабатываемый в настоящей работе.

Прежде всего рассмотрим требуемый ландшафт функциональности системы противодействия (рис. 1.12)

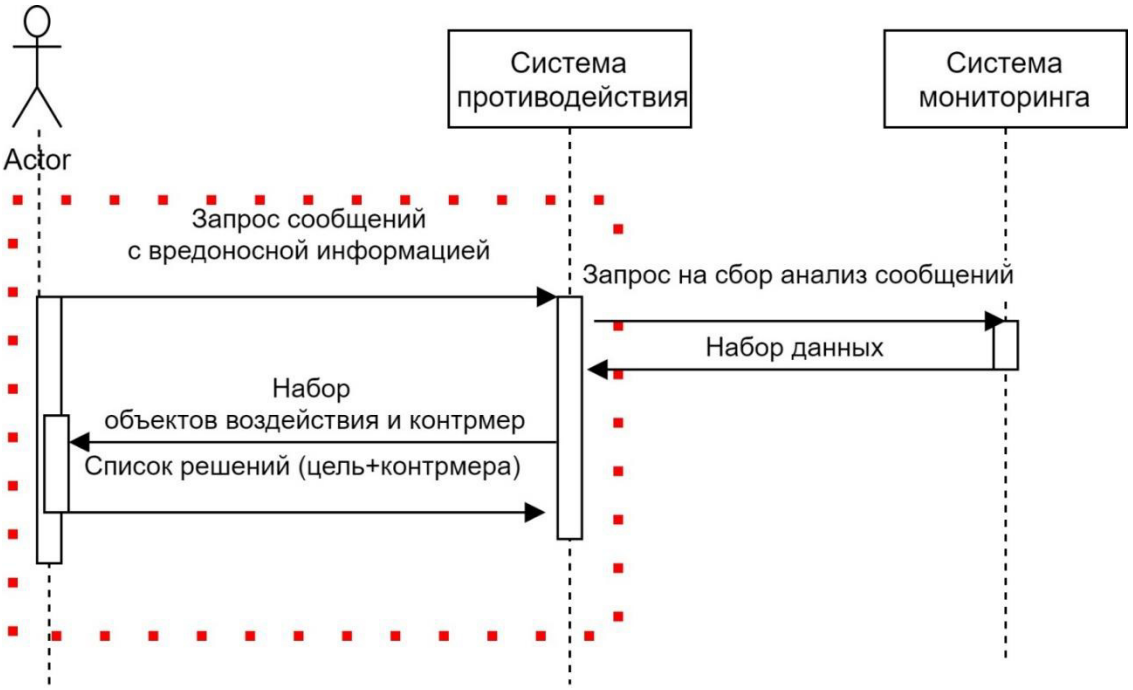


Рисунок 1.12 – Ландшафт функциональности системы противодействия вредоносной информации в социальных сетях

Как показано на диаграмме, система противодействия может быть центральным элементом в общем процессе. Процессы в системе противодействия могут быть автоматизированы за счет разработки алгоритмов и программных компонентов.

Для формирования требований к системе противодействия разделим требования на две группы: 1) функциональные; 2) не функциональные.

Функциональные требования к системе могут быть реализованы путем разработки архитектуры, компонентов и программных прототипов. Нефункциональные требования к системе противодействия могут быть реализованы путем разработки моделей и алгоритмов [141].

Функциональные требования представляют собой перечень функций, которые должна выполнять система. Нефункциональные требования

описывают целевые характеристики системы, такие как оперативность, требования по обоснованности и ресурсопотреблению и т.д.

Определим множество функциональных требований к системе противодействия вредоносной информации следующим образом:

- возможность формирования задачи на сбор сообщений, дополнительных данных и анализ сообщений для системы мониторинга;
- возможность настройки доступных мер противодействия в системе;
- возможность анализа источников сообщений в полученном наборе данных;
- возможность ранжирования и сортировки объектов воздействия в полученном наборе данных;
- возможность ранжирования и сортировки доступных контрмер из базы контрмер для каждого набора данных;
- возможность выбора цели воздействия для противодействия;
- генерация отчетов о полученных результатах в виде, адаптированном для эксперта по информационной безопасности;
- генерация отчетов о работе системы в виде, адаптированном для администратора системы;
- учет специфики режима работы системы (настройка, эксплуатация).

Система противодействия должна учитывать успешное применение существующих подходов противодействия вредоносной информации и анализа источников [126, 132-136, 138].

Как сказано в работах [140, 141], множество нефункциональных требований к системе противодействия вредоносной информации можно определить как три классические компоненты эффективности.

Оперативность – время, необходимое для противодействия вредоносной информации в социальных сетях. Требование к оперативности задается в виде (1.7):

$$T_m \leq T_s, \qquad (1.7)$$

где T_m – время, необходимое для противодействия вредоносной информации в социальных сетях с использованием предлагаемой методики, S – множество систем противодействия.

Для того, чтобы система противодействия вредоносной информации в социальных сетях использовалась в режиме, близком к существующим аналогам, она должна обеспечить противодействие за время, не превышающее некоторой границы. Такое требование к оперативности задается следующим образом (1.8) [142]:

$$P_{operability}(T_m \leq T^{acceptable}) \geq P_{operability}^{acceptable}, \quad (1.8)$$

где $P_{operability}$ – вероятность противодействия вредоносной информации за заданной время, $T^{acceptable}$ – допустимое время противодействия, $P_{operability}^{acceptable}$ – допустимое значение вероятности.

Основываясь на результатах опросов экспертов и серии проведенных исследований, было выбрано $T^{acceptable}$ = 102 минутам.

Обоснованность – совокупность учитываемых параметров для выбираемых объектов воздействия и контрмер в процессе противодействия вредоносной информации. Требование к обоснованности является основным в данной диссертации и задается в виде (1.9):

$$\begin{aligned} &N_{param} \rightarrow max, \\ &N_{param} > max N_{param}^{s}, \end{aligned} \quad (1.9)$$

где N_{param} – количество учитываемых параметров при выборе объекта воздействия и контрмер, S – множество систем противодействия, N_{param}^{s} – количество учитываемых параметров для системы $s \in S$.

Таким образом, разрабатываемая система должна учитывать большее количество параметров для объектов воздействия и контрмер, нежели существующие аналоги. Это позволит говорить о том, что новая система превосходит существующие аналоги качеству анализа.

Ресурсопотребление, как сказано в [142], характеризует номенклатуру и количество необходимых программных и аппаратных средств, кадровые

и другие ресурсы, затрачиваемые на реализацию процесса противодействия вредоносной информации. Требования к ресурсопотреблению задаются следующим образом (1.10):

$$P_{\text{res}}\left(r \le R^{\text{acceptable}}\right) \ge P_{res}^{\text{acceptable}}, \quad (1.10)$$

где P_{res} – вероятность того, что ресурсы, затрачиваемые на противодействие вредоносной информации r, не превышают допустимого значения $R^{\text{acceptable}}$, $P_{res}^{\text{acceptable}}$ – допустимое значение вероятности. Для выполнения противодействия вредоносной информации предполагается выделение отдельного компьютера и одного оператора. При этом на компьютере часть ресурсов будет занята операционной системой, а в рабочем времени оператора – 25% будет уходить на другие обязанности. Поэтому $R^{\text{acceptable}} = 0{,}75$, то есть процесс противодействия вредоносной информации должен занимать не более 75% от общего объема ресурсов.

Все перечисленные свойства и требования к ним приведены в таблице 1.3

Таблица 1.3. − Требования к системе противодействия вредоносной информации в социальных сетях

Свойство	**Показатели**	**Требования**
Оперативность	Время, необходимое для противодействия вредоносной информации	$T_m \le T_s$ $P_{operability}(T_m \le T^{acceptable}) \ge P_{operability}^{acceptable}$
Обоснованность	Совокупность учитываемых параметров для выбираемых объектов воздействия и контрмер в процессе противодействия	$N_{param} \to max$ $N_{param} > maxN_{param}^{s}$
Ресурсопотребление	Вероятность того, что количество использованных ресурсов не будет превышать допустимое значение	$P_{\text{res}}\left(r \le R^{acceptable}\right) \ge P_{res}^{acceptable}$

Таким образом, целевой функцией разрабатываемой системы является максимизация параметра обоснованности с учетом требований к оперативности и ресурсопотреблению.

1.4 Постановка задачи исследования

Сформулирована задача исследования. Она заключается в разработке: (1) комплекса моделей социальной сети, источника и вредоносной информации; (2) комплекса алгоритмов анализа источников вредоносной информации в социальных сетях и ранжирования контрмер; (3) методики противодействия вредоносной информации в социальных сетях с учетом требований к обоснованности; (4) архитектуры и программных прототипов компонентов системы противодействия вредоносной информации в социальных сетях.

Целью исследования является повышение эффективности противодействия вредоносной информации в социальных сетях. В диссертации показатель эффективности определяется через показатель обоснованности, а также с учетом требований к оперативности и к ресурсопотреблению.

Целевой функцией методики противодействия вредоносной информации в социальных сетях с учетом требований к обоснованности является максимизация количества учитываемых параметров для выбираемых объектов воздействия и контрмер в ходе противодействия вредоносной информации в социальных сетях $N_{param} \to max$, при соблюдении требований к другим свойствам.

К оперативности [142] $P_{operability}(T_m \leq T^{acceptable}) \geq P_{operability}^{acceptable}$, где $P_{operability}^{acceptable} = 0{,}99$, допустимое время противодействия вредоносной информации $T^{acceptable} = 102$ минуты.

К ресурсопотреблению [142] $P_{\text{res}}\left(r \leq R^{acceptable}\right) \geq P_{res}^{acceptable}$, где $P_{res}^{acceptable}$ =0,99, $R^{acceptable} = 0{,}75$ (75% от общего объема ресурсов).

Следуя установленным требованиям к системе противодействия вредоносной информации в социальных сетях, определим общий путь диссертационного исследования. Необходимо разработать:

1. Комплекс моделей социальной сети, источника и вредоносной информации, содержащий всю необходимую информацию об информационном обмене вредоносной информацией в социальной сети, в том числе об источнике вредоносной информации и об обратной связи со стороны аудитории источника: количество просмотров, количество отметок мне нравится (лайк), количество отметок мне не нравится, количество репостов, количество комментариев, количество подписчиков источника в социальной сети и т.д.

2. Комплекс алгоритмов анализа и ранжирования контрмер, сортирующий объекты воздействия по приоритету и ранжирующий контрмеры для противодействия вредоносной информации в социальных сетях.

3. Методику противодействия вредоносной информации в социальных сетях, использующую авторские алгоритмы анализа источников и ранжирования контрмер, формирующую пары цель-контрмера для принятия решения оператором о противодействии вредоносной информации в социальных сетях.

4. Архитектуру и программные прототипы компонентов системы противодействия вредоносной информации, ориентированную на ранжирование и выбор доступных контрмер в системе для заданных типов вредоносной информации.

Входными для исследования являются следующие данные:

$$DATASET \subseteq \{messages, sources\},$$

где *messages* – множество сообщений, содержащих вредоносную информацию, sources – множество источников этих сообщений.

$$MESSAGE = <messageURL, source, activity, messageType>,$$

где $messageURL$ – адрес сообщения в СС, $source$ – источник сообщения, $messageType$ – тип сообщения (пост, комментарий или ответ на комментарий), $activity$ – характеристики сообщения.

$$SOURCE = <sourceID, sourceURL>,$$

где $sourceID$ – уникальный идентификатор источника, $sourceURL$ – адрес источника в СС.

$$ACTIVITY = <countLike, countRepost, countView, countComment>,$$

где $countLike$ – количество отметок «мне нравится», $countRepost$ – количество «репостов» (копий со ссылкой на источник), $countView$ – количество просмотров, а $countComment$ – количество комментариев.

Требуется найти:

$$DATASET\_MAX \subseteq \{messages\_max, sources\_max\},$$

где $messages\_max$ – множество сообщений $(message)$, у которых характеристики $activity$ будут самыми высокими по сравнению с другими сообщениями в множестве messages, а $sources\_max$ – множество источников $(source)$, которые связаны с максимальным количеством сообщений $(message)$, входящих в множество $messages\_max$.

Сформулируем научную задачу следующим образом: для имеющегося набора входных данных найти следующий кортеж:

$$< Models, Algorithms, Methodology, Architecture >,$$

где $Models$ – комплекс моделей социальной сети, источника и вредоносной информации; $Algorithms$ – комплекс алгоритмов анализа источников и ранжирования контрмер; $Methodology$ – методика противодействия вредоносной информации в социальных сетях, $Architecture$ – архитектура и программные прототипы компонентов системы противодействия вредоносной информации в социальных сетях.

При этом необходимо добиться максимизации показателя обоснованности с учетом требований оперативности и к ресурсопотреблению.

1.5 Выводы по главе 1

1. Противодействие вредоносной информации в социальных сетях является важным элементом информационной безопасности государства, общества и личности, однако большинство систем не учитывает ландшафт функциональности системы противодействия. Существующие системы разделены два разделенных модуля: (1) мониторинг; (2) противодействие. Оператор находится между ними, необходима автоматизация процесса противодействия. Также современные социальные сети имеют сложную структуру, состоят из сообщений и источников, на страницах которых они опубликованы. И параметры сообщений и источников недостаточно учитываются при выборе цели противодействия.

2. Методика противодействия вредоносной информации, разработка которой обосновывается в данной главе, будет учитывать характеристики источника, количество сообщений на его странице, обратную связь от аудитории сообщения и источника. Будет ранжировать контрмеры с учетом коэффициентов сложности, а также поддерживать 2 стадии: 1) настройка; 2) эксплуатация. Для разработки данной методики необходимо разработать модели социальной сети, источника, вредоносной информации, комплекс алгоритмов анализа источников и сортировки объектов воздействия. Внедрение методики противодействия вредоносной информации в социальных сетях позволит: 1) сортировать объекты воздействия по приоритету для оператора; 2) повысить качество принимаемых решений в процессе противодействия вредоносной информации; 3) задать исходные настройки для системы противодействия.

3. Формальная постановка задачи определяет цель разработки методики как максимизацию количества учитываемых параметров для выбираемых объектов воздействия и контрмер в ходе противодействия вредоносной информации в социальных сетях N_{param}, при соблюдении требований к оперативности и ресурсопотреблению.

ГЛАВА 2. Комплексы моделей и алгоритмов анализа источников вредоносной информации и выбора контрмер

Существующие подходы к построению моделей можно условно разделить на 3 концепции: 1) модели представления данных, 2) модели информационного обмена в социальных сетях, 3) модели распространения информации. Каждая концепция по-своему уникальна и позволяет описывать различные характеристики. Так, например, первая лучше всего подходит для разработки модели социальной сети.

2.1. Комплекс моделей социальной сети, источника и вредоносной информации

Модели данных (МД) – это совокупность правил создания и взаимосвязи структур данных социальных сетей в базе данных, возможных операций, а также ограничений (например, количество взаимных друзей в Facebook, VK) [143]. Модель данных всех социальных сетей состоит из трех наборов: (1) набор типов структур данных; (2) операторов и правил вывода; (3) набор общих правил целостности.

В основе структуризации данных СС лежат концепции «агрегации» и «обобщения». При этом наименьшей единицей в модели данных социальной сети (МДСС) является «элемент данных» (сетевая модель данных (версия CODASYL) [144]) или «атрибут» (реляционная модель данных (РМД) [145]). Поименованная совокупность всех элементов данных внутри МДСС, которую можно рассматривать как единое целое называется «агрегат данных» (CODASYL). В МД социальной сети «запись» (CODASYL) или «кортеж» (РМД) может иметь несколько атрибутов. Так, например, запись: пользователь <user> имеет несколько элементов данных (связанный друг, связанный пост (репост), пол) и несколько агрегатов: простые агрегаты – ФИО, адрес и повторяющиеся агрегаты – интересы.

Среди атрибутов МДСС выделяются одно или несколько ключевых полей в качестве основного ключа, именно они характеризуют домены

(классы) МДСС. Если модель данных описывает класс реципиента, тогда основной ключ – это имя или идентификатор. Для МД социальной сети основными ключами будут идентификатор источника и идентификатор сообщения.

Рассмотрим три основных структуры данных социальных сетей.

Первый тип СС – это сети, структура которых представляет собой полносвязный граф (Полносвязанные СС). К таким сетям относятся Facebook, VK, OK и другие схожие (рис. 2.1). Особенностью таких сетей является связь между идентификаторами страниц двунаправленная или однонаправленная (на рисунке – не прерывистые линии), которая позволяет сторонним страницам (будь то страница пользователя или группы) взаимодействовать с сообщениями или страницами, с которыми они не связаны – косвенная связь (на рисунке – пунктирные линии).

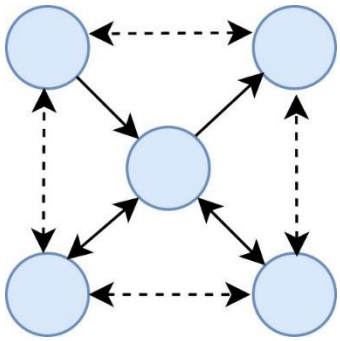


Рисунок 2.1 – Структура данных полносвязных социальных сетей

Второй тип – социальные сети, медиа-трансляторы информации (будь то пользователь или организация) (медиа-трансляторы). По своей сути структура данных таких СС также, как и первого типа содержит связи между идентификаторами страниц, однако медиа-транслятор имеет свою отличительную особенность – это сторонний информационный канал, формируемый системой на основе анализа предпочтений получателей и связанных с ними идентификаторов (рис. 2.2). Но только, если в сетях первого типа вершинами графа могут быть как сообщения, так и страницы пользователей, групп, то в СС второго типа связь устанавливается только

между пользователями или стоящими за их страницами организациями. Одним из ярких примеров таких СС является Instagram. Новая сеть медиа-транслятор – Tik Tok.

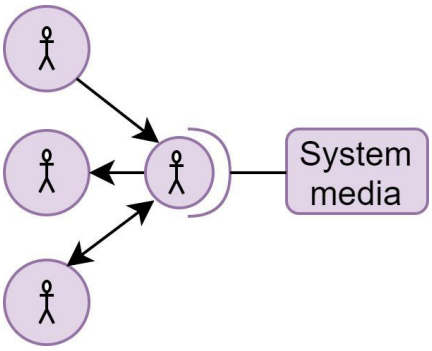


Рисунок 2.2 – Структура данных социальных сетей медиа-трансляторов

К третьему типу сетей относятся такие СС, структура данных у которых однонаправленная от многих источников (сообщений) ко одному реципиенту (рис. 2.3). Отличительной особенностью таких сетей является то, что получатель информации выбирает набор источников, с которыми он связан (подписан на них), система предлагает получателю схожие сообщения и источники по теме. Однако источники не связаны с получателями обратной связью и не видят созданный ими контент (исключения составляют комментарии, но и они находятся на странице источника). Яркими примерами таких СС являются Youtube и Telegram.

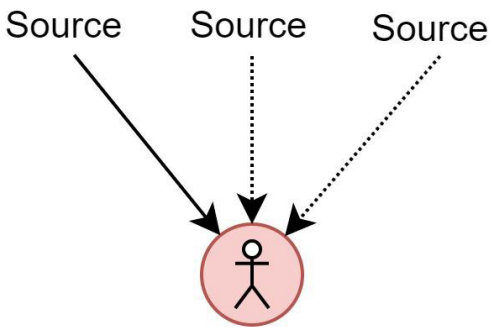


Рисунок 2.3 – Структура данных однонаправленных социальных сетей

Все три типа структур СС содержат общие атрибуты, отличие одной структуры от другой начинается на уровне отношений.

2.1.1 Модель социальной сети

В процессе анализа основных структур данных СС можно выделить общие атрибуты, которые могут быть применимы для описания взаимосвязи между источником, реципиентом и вредоносной информацией.

Пусть в момент присоединения к социальной сети субъект проходит процесс регистрации, создает сетевой профиль – аккаунт (account), который принадлежит множеству $ACCOUNT$. В процессе регистрации аккаунта субъект получает уникальный идентификатор (id), который принадлежит множеству ID. Субъект всегда к нему привязан, хотя находится на границе между виртуальным и реальным миром. Субъект может не добавлять о себе никакой информации после регистрации, однако это не помешает ему создавать и передавать информацию в социальной сети [146].

Далее субъект заполняет аккаунт путем внесения данных о себе в сетевом профиле и тогда он заполняет свою собственную страницу ($page_{ac}$). Страница аккаунта принадлежит множеству $PAGE_{ac}$.

Субъект может создать сообщество, и в этот момент сообщество получает свой уникальный адрес и профиль (group). Сообщества в социальных сетях образуют множество $GROUP$, после заполнения профиля сообщества формируется его страница ($page_g$). Страницы групп принадлежат множеству $PAGE_g$.

При этом, $PAGE = PAGE_{ac} \cup PAGE_g$ [146].

В процессе анализа любого сообщения в СС можно определить страницу $Page$, на которой оно опубликовано. Страница, на которой опубликовано сообщение с вредоносной информацией – это источник (source). Все источники в социальных сетях образуют множество $SOURCE$.

Распространение вредоносной информации в социальных сетях возможно путем публикации сообщения (message). Все сообщения социальной сети образуют множество $MESSAGE$. Сообщение – это любой пост на стене аккаунта, на стене группы, запись в комментариях и др. Тогда, когда

субъект создает сообщение и публикует его в открытом доступе от имени своего аккаунта или от имени группы, он (субъект) является автором (author). Все пишущие субъекты социальных сетей образуют множество *AUTHOR*. По закону РФ [28] ответственность за распространение вредоносной информации лежит не на авторе, а на источнике, в котором такое сообщение опубликовано [146].

Множество источников и сообщений может быть выделено в отдельные домены (классы), совместно образующие модель социальной сети (рис. 2.4).

Определим общие атрибуты, присутствующие в большинстве структур данных (таблица 2.1).

Таблица 2.1. − Основные атрибуты модели данных социальной сети

Элемент структуры данных	Полносвязанные СС	Медиа-трансляторы	Однонаправленые СС
id_source	+	+	+
followers	+	+	+/-
message	+	+	+
url_message	+	+	+
type_message	+	+	+
like	+	+/-	+
comment	+	+	+
repost	+/-	+/-	-
answer	+	+	+
views	+	+	+

Рассмотрим модель социальной сети (рис. 2.4).

В модели социальной сети представлены данные в виде набора отношений, каждое из которых является подмножеством декартова произведения определенных множеств.

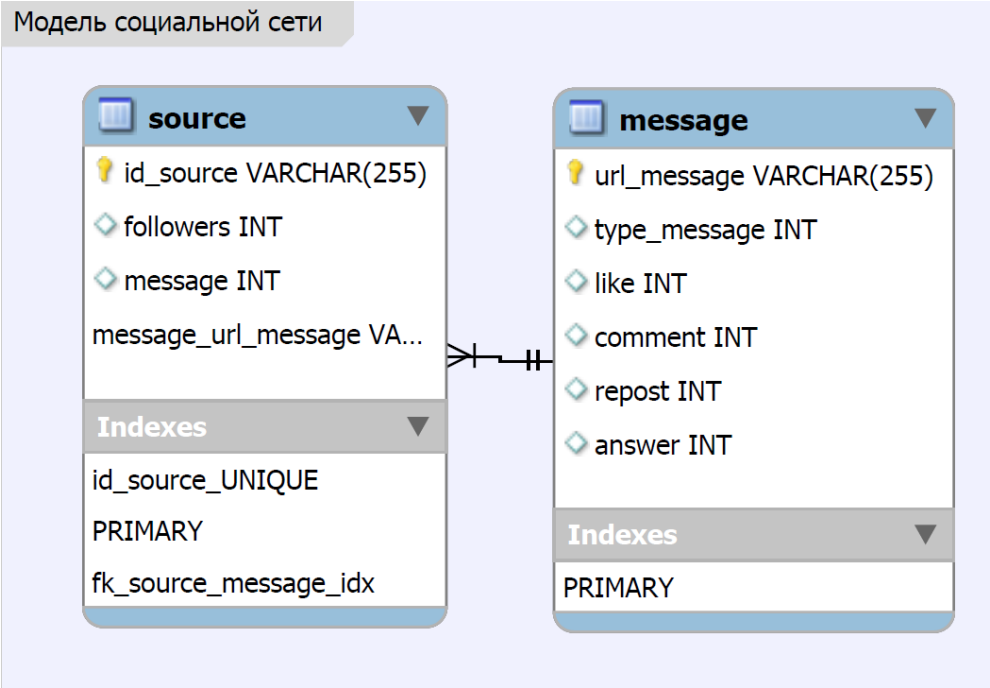


Рисунок 2.4 – Модель социальной сети (МСС)

Элементы отношения в модели данных социальной сети называются «кортежами». Элементы кортежа – атрибуты (поля), к ним относятся:

1. *id_source* - адрес веб_страницы, на которой публикуются сообщения;
2. *followers* - количество подписчиков источника в СС;
3. *message* - количество сообщений;
4. *url_message* - адрес сообщения;
5. *type_message* - тип сообщения (post, comment, answer);
6. *like* - количество отметок «мне нравится»;
7. *comment* - количество комментариев;
8. *repost* - количество репостов;
9. *answer* - количество ответов на комментарии;
10. *views* - количество просмотров.

Модель социальной сети обладает основными свойствами:

1. $f: MESSAGE \rightarrow SOURCE$
2. в СС нет одинаковых кортежей;

3. порядок кортежей не существенен;
4. число кортежей в отношении $R\,(SOURCE, MESSAGE)$– это мощность отношения;

Предложенная модель социальной сети отличается от существующих тем, что содержит новые классы, атрибуты и отношения между ними.

2.1.2 Модель источника информации

В первой главе в пункте 1.2.1 рассматривались релевантные модели информационного обмена, такие как SMCRE [50], математическая модель Шеннона и Уивера [51], A-B-X модель Теодора Ньюкомба [50] и Интегральная (обобщенная) модель Б. Вестли и М. Маклина [50, 52]. Согласно таблице 2.1 для социальных сетей, в независимости от их структуры, общими атрибутами являются источники, сообщения и некоторые признаки обратной связи на сообщения. Именно наличие признаков обратной связи у сообщения позволяет характеризовать источник сообщения.

Пусть

$ACTIVITY\ \{countLike, countRepost, countView, countComment\}$

это множество всех признаков обратной связи у сообщения от реципиентов информации в социальной сети, где где $countLike$ – количество отметок «мне нравится», $countRepost$ – количество «репостов» (копий со ссылкой на источник), $countView$ – количество просмотров, а $countComment$ это количество комментариев.

Исходя из требований определенных в пункте 1.4 диссертации необходимо найти такие атрибуты элементов множества $ACTIVITY$ и отношения $R\,(SOURCE, MESSAGE)$, которые позволят в дальнейшем анализировать источники и сообщения, содержащие вредоносную информацию и выбирать объект для противодействия.

Предположим, что сумма элементов активности к одному сообщению позволяет вычислить индекс активности сообщения, на основе чего может быть получен интегральный показатель индекса активности, зависимый

от количества сообщений источника, следовательно, одним из атрибутов модели источника будет *index_active*.

Пусть сумма просмотров одного сообщения позволяет вычислить индекс просматриваемости сообщения, на основе чего может быть получен интегральный показатель индекса просматриваемости для источника, тогда одним из атрибутов модели источника будет *index_viewability*.

При этом, функция $f: MESSAGE \rightarrow SOURCE$ позволяет задать область определения, входные значения (аргументы) и выходные значения. Функция сюръективна, то есть имеется отображение множества $MESSAGE$ на множество $SOURCE$, при котором каждый элемент множества $SOURCE$ является образом хотя бы одного элемента $MESSAGE$, таким образом $\forall source \in SOURCE\ \exists\, message\ MESSAGE : source = f(message)$.

Однако сообщения на стене источника (аргументы) могут быть разного типа (пост, комментарий, ответ). Следовательно, для отдельных сообщений (аргументов) может быть задан рейтинг (числовой коэффициент) в дереве сообщений (рис. 2.5).

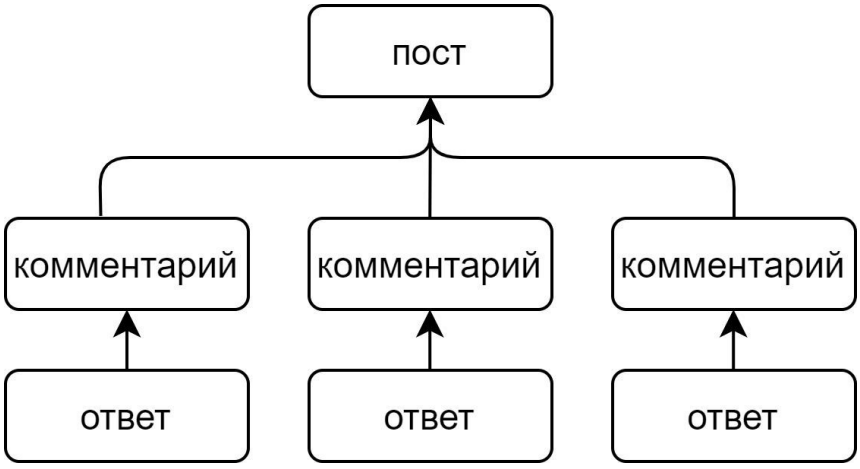


Рисунок 2.5 – Дерево сообщений социальной сети

В зависимости от суммы аргументов источник может быть оценен по потенциалу *Potential*:

1) источник с высоким потенциалом;
2) источник со средним потенциалом;

3) источник с низким потенциалом.

Если источник имеет такие атрибуты как *index_active*, *index_viewability*, может быть задан индекс влиятельности – *index_impact*, отражающий уровень влиятельности источника информации на его аудиторию.

Рассмотрим модель источника (рис. 2.6)

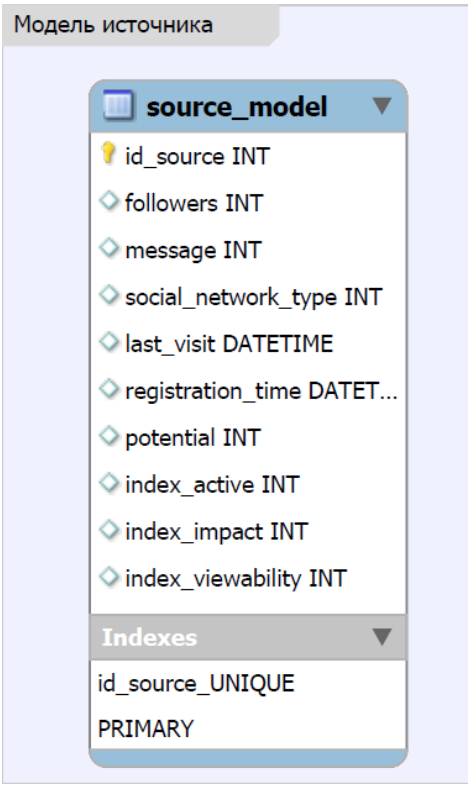


Рисунок 2.6 – Модель источника информации

Выделим кортеж атрибутов характеризующий $SOURCE$ через элементы множества $ACTIVITY$ и отношение $R\,(SOURCE, MESSAGE)$ – $< \mathrm{index}_{\mathrm{active}}, \mathrm{index}_{\mathrm{viewability}}, \ \mathrm{potential}, \mathrm{index}_{\mathrm{impact}} >$. Другими атрибутами модели источника информации являются: *followers* – количество связанных пользователей; *social_network_type* – тип структуры данных социальных сетей; *registration_time* – время регистрации источника и др.

Модель источника информации отличается от существующих наличием новых классов, атрибутов и отношений.

2.1.3 Модель вредоносной информации

Основой для формирования понятия вредоносная информация являются два термина [1]:

1) Информация (I – information) [20].
2) Информационный объект (IO – information object) – это логически цельный блок информации, представленный в определенной фиксированной форме, который создан и используется в ходе информационной составляющей деятельности человека [25].

Формально оба этих термина связаны между собой, таким образом, что $IO \subseteq I$ (рис. 2.7 а) т.е. информационный объект является элементом множества всей анализируемой информации.

При этом термин «информация» также связан с термином «информационное пространство» (IA – information area) [1, 22,23], а множества I и IO являются собственными подмножествами информационного пространства.

Социальные сети (SN) – совокупность взаимосвязанных узлов, которыми являются аккаунты, сообщества, страницы, посты, вложения и т.д., а связи между объектами – это одноуровневые отношения (состоят «в друзьях», состоят в сообществе, и т.п.) и отношения вложенности (стена содержит пост, страница аккаунта содержит ссылку на пост и т.п.). Социальные сети ассоциируются с графами, где часть объектов информационные, и представляют собой вершину графа, а часть – связи между такими объектами – представляют собой ребра между вершинами [1].

Справедливо, что $IO \subseteq I \subseteq IA$, $SN \subseteq I \subseteq IA$, и именно область пересечения между SN и IO и является предметом исследования на пути разработки модели вредоносной информации в социальных сетях (рис. 2.7, б).

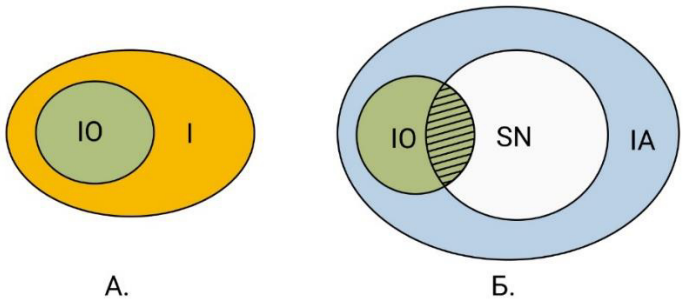


Рисунок 2.7 – Графическое представление взаимоотношений множеств информационного пространства

Пусть MIO – это информационный объект (вредоносный информационный объект), который содержит признаки, позволяющие принять решение о том, что информация наносит вред обществу, личности, государству или бизнесу.

При этом, признаки $Token$ информационной угрозы T устанавливает эксперт, в зависимости от условий. Например, когда используется система родительского контроля, родитель сам выбирает ограничения для ребенка. Если же представитель бизнеса заинтересован в защите конфиденциальной информации, тогда также он сам задает их.

Следовательно, теоретико-множественная модель вредоносной информации в социальной сети, включает такие базовые элементы как:

1) IO – информационный объект (от англ. information object),

2) T – информационная угроза (от англ. threat),

3) MIO – вредоносный информационный объект,

4) $Token$ – признак информационной угрозы, содержащийся во вредоносном информационном объекте (от англ. token),

5) $Feature$ – признак наличия информационного объекта [1,0] (от англ. feature)

6) связи между объектами.

Теоретико-множественная модель формально представлена следующим образом:

$$IO = \{io\}; MIO = \{io\}; MIO_i = \{io\}$$
$$MIO \subset IO; \forall\ io \in MIO: io \in IO$$
$$MIO_i \subseteq MIO; \forall\ io \in MIO_i: io \in MIO$$
$$Token_{mio_i} \subset T;\ Token_{mio_i} = \{t\}$$
$$CheckFeature(io, t) = \{True; False\}$$
$$io\ \in MIO_i\ \Leftrightarrow\ \exists\ Token_{mio_i}: checkFeature\ (io,\ t) = True.$$

где IO – множество информационных объектов, io_1–один информационный объект, T –множество всех возможных признаков информационной угрозы, t_n –один признак информационной угрозы, MIO – множество вредоносных информационных объектов, MIO_i– отдельный класс вредоносной информации, $Token_{mio_i}$– множество признаков характеризующих MIO .

Таким образом, для противодействия вредоносной информации необходимо задать набор признаков, характерных для информационной угрозы. Отличительной особенностью является то, что модель допускает наличие дискретных признаков в наборе признаков, таких как: дата создания информационного объекта, связь информационного объекта с другими объектами в социальной сети, частота повторения признака и др.

2.1.4 Информационно-признаковая модель вредоносной информации

Противодействие вредоносному информационному объекту может осуществляться на уровне сообщений или на уровне источников. Необходимо выделить такие угрозы и их информационные признаки сообщения в социальной сети, характеризующие его как вредоносное.

В общем виде под информационно-признаковой моделью [1] (таблица 2.2.) в исследовании понимается упорядоченная совокупность сведений о связях сообщений и их информационных признаков с содержанием

сообщений. В свою очередь под информационными признаками сообщений понимается отдельные свойства сообщений, а точнее их содержание.

Таблица 2.2. − Информационно признаковая модель вредоносной информации

Информационные угрозы (1)	**Вредоносная информация в социальных сетях (2)**	**Информационные признаки (3)**
Самоубийство *(Пример)*	Сообщение, содержащее предложение, просьбу и/или приказ совершить самоубийство, и/или описывающее самоубийство как способ решения проблем *(Пример)*	t_1
	Сообщение, содержащее положительную оценку либо одобрение совершения самоубийства и/или действий, направленных на самоубийство *(Пример)*	t_2

(1) информационная угроза – задается оператором системы;

(2) вредоносная информация в СС – задается оператором путем формирования набора ключевых слов;

(3) информационные признаки, формирующие множество всех возможных признаков t.

На рисунке (рис. 2.8) изображено соотношение различных уровней информационно-признаковой модели вредоносной информации. Показано, что автор формирует сообщение и размещает его в источнике распространения, на странице либо аккаунта, либо группы. Сообщения могут содержать или не содержать признаки вредоносной информации. Признаки (таблица 2.2) формируют уровень информационных угроз. Таким образом, собрав информацию на странице какого-либо источника возможно определить, какие из этих сообщений относятся к вредоносным. С учетом выявленных угроз и их количества может быть принято решение о противодействии сообщению или источнику.

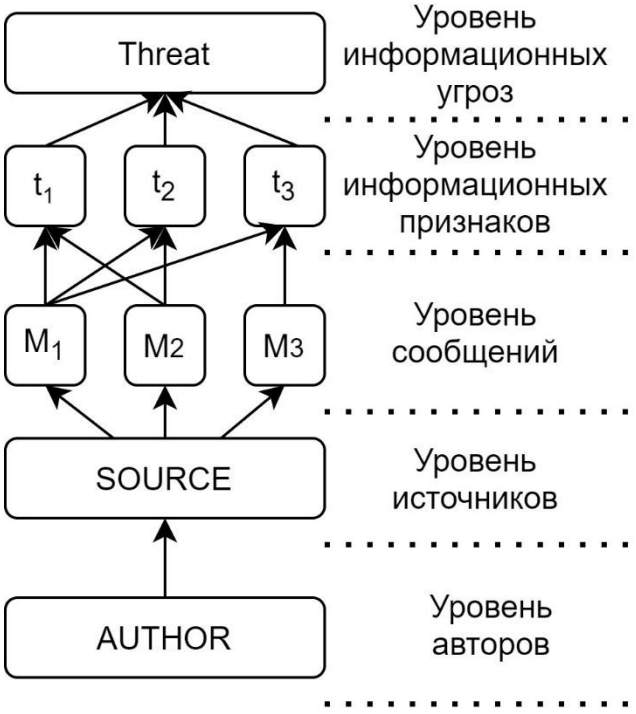


Рисунок 2.8 – Графическое представление информационно-признаковой модели вредоносной информации

Разработанная информационно-признаковая модель вредоносной информации позволяет сформировать исходные данные для противодействия вредоносной информации.

Разработанный комплекс моделей состоит из модели социальной сети, модели источника информации, модели вредоносной информации и информационно-признаковой модели вредоносной информации. Каждая из моделей, входящих в комплекс, содержит уникальные, предложенные автором комплекса множества, атрибуты и отношения между элементами. Одновременно с этим комплекс моделей позволяет сформировать требования к алгоритмам анализа и оценки источников и выбора контрмер.

2.2 Комплекс алгоритмов анализа источников и ранжирования контрмер

С 01 февраля 2021 года вступила в силу поправка в 149 ФЗ и появилась статья 10.6 «Особенности распространения информации в социальных сетях» [28], согласно которой законодатель обязывает собственников

социальных сетей противодействовать вредоносной информации. Однако же, существующие меры и требования не учитывают источник, в котором опубликовано сообщение и аудиторию, противодействие направлено на сообщение. В некоторых случаях блокировка самих источников оказывается более эффективной, чем ограничение доступа пользователей социальной сети к сообщению. Необходимо разработать такие алгоритмы, которые в отличие от существующих будут учитывать атрибуты и характеристики источника и сообщения.

В данном разделе диссертации представлены результаты разработки диаграмм и алгоритмов, позволяющих оценивать источник и ранжировать объекты воздействия и контрмеры для реализации мер противодействия вредоносной информации в социальных сетях. Предполагается, что это позволит повысить эффективность противодействия за счет применения целевых мер противодействия.

2.2.1 Алгоритм ранжирования источников по потенциалу

В работах [59, 61, 78, 137, 147-150] рассматриваются алгоритмы выявления лидеров мнений, каналов распространения, источников слухов и т.д. Однако большинство существующих алгоритмов использует в своей основе контент анализ. Разработанный алгоритм ранжирования источников опирается на зависимость от количества сообщений.

Предположим, что совокупность сообщений во множестве *DATASET* может быть разделена по множеству *SOURCES*, которым принадлежат разные количества сообщений из множества *MESSAGES*. Тогда список сообщений от отдельно взятого источника может быть представлена в виде узла фрейма иерархическим графом [1] (рис. 2.9).

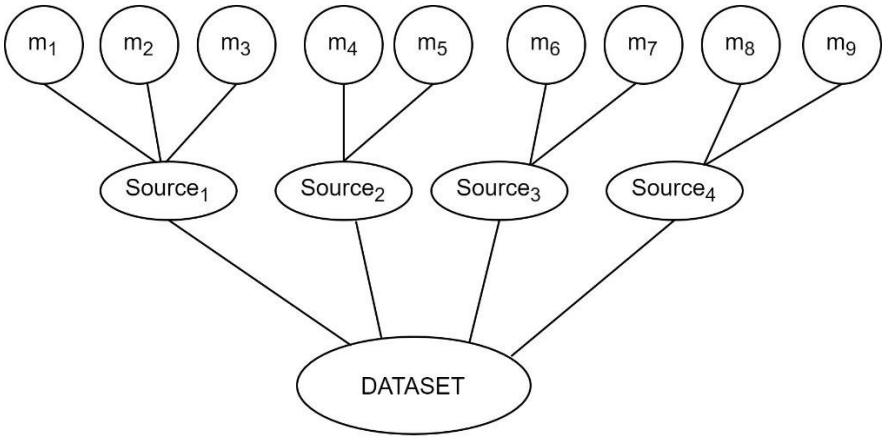


Рисунок 2.9 – Иерархический граф ранжирования источников

Каждое сообщение располагается на некотором уровне глубины дерева сообщений на стене источника. Если это пост – оно является корнем дерева. Если это ответ на пост, то сообщение располагается на втором уровне дерева, Ответ на ответ занимает третий уровень. Следовательно, каждому сообщению может быть присвоен числовой коэффициент (таблица 2.3).

Таблица 2.3. − Числовые коэффициенты сообщений на стене источника

№	Тип сообщения	Числовой коэффициент
1	Пост	1
2	Комментарий	0,5
3	Ответ на комментарий	0,25

В зависимости от суммы сообщений на стене, источники могут быть разделены потенциалам, следующим образом:

1. Потенциал источника является низким P_{LI}, тогда, когда он соответствует неравенству (2.1):

$$f_1(S_p) \le \overline{X_1} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad (2.1)$$

где $\sum_{i=1}^{n} x_i$ – сумма числовых коэффициентов всех сообщений на стене источника, n – количество сообщений, принадлежащих источнику, а $\overline{X_1}$ – среднеарифметическое в наборе данных для всех источников *DATASET*.

2. Потенциал источника является средним P_{MI}, тогда, когда соблюдается неравенство (2.2):

$$f_2(S_p) \leq \overline{X_2} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad (2.2)$$

где $\sum_{i=1}^{n} x_i$ – сумма числовых коэффициентов всех сообщений на стене источника, n – количество сообщений, принадлежащих источнику, а $\overline{X_2}$ – среднеарифметическое в наборе данных, полученное после отделения источников с низким потенциалом P_{LI}.от исходного набора данных *DATASET*.

3. Потенциал источника является высоким P_{HI}, если соблюдается неравенство (2.3):

$$f_3(S_p) > \overline{X_2} = \frac{\sum_{i=1}^{n} x_i}{n}. \quad (2.3)$$

Таким образом, все источники в наборе данных в зависимости от количества и глубины сообщений на стене источника могут быть ранжированы по потенциалу (таблица 2.4).

Таблица 2.4. − Классификация источников по потенциалу

Значение потенциала	Потенциал	Описание
0	P_{LI}	Низкий (low index) потенциал источника
1	P_{MI}	Средний (medium index) потенциал источника
2	P_{HI}	Высокий (high index) потенциал источника

Рассмотрим блок схему алгоритма ранжирования источников по потенциалу (рис. 2.10).

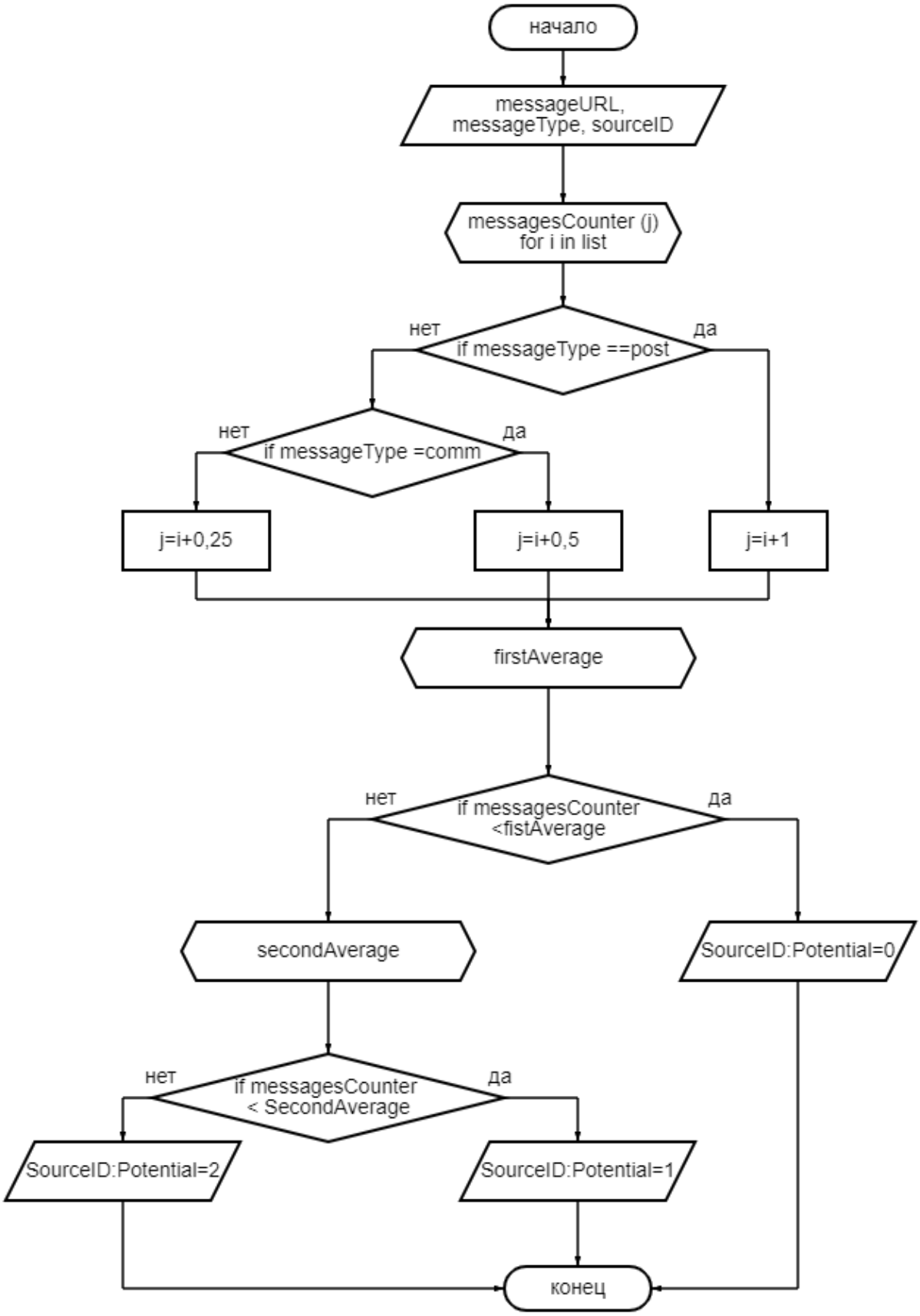


Рисунок 2.10 – Блок-схема алгоритма ранжирования источников

На вход в алгоритм ранжирования источников по потенциалу подается набор кортежей *<messageURL, messageType, sourceID>*. Далее происходит обработка данных по шагам:

Шаг 1. Присваивание каждому сообщению в наборе числового коэффициента в зависимости от атрибута *messageType*. И суммирование числовых коэффициентов всех сообщений для каждого источника. На выходе формируется кортеж *<sourceID, message_Count>.*

Шаг 2. Расчет первого среднеарифметического по количеству сообщений, принадлежащих источникам. Для источников со значением *message_Count* меньшим, чем первое среднеарифметическое присваивается показатель низкого потенциала равный 0. Источники с низким потенциалом отделяются, формируется новый кортеж *<sourceID, message_Count>.*

Шаг 3. Расчет второго среднеарифметического по количеству сообщений источников. Для источников со значением *message_Count* меньшим или равным второму среднеарифметическому присваивается показатель потенциала равный 1. Для источников со значением *message_Count* большем, чем второе среднеарифметическое – показатель потенциала равен 2.

На выходе из алгоритма ранжирования источников по потенциалу формируется кортеж *<sourceID, potentialIndex>.*

Алгоритм ранжирования источников по потенциалу в отличие от существующих учитывает количество опубликованных сообщений, глубину их расположения на странице в социальной сети при ранжировании источников.

2.2.2 Алгоритм оценки источников

Противодействие вредоносной информации может осуществляться на основе комплексного подхода к анализу источников, сообщений. Обратим внимание на тот факт, что объектом деструктивного воздействия вредоносной информации являются пользователи социальной сети [74, 82, 137, 151, 152].

При этом каждый пользователь оставляет след во время просмотра сообщения, и может выражать свою реакцию на него. Алгоритм оценки источников учитывает обратную связь от аудитории вредоносной информации на странице источника в процессе информационного обмена.

Пусть множество

$ACTIVITY$ $\{countLike, countRepost, countView, countComment\}$ – включает в себя все признаки обратной связи от аудитории вредоносной информации в социальной сети, при этом $countLike$ – количество отметок «мне нравится», $countRepost$ – количество «репостов» (копий со ссылкой на источник), $countView$ – количество просмотров, а $countComment$ – количество комментариев. Все элементы множества имеют целочисленный тип данных (integer). Множество таких чисел (integer) является конечном подмножеством, состоящим из неотрицательных чисел.

Во множество $SOURCE$ $\{\text{sourceID}, \text{messageURL}\}$ входят идентификатор источника и адрес сообщений в социальной сети.

В соответствии с требованиями, определенными в пункте 1.4 Главы 1 необходимо найти кортеж атрибутов характеризующий $SOURCE$ через элементы множества $ACTIVITY$ и отношение $R\,(SOURCE, MESSAGE)$ – $< \text{index}_{\text{active}}, \text{index}_{\text{viewability}}, \text{index}_{\text{impact}} >$,

где $\text{index}_{\text{active}}$ – индекс активности, который может быть задан через целевую функцию (2.4):

$$f(S_{act}) \to I_{act}^{s}[0,1,2], \left(I_{act} = \frac{I_{act}}{\max I + 1}\right). \qquad (2.4)$$

$\text{index}_{\text{viewability}}$ – индекс просматриваемости может быть задан функцией (2.5):

$$f(S_{view}) \to I_{view}^{s}[0,1,2], \left(I_{view} = \frac{I_{view}}{\max I + 1}\right). \qquad (2.5)$$

$\text{index}_{\text{impact}}$ – индекс влиятельности источника, который может быть задан целевой функцией (2.6):

$$f\left(S_{impact}\right) \to I_{impact}^{s}[0,1,2], \left(I_{impact} = \frac{I_{impact}}{\max I + 1}\right). \qquad (2.6)$$

Значение индексов активности, просматриваемости и влиятельности источника находится между 0 и 2, при этом к значениям индексов применяется нормировка, такая что ко всему множеству применяется общее начало отсчета равное 0, преобразование монотонное и сохраняет отношение предпочтения на множестве допустимых решений. Метод нормировки – сравнительная нормализация, при которой за идеальное значение выбирается максимум. Применение этого метода обосновано тем, что все значения для всех индексов являются неотрицательными величинами.

Рассмотрим блок схему алгоритма оценки источников (рис. 2.11).

На вход алгоритм оценки источников подается кортеж *<messageURL, sourceID, likesCount, commentCount, repostCount, viewCount>*, обработка данных происходит в несколько шагов.

Шаг 1. Вычисление индекса активности источников:

1.1 Формируются хеш-таблицы (key-value), в первой key – это *sourceID,* value –*urlCOUNTER*. Значения элементов во второй *likesCount, commentCount, repostCount* связаны в свою очередь с ключом *messageURL.*

1.2 Во второй хеш-таблице суммируются показатели *likesCount, commentCount, repostCount* для *messageURL,* формируется кортеж *<message.SourceID, activityIndex>*.

1.3 Значения из кортежа *<message.SourceID, activityIndex>* суммируются и затем делятся на показатель *urlCOUNTER* из первой хеш-таблицы. Формируется набор индексов активности источника, к которым применяется сравнительная нормировка.

Шаг 2. Вычисление индекса просматриваемости источников:

Формируются хеш-таблицы (key-value), *{'sourceID': urlCOUNTER, 'messageURL': viewCount}*. Все значения *viewCount* для всех *messageURL* суммируются и затем сумма делится на *urlCOUNTER.* На выходе формируется кортеж *<SourceID, viewIndex>*. К индексам просматриваемости применяется нормировка.

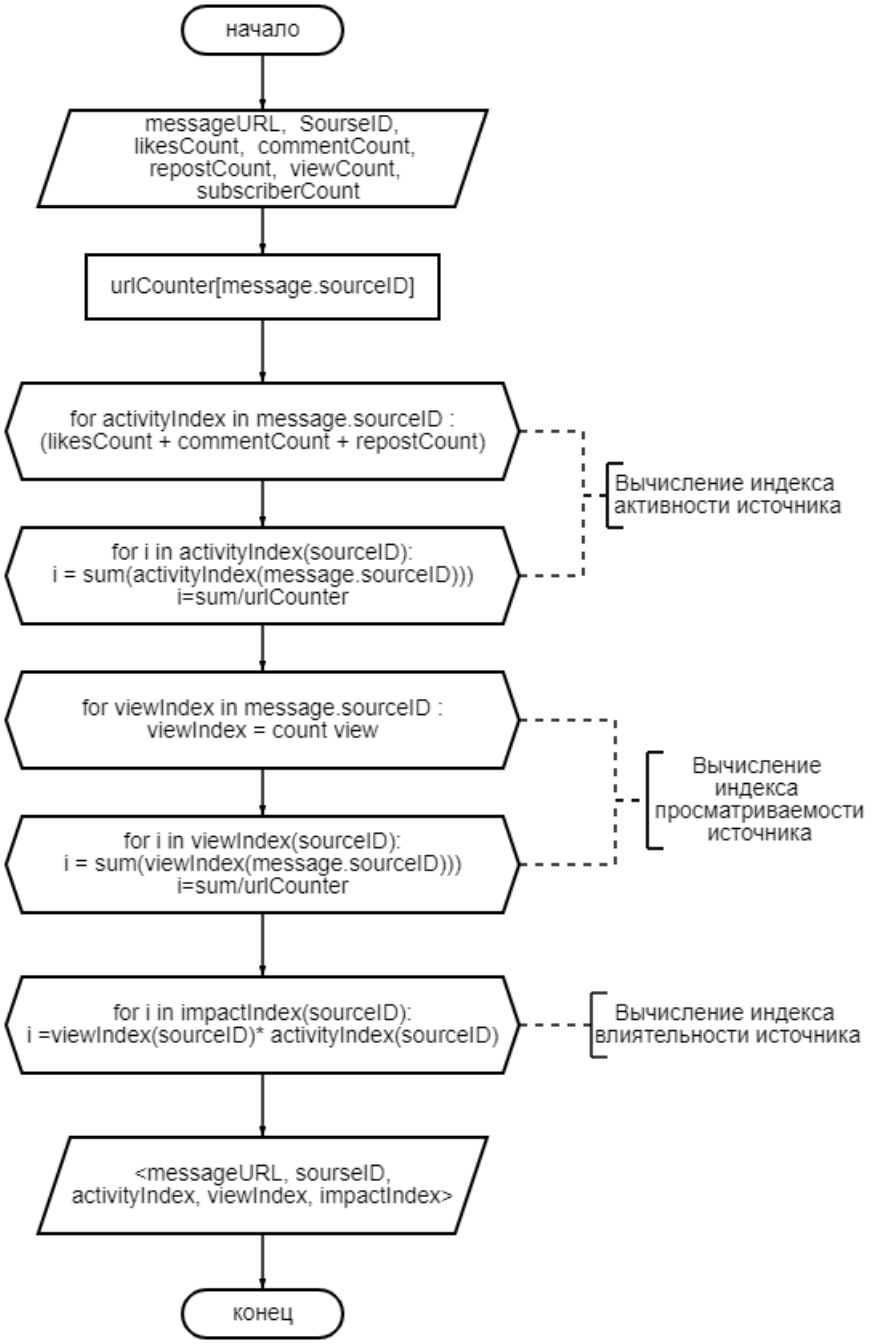


Рисунок 2.11 – Блок-схема алгоритма оценки источников

Шаг 3. Вычисление индекса влиятельности источника:

Для каждого источника перемножаются индекс активности и индекс просматриваемости, получается значение индекса влиятельности, к которому применяется сравнительная нормировка.

На выходе из алгоритма формируется кортеж *<sourceID, activityIndex, viewIndex, impactIndex>*.

Алгоритм оценки источников в отличие от аналогов учитывает количественные характеристики обратной связи от аудитории вредоносной информации в процессе информационного обмена и преобразует их в качественные (индексы).

2.2.3 Алгоритм сортировки объектов воздействия

В основе существующих методов и решений противодействия вредоносной информации [120, 122, 126, 132-134, 138, 139, 153] лежат алгоритмы выявления множества информационных объектов с вредоносной информацией. Однако все эти методы и решения опираются на концепцию: «обнаружение-противодействие». А выбор объекта противодействия осуществляется по принципу FIFO («первым пришёл – первым ушёл»). При этом информационных объектов, содержащих вредоносную информацию, источников, где эта информация опубликована в социальных сетях, миллионы. И все информационные объекты условно могут быть разделены между собой по потенциалу и индексам активности источника, следовательно, возможно применить фильтр в процессе выбора объекта противодействия и задать приоритетную очередь.

Алгоритм сортировки объектов воздействия связан с алгоритмами ранжирования по потенциалу и оценки источников таким образом, что на вход получает выходные данные из них и сортирует объекты воздействия по приоритету на выходе.

Формально целевая функция приоритезации объектов воздействия алгоритма сортировки объектов воздействия может быть задана формулой (2.7):

$$f(S) \to I_{pr}^{s} = I_{p}^{s} + I_{i}^{s} = [0,4], \qquad (2.7)$$

где S – источник, I_{pr}^{s} – приоритет источника, I_{p}^{s} – потенциал а I_{i}^{s}– индекс влиятельности.

При этом правила для выбора объекта воздействия *Target*, следующие:

1) $\{source \in TARGET | I_{pr}^{s} \cong max\}$,

2) $\{message \in TARGET | I_{pr}^{s} \cong min\}$,

где $TARGET$ – это множество объектов воздействия.

Рассмотрим блок-схему алгоритма сортировки объектов воздействия (рис. 2.12)

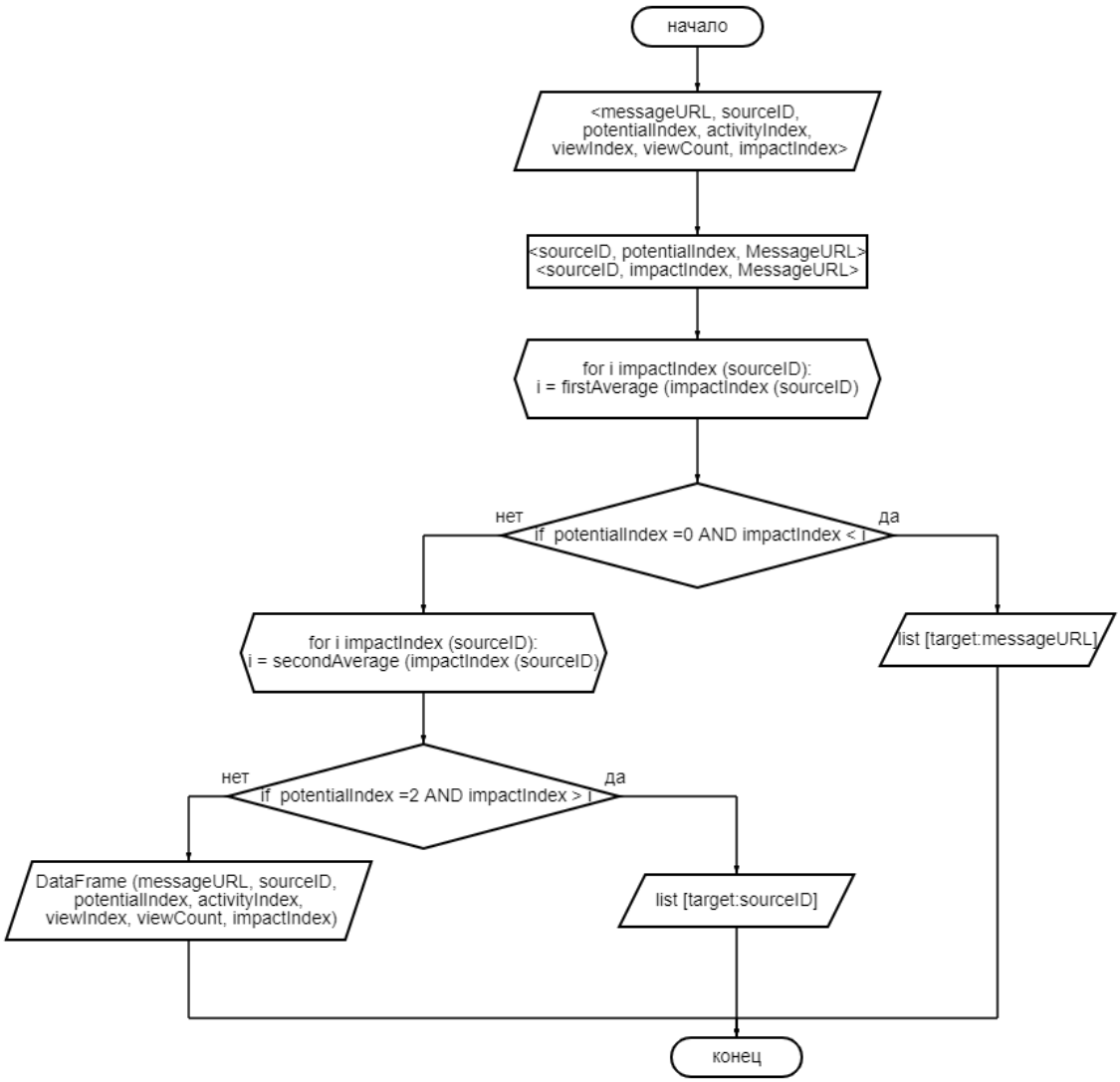


Рисунок 2.12 – Блок-схема алгоритма сортировки объектов воздействия

На вход в алгоритм сортировки объектов воздействия передается набор кортежей *<messageURL, sourceID, potentialIndex, activityIndex, viewIndex, impactIndex>*.

В основе алгоритма лежит сортировка с бинарным поиском, для чего на первом шаге вычисляется среднее арифметическое значение индекса влиятельности всех источников в массиве. Далее выделяются объекты

с высоким и низким приоритетом. Отдельно формируется набор кортежей *<messageURL, sourceID, potentialIndex, activityIndex, viewIndex, impactIndex>* с индексом приоритета $1 \leq I_{pr}^{s} \leq 3$.

На выходе формируются два списка и набор кортежей: (1) список Priority_High – цели *Target*, где объектом воздействия является *sourceID*, имеющие высокий приоритет для принятия мер противодействия; (2) список Priority_Low – цели *Target*, где объектом воздействия является *messageURL*, имеющие низкий приоритет для принятия мер противодействия; (3) Priority_Medium набор кортежей, который передается оператору для дополнительной оценки и выбора объекта воздействия между адресом сообщения и адресом страницы в социальной сети, на которой оно опубликовано.

Таким образом алгоритм сортировки объектов воздействия формирует приоритетные списки для противодействия.

2.2.4 Алгоритм ранжирования контрмер

Для ранжирования мер противодействия используется оценка сложности, основанная на экспертной оценке доступных ресурсов и особенностей мер противодействия. В результате, на основе сведений о доступных мерах противодействия и возможных объектов воздействия формируется список возможных целей мер противодействия, ранжированный на основе метрики, учитывающей как сложность самой меры противодействия, так и свойства объекта воздействия [3].

В процессе разработки алгоритма ранжирования контрмер используется диаграмма противодействия (рис. 2.13), такая что:

Класс контрмер является обобщающим классом для всех классов мер обеспечения информационной безопасности, таких как:

1) организационные меры;
2) технические меры;
3) управленческие меры.

Одновременно с этим класс контрмеры связан с классами:

1) цель воздействия;
2) тип воздействия;
3) метод воздействия

Класс контрмеры связан с ними таким образом, что изменение в атрибутах или в операциях в любом из этих классов приводит к необходимости внесения изменений в класс контрмеры.

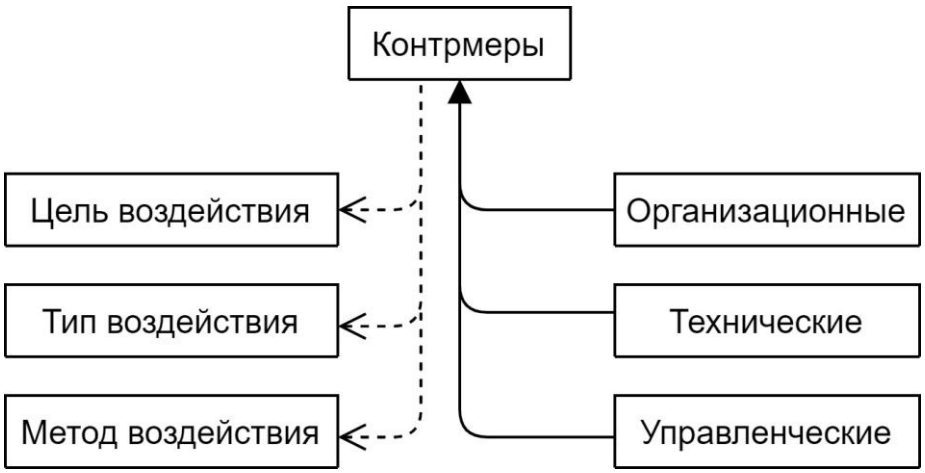


Рисунок 2.13 – Диаграмма классов комплекса алгоритмов выбора мер противодействия

Анализ возможных мер противодействия показал, что при выборе контрмеры должны учитываться такие параметры как сложность реализации мер противодействия (например, автоматическая мера противодействия в общем случае значительно проще в реализации, чем ручная), доступные ресурсы (например, наличие или отсутствие возможности подачи жалобы в суд).

Алгоритм ранжирования контрмер состоит из нескольких состояний объектов, входящий в суперкласс {контрмеры}, цель алгоритма выбрать базовые настройки контрмер (рис. 2.14).

Рисунок 2.14 – Диаграмма алгоритма ранжирования контрмер

Шаг.1. «Формирование списков». Происходит формирование списка контрмер. Список задается оператором, он может быть создан с нуля, может быть выбран из списка возможных вариантов, представленных в некоторой базе данных.

Шаг 2. Классификация контрмер

Выделим ряд свойств, по которым осуществляется классификация контрмер, и зададим классы и их атрибуты, связанные с классом контрмер [3]:

1. тип воздействия:

а. *позитивный (замещение)* (способ противодействия предполагает наполнение информационного пространства социальных сетей положительной информацией, которая не связана с обнаруженной вредоносной информацией);

б. *негативный (блокировка)* (способ противодействия предполагает снижение объема вредоносной информации в информационном пространстве

социальных сетей, а значит направлен на блокировку источников и/или сообщений);

в. *нейтральный (шум)* (способ противодействия предполагает снижение популярности вредоносной информации в информационном пространстве СС посредством распространения альтернативной точки зрения или размытия внимания).

2. метод воздействия:

а. *автоматический метод воздействия* (могут быть разработаны специальные сценарии, применяемые системой противодействия без участия оператора);

б. *автоматизированный метод воздействия* (требуется подтверждение от оператора системы);

в. *ручной метод воздействия* (требуется привлечение оператора системы для разработки уникального сценария).

Шаг 3. Формирование набора комбинаций свойств контрмер

Пусть для контрмер, направленных на вредоносную информацию и/или источник, определен ряд классов свойств контрмер [3]:

$$KC_i \in KC,$$

где KC – множество классов свойств контрмер, а KC_i – конкретный класс свойств контрмер. Для каждого конкретного класса свойств контрмер задан вес w_i, который определяет вклад в сложность меры от данного класса. Также, каждый конкретный класс свойств контрмер содержит набор экземпляров, которые определяют значение данного свойства контрмеры:

$$kc_{i,j} \in KC_i,$$

такое что KC_i – конкретный класс свойств контрмеры, а $kc_{i,j}$ – экземпляры конкретного класса свойств мер противодействия. Для каждого экземпляра конкретного класса свойств контрмер задан уровень сложности $lc_{i,j}$, который определяет вклад в сложность меры противодействия от данного экземпляра класса.

При этом любая из контрмер определяет свою начальную сложность как cw_x.

Кроме собственно своего содержания контрмера определяется как выбор одного из доступных экземпляров для каждого конкретного класса свойств мер. Таким образом, для каждой контрмеры $c_x \in C$, значение свойства $cp_{x,i,j}$, определенного конкретным классом свойств мер KC_i – может принимать значение $\{1,0\}$, в зависимости от того, относится ли данная контрмера к экземпляру конкретного класса свойств мер противодействия $kc_{i,j}$. При этом выполняется следующее утверждение (2.8) [3]:

$$\forall c_x, \forall KC_i, \sum_{j=1}^{|KC_i|} cp_{x,i,j} = 1, \qquad (2.8)$$

так как каждая конкретная мера может иметь только одно значение, определяемое любым из конкретным классом свойств контрмер.

Тогда все множество контрмер представляет собой набор комбинаций свойств мер противодействия. Полный список контрмер может быть представлен в форме таблицы (см. таблица 2.5).

Таблица 2.5. − Формальное представление списка контрмер

C		***KC***								
		KC₁			...			KC_p		
		w₁			...			$w_{\|KC\|}$		
		$kc_{1,1}$	.. .	$kc_{1,\|KC1\|}$	...	...	...	$kc_{\|KC\|,1}$	.. .	$kc_{\|KC\|,\|KCp\|}$
		lc₁,₁	.. .	$lc_{1,\|KC1\|}$	...	...	...	$lc_{\|KC\|,1}$	.. .	$lc_{\|KC\|,\|KCp\|}$
c_1	***cw₁***	$cp_{1,1,1}$	.. .	$cp_{1,1,\|KC1\|}$	...	...	...	$cp_{1,\|KC\|,1}$	.. .	$cp_{1,\|KC\|,\|KCp\|}$
...	...	...	.. .	...	...	...	...	...	.. .	...
cn	cw_n	$cp_{n,1,1}$	.. .	$cp_{n,1,\|KC_1\|}$	...	...	...	$cp_{n,\|KC\|,1}$	.. .	$cp_{n,\|KC\|,\|KC_p\|}$

Шаг 4. Оценка сложности реализации конкретной контрмеры

Метрика сложности для реализации контрмеры $c_x \in C$, может быть задана в виде следующей функции (2.9):

$$complexity\ (c_x) = cw_x * \sum_{i=1}^{|KC|} \quad w_i * \left(\sum_{j=1}^{|KC_i|} \quad \left(cp_{x,i,j} * \ lc_{x,i,j} \right) \right). \quad (2.9)$$

Результатом выполнения данного алгоритма является набор кортежей с оценками сложности для каждой контрмеры $<c_x, complexity(c_x)>$. В результате каждая контрмера c_x имеет свою метрику сложности и может быть ранжирована в системе по этой метрике.

2.3 Формальное представление комплекса алгоритмов анализа источников и ранжирования контрмер

Представим в формальном виде атрибуты комплекса анализа источников и ранжирования контрмер с целью выбора объекта воздействия в рамках противодействия вредоносной информации в социальных сетях:

1. $Sources =< messageURL, messageType, sourceID >.$

В кортеже $Sources$ имеются следующие поля: $messageURL, messageType, sourceID$.

2. $Messages =< messageURL,\ soutceID, likesCount,$

$commentCount, repostCount, viewCount, subscriberCount >$.

Кортеж $Messages$ содержит следующие поля: $messageURL$ – адрес сообщения в сети Интернет и в социальной сети, $sourceID$ – идентификатор источника, где опубликовано сообщение, $likesCount$– количество отметок мне нравится, $commentCount$ – количество комментариев, $repostCount$ – количество репостов, $viewCount$– количество просмотров, $subscriberCount$– количество подписчиков.

3. $Priorities =< messageURL,$

$sourceID, potentialIndex, viewIndex, impactIndex >.$

В кортеже $Priorities$ имеются следующие поля: $messageURL, sourceID,$ $potentialIndex$ – потенциал источника, $activityIndex$ – индекс активности

источника, $viewIndex$ – индекс просматриваемости, $impactIndex$ – индекс влиятельности.

$$4.\ Countermeasures = < type_{action}, method_{action},$$
$$level_{complexity}, class_{weight}, applicability_{factor}, c_w, complexity\ >$$

Кортеж $Countermeasures$ включает следующие поля: $type_{action}$ – тип воздействия, $method_{action}$– метод воздействия, $level_{complexity}$– уровень сложности реализации контрмеры, $class_{weight}$ – вес класса контрмер, $applicability_{factor}$ – применимость контрмеры, $complexity$ – сложность реализации конкретной меры, c_w – начальная сложность контрмеры.

Представим задачу анализа источников и выбора контрмер в виде (2.10):

$$Z = SC \to max,$$
$$\begin{cases} f_1(S) \to I_p^s = \{0,1,2\}, \\ f_2(S) \to I_i^s[0,1,2], \left(I_i = \frac{I_i}{\max I + 1}\right), \\ f_3(S) \to I_{pr}^s = I_p^s + I_i^s = [0,4], \\ f(C) \to complexity(c_x) = \frac{cw_x * \sum_{i=1}^{|KC|} w_i * \left(\sum_{j=1}^{|KC_i|} (cp_{x,i,j} * lc_{x,i,j})\right)}{100 * |KC|}, f(c) \to (0;1] \end{cases}, \quad (2.10)$$

где:

1. $f_1(S)$ – индекс потенциала источника равный 0,1,2 в зависимости от количества сообщений в анализируемом наборе данных, принадлежащих источнику. Вычисляется по схеме «алгоритма ранжирования источников по потенциалу».

2. $f_2(S)$– индекс влиятельности источника, значение которого находится между 0 и 2, при этом к значениям применяется нормировка, такая что ко всему множеству применяется общее начало отсчета равное 0, преобразование монотонное и сохраняет отношение предпочтения на множестве допустимых решений. Метод нормировки – сравнительная нормализация, при которой за идеальное значение выбирается максимум. Применение этого метода обосновано в связи с тем, что все значения I_i^s являются неотрицательными

величинами. Вычисление индекса влиятельности происходит по схеме «алгоритма оценки источников».

3. $f_3(S)$ – приоритет источника в качестве объекта воздействия в анализируемом наборе данных. Для получения значения применяется алгоритм сортировки источников.

4. $f(C)$ – ранжированные контрмеры с учетом их сложности. Ранжирование происходит согласно алгоритму ранжирования контрмер.

Рассмотрим диаграмму комплекса алгоритмов анализа источника и ранжирования контрмер (рис. 2.15)

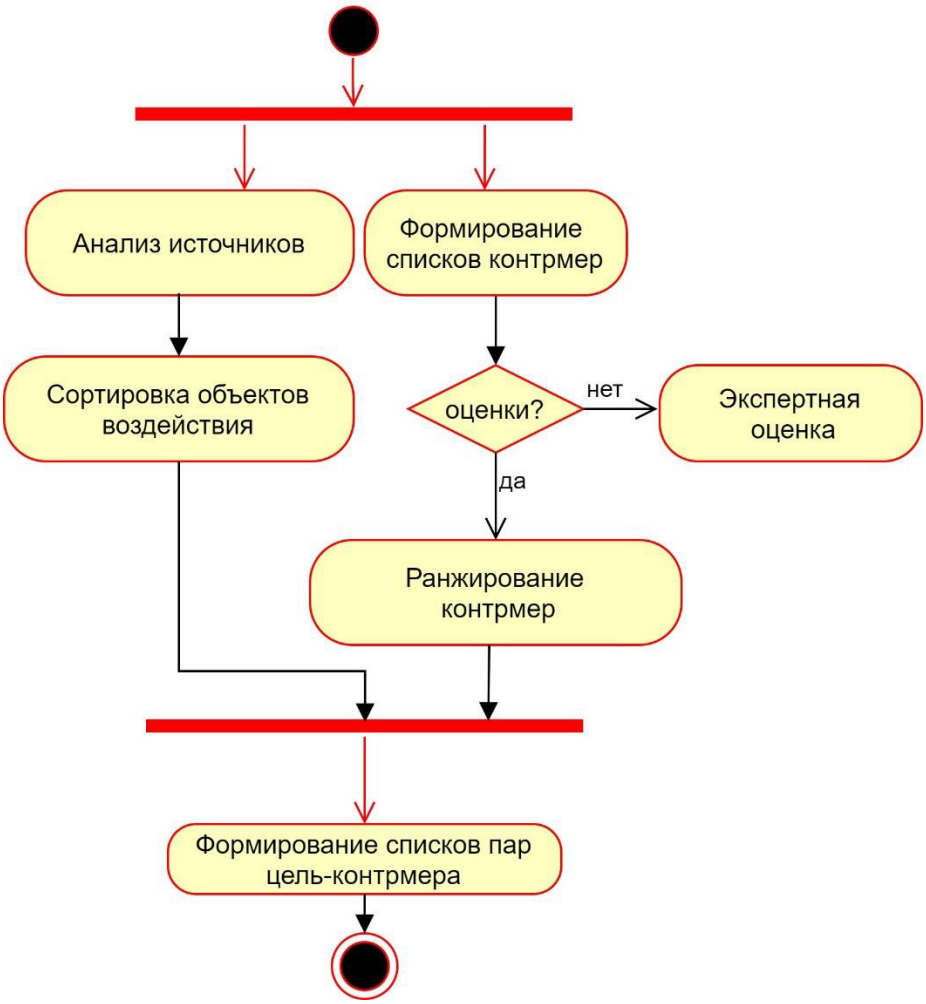


Рисунок 2.15 – Диаграмма комплекса алгоритмов анализа источников и ранжирования контрмер

Разработанный комплекс алгоритмов анализа источников вредоносной информации и ранжирования контрмер отличается от существующих

аналогов учетом таких атрибутов как потенциал источника, активность пользователей на странице источника, количество просмотров сообщения с вредоносной информацией, количество друзей и подписчиков источника.

2.4 Вывод по главе 2

1. Разработана модель социальной сети, включающая в себя сообщения, источники и связи между ними, которая отличается наличием новых структурных элементов и связей. Также разработана модель источника, в которой впервые учитываются такие параметры как индекс активности, потенциал, индекс влиятельности, индекс просматриваемости. Разработана теоретико-множественная модель вредоносной информации, состоящая из взаимосвязанных объектов и признаков вредоносной информации, вместе формирующих вредоносно-информационные объекты. Также предложена авторская информационно-признаковая модель вредоносной информации.

2. Разработан комплекс алгоритмов анализа источников вредоносной информации и ранжирования контрмер, который отличается от существующих аналогов учетом таких атрибутов как потенциал источника, активность пользователей на странице источника, количество просмотров сообщения с вредоносной информацией, количество друзей и подписчиков источника. Входящий в комплекс алгоритм ранжирования контрмер отличается от аналогов учетом коэффициентов и уровней сложности для каждой меры противодействия. При этом разработанный комплекс алгоритмов позволяет сформировать требования к методике противодействия вредоносной информации и является основой для системы принятия решений.

ГЛАВА 3. Методика и архитектура противодействия вредоносной информации в социальных сетях

3.1 Методика противодействия вредоносной информации в социальных сетях

В данном разделе диссертации представлены методика противодействия вредоносной информации в социальных сетях. Противодействие вредоносной информации является частью информационного обеспечения видов человеческой деятельности. Методика в свою очередь описывает информационную поддержку процесса принятия решений.

Информационная поддержка процесса принятия решений о противодействии вредоносной информации включает в себя следующее [154]:

- анализ собранной и обработанной информации;
- выработку на основе выпаленного анализа вариантов решений;
- оценку этих вариантов;
- выбор из них наилучшего;
- предоставление лицу, принимающему решение, выбранного и альтернативных вариантов с обоснованием выбора.

Методику можно разделить на две стадии в соответствии с жизненным циклом информационных систем: (1) настройки и (2) эксплуатации.

1. Стадия настройки противодействия и формирования исходных данных. На этой стадии задаются базовые списки информационных угроз, задаются списки доступных агентов реализации, списки доступных в системе контрмер и их коэффициенты, формируется список ранжированных контрмер.

2. Стадия эксплуатации. Сюда входят получение информации от внешней системы мониторинга, анализ объектов воздействия и сортировка. Формирование пар цель-контрмера, запуск противодействия.

Рассмотрим стадии методики более подробно. На рисунках 3.1 и 3.2 приведено общее представление методики.

3.1.1 Стадия настройки методики противодействия вредоносной информации и формирование исходных данных

Стадия настройки методики противодействия и формирование исходных данных может быть разделена на следующие шаги:

Шаг 1. Настройка системы запросов.

Оператор формирует список информационных угроз и их признаков, в соответствии с информационно-признаковой моделью угроз (п. 2.1.4 диссертации).

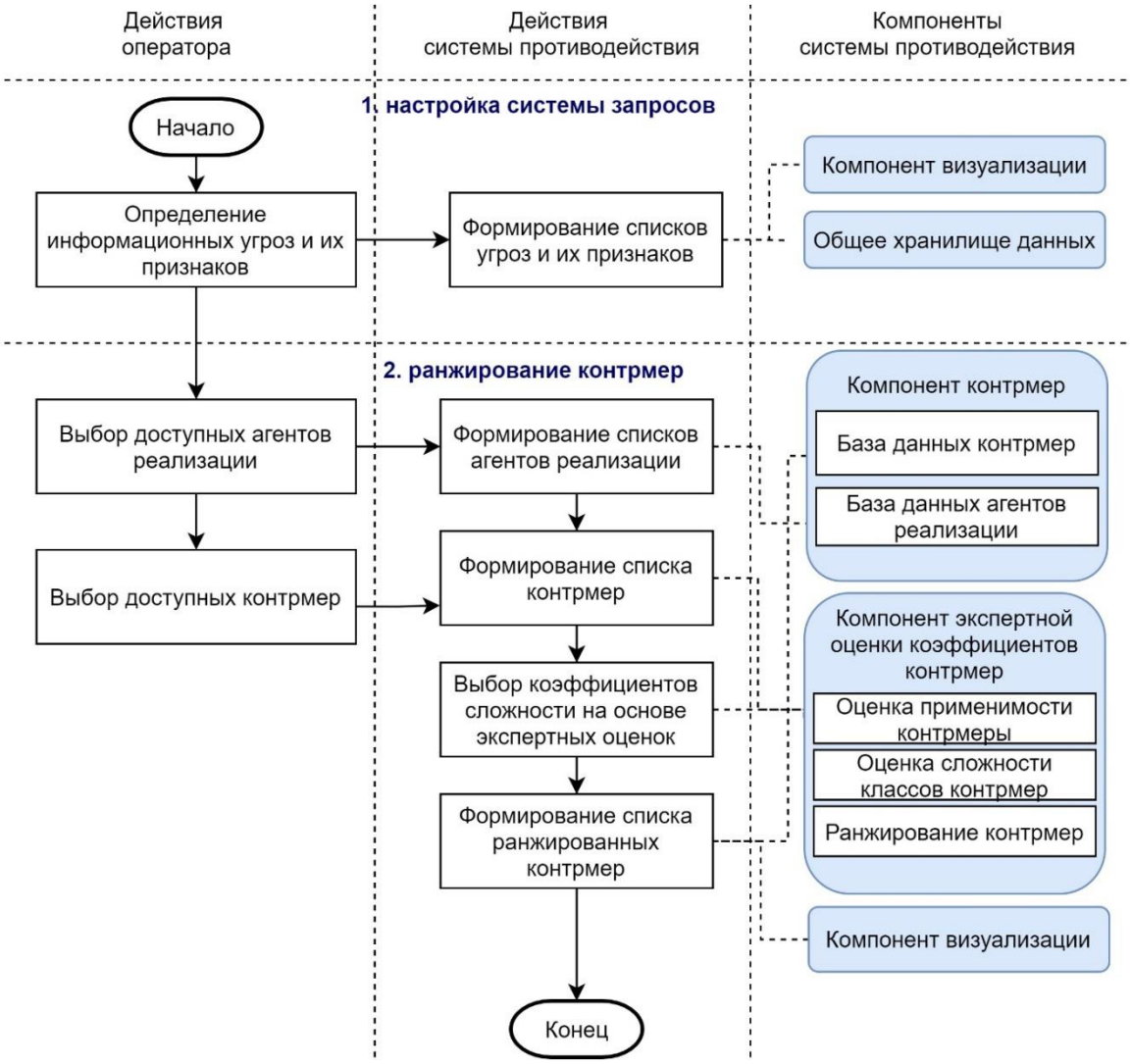


Рисунок 3.1 – Представление методики противодействия вредоносной информации на стадии настройки

После получение набора информационных угроз и их признаков от оператора, формируется список угроз и их признаков, пример таблица 3.1.

Таблица 3.1. − Пример списка угроз и их признаков

Пример угроз	**Вредоносная информация в социальных сетях**	**Информационные признаки**
T_1	Наркотики купить	a_1
	Наркотики рецепт изготовления	a_2
T_2	Взрывное устройство набор для сборки с инструкцией	b_1
T_3	Секретный алгоритм защиты телефонных звонков	c_1

В конце *Шага 1* списки угроз и их признаков сохраняются в общее хранилище данных.

Шаг 2. Ранжирование контрмер

В начале оператор выбирает доступные агенты реализации, например такие как: (1) оператор связи; (2) black_list; (3) браузер; (4) антивирус; (5) операционная система (ОС); (6) система родительского контроля (РК). Формируется список доступных агентов реализации, список сохраняется.

Затем оператор выбирает доступные контрмеры, например такие как [155]:

1) блокировка через оператора связи;
2) блокировка через социальную сеть;
3) блокировка через black_list;
4) блокировка через специальное программное обеспечение;
5) фильтрация через антивирус;
6) фильтрация через систему РК.

Формируется список контрмер (таблица А.1. в Приложении А)

Затем формируются коэффициенты сложности на основе экспертных оценок согласно *алгоритму выбора коэффициентов сложности контрмер на основе экспертных оценок*

В алгоритме используются следующие величины:

а. вес w_i, *(weight)*, который определяет вклад в сложность меры противодействия от класса KC_i;

б. уровень сложности $lc_{i,j}$ *(level of complexity)*, который определяет вклад в сложность меры противодействия от экземпляра класса $kc_{i,j}$;

в. начальная сложность cw_x контрмеры.

Величины задаются экспертным путем и зависят от квалификации сотрудников, доступных ресурсов и т.д.

В алгоритме для выбора данных значений предлагается использовать вариант экспертного метода Дельфи. Сущность Дельфи-метода экспертных оценок заключается в том, что в результате серии действий независимых экспертов формируется некое обобщенное мнение, что позволяет избежать субъективных оценок.

Алгоритм состоит из следующих шагов [3].

Шаг 1. Выбор экспертов.

Выбирается группа экспертов, которым предоставляются сведения о возможных контрмерах.

Шаг 2. Голосование.

На данном шаге определяется какие свойства применимы к контрмерам. Для каждой уточняемой величины $cp_{x,i,j}$ эксперты выставляют оценки применимости от 1 до 10.

Шаг 3. Обработка результатов.

На данном шаге выполняется усреднение полученных значений (3.1):

$$cp_{x,i,j} = \frac{\sum_{l=1}^{N} cp_{x,i,j,l}}{10*N} \quad (3.1)$$

Далее, полученное значение округляется до 0 либо 1 и это значение определяет, применим ли данный экземпляр $kc_{i,j}$ конкретного класса свойств контрмер для данной меры противодействия для данного объекта воздействия.

Шаг 4. Голосование.

Для каждой уточняемой величины (w_i, $lc_{i,j}$) эксперты выставляют оценки сложности от 1 до 10.

Шаг 5. Обработка результатов.

На данном шаге выполняется усреднение полученных значений (3.2, 3.3):

$$w_i = \frac{\sum_{l=1}^{N} w_{i,l}}{N}, \quad (3.2)$$

$$lc_{i,j} = \frac{\sum_{l=1}^{N} lc_{i,j}}{N} \quad (3.3)$$

Шаг 6. Голосование.

Для каждой контрмеры эксперты выставляют оценки начальной сложности cw_x от 1 до 10.

Шаг 7. Обработка результатов.

На данном шаге выполняется усреднение полученных значений для начальной сложности (3.4).

$$coefficient(\mathrm{c}w_x) = \frac{\sum_{l=1}^{N} \mathrm{c}w_{x,l}}{N} \quad (3.4)$$

Результатом применения данного алгоритма являются заполненные показатели для определения сложности применения контрмер.

Далее происходит ранжирование контрмер согласно алгоритму описанному в п. 2.2.4 диссертации.

В конце второго шага первой стадии методики формируется список ранжированных контрмер.

3.1.2 Стадия эксплуатации методики противодействия вредоносной информации

Рассмотрим методику противодействия вредоносной информации на стадии эксплуатации.

Стадия запроса информации и анализа объектов воздействия состоит из следующих шагов:

Шаг 1. Запрос на сбор данных

Оператор выбирает информационные угрозы из сохраненного списка (пункт 3.1.1 диссертации), если это необходимо – задает новые информационные признаки.

Оператор запускает сбор информации, система противодействия посылает запрос ко внешним системам мониторинга и получает них набор данных с сообщениями, содержащими вредоносную информацию, с источниками и параметрами, необходимыми для дальнейшего анализа.

Шаг 2. Ранжирование и сортировка объектов воздействия

В методике противодействия по алгоритму (п. 2.2.1 диссертации) источники ранжируются по потенциалу.

Далее по алгоритму (п. 2.2.2 диссертации) оцениваются источники, формируются кортежи *<messageURL, sourceID, potentialIndex, activityIndex, viewIndex, impactIndex>.*

Затем согласно алгоритму (п. 2.2.3 диссертации) сортируются объекты воздействия по приоритету и формируются списки, которые передаются оператору.

Шаг 3. Противодействия вредоносной информации

Оператор получает информацию о потенциале источника, на который влияет количество сообщений, опубликованных на его странице в социальной сети, информацию о приоритете воздействия, на который влияет уровень активности аудитории источника и количество просмотров.

Происходит корректировка оператором списков объектов воздействия, формирование пар цель-контрмера, затем проверка таких пар оператором и запуск противодействия.

Оператор передает команду на запуск противодействия, система запускает противодействие через агентов реализации и демонстрирует промежуточные результаты процесса противодействия оператору.

Далее формируется отчет о результатах противодействия, выбранной информационной угрозе и определенным в ходе эксплуатации системы объектам воздействия. Процесс противодействия завершается.

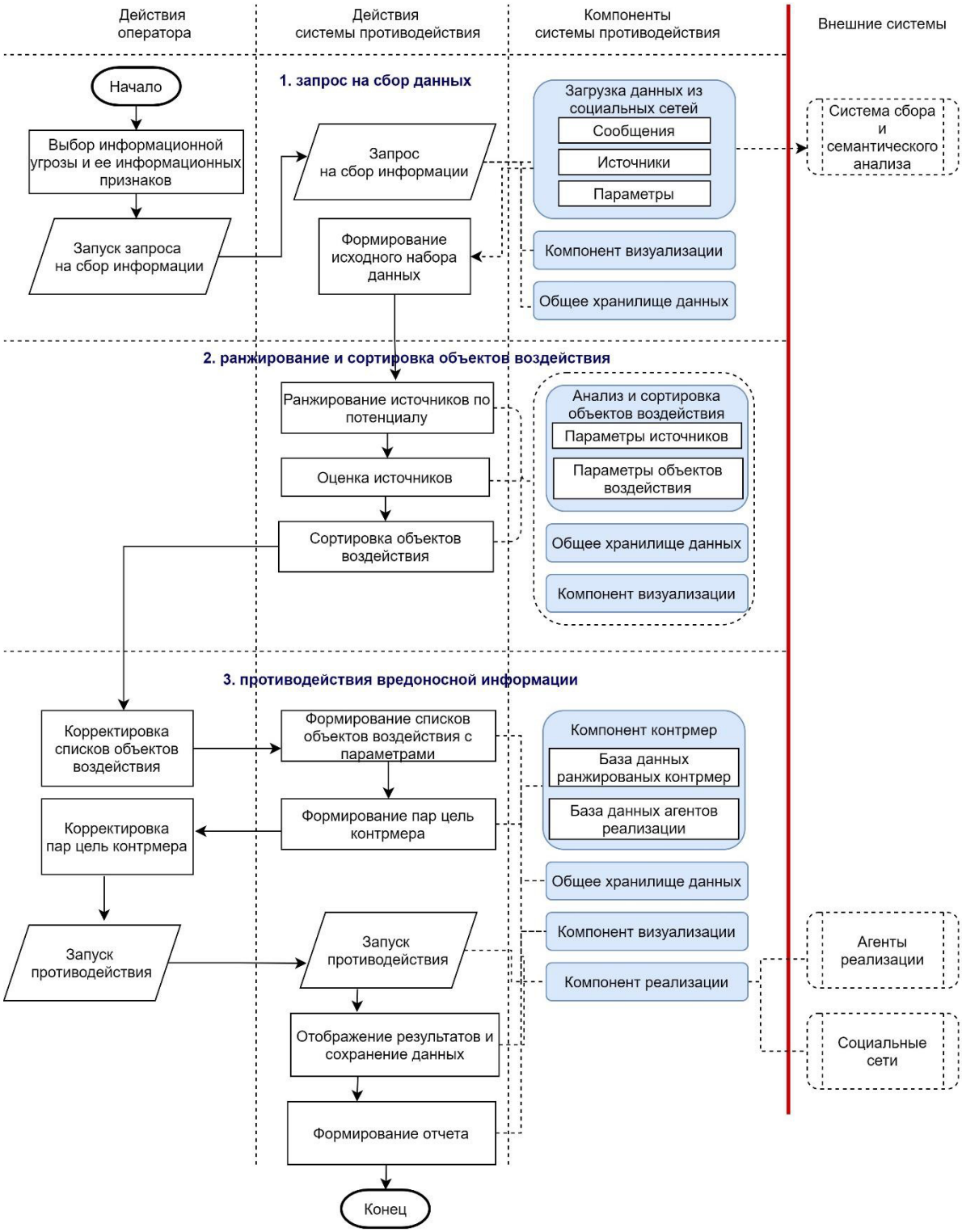


Рисунок 3.2 – Представление методики противодействия вредоносной информации на стадии эксплуатации

Выходными данными методики являются:

- возможные информационные угрозы, признаки, контрмеры и их коэффициенты, доступные агенты реализации мер противодействия;

- различные параметры объектов воздействия, согласно которым оператор распределяет свое внимание и очередность принятия решения о противодействии;

- сформированные пары цель-контрмера для противодействия вредоносной информации в социальных сетях через доступные агенты реализации.

Таким образом, предлагаемая методика противодействия вредоносной информации на разных стадиях ее жизненного цикла с учетом требований к обоснованности позволяет:

1) определить потенциал источника, зависимый от количества сообщений на его странице;

2) оценить индекс активности источника, на который влияет уровень активности аудитории сообщений с вредоносной информацией;

3) оценить индексы просматриваемости сообщений и затем источника;

4) определить индекс влиятельности источника, зависимый от активности и просматриваемости в целом;

5) определить приоритет объекта воздействия, на который влияют потенциал источника и индекс влиятельности;

6) сортировать объекты воздействия по приоритету для поддержки принятия решения оператором о выборе объекта воздействия;

7) сформировать пары цель-контрмера для поддержки принятия решения о противодействии вредоносной информации.

Разработанная методика отличается от существующих использованием авторских алгоритмов анализа и оценки источников и ранжирования контрмер, благодаря чему повышается обоснованность принятия решения о противодействии цели и выбора контрмеры и сокращается время работы оператора в процессе противодействия вредоносной информации в СС.

3.2 Архитектура и программные прототипы компонентов системы противодействия вредоносной информации в социальных сетях

В системе мониторинга происходит сбор и обработка сообщений, семантический анализ текстов. В системе противодействия – анализ источников, сортировка объектов воздействия, ранжирование контрмер и формирование для оператора, выбранного и альтернативного вариантов для противодействия.

Обзор существующих решений показывает, что мировыми лидерами в разработке систем и архитектур для мониторинга и противодействия сегодня являются крупные корпорации. Рассмотрим некоторые примеры.

Creopoint Inc [156] – компания, занимающаяся фильтром фальшивых новостей и блокировкой информации для крупных брендов и известных миллиардеров, в 2020 году зарегистрировала патент на «Сдерживание распространения дезинформации с использованием настраиваемых каналов разведки» [138]. Этой компании принадлежит также другой патент – «Распространение сообщений о дезинформации с помощью настраиваемых каналов разведки» [157].

BG Negev Technologies and Applications Ltd – венчурная компания Университета Бен-Гуриона. Им принадлежит множество разработок, патентов в области информационной безопансости. В частности в 2017 получен патент на способ обнаружения спамеров и поддельных профилей в социальных сетях [136].

Zerofox Inc – крупный игрок на рынке систем мониторинга и противодействия вредоносной информации. Их продукт представляет из себя технологию SaaS, которая собирает и обрабатывает данные из Instagram, Facebook, Slack, Twitter, Instagram, Pastebin, YouTube, с сайтов магазинов мобильных приложений, deep & dark web, различных доменов, из электронной почты и тд. Им принадлежит патент на систему защиты от подозрительных социальных субъектов [133].

International Business Machines Corp (IDM) также постоянно разрабатывает новые продукты в области информационной безопансости, например ими получен патент на систему аналитики для смягчения злоупотреблений со стороны интернет-троллей [135] в 2018 году.

Другие релевантные системы и архитектуры, которые также рассматривались в первой главе принадлежат таким компаниям как WEBSAFETY Inc [134] Ithreat Cyber Group Inc [153].

В России крупным разработчиком систем противодействия вредоносной информации является Лаборатория Касперского [126].

Большинство из рассмотренных решений представляет собой системы, разделенные на компоненты и уровни, для которых отдельно описываются принципы обработки потоков данных.

3.2.1 Архитектура системы противодействия вредоносной информации в социальных сетях

На основании проведенного исследования решений мировых производителей в области разработок систем противодействия вредоносной информации в социальных сетях сформируем общие для СПД и специфические требования.

Общие требования:

1) разделенность на уровни;
2) разделенность на компоненты;
3) взаимодействие с внешними сервисами и системами

Специфические требования:

1) поддержка процессов сбора информации по запросу оператора;
2) поддержка процессов анализа источника и сортировки объектов воздействия;
3) поддержка процессов формирования исходных данных, ранжирования контрмер;
4) поддержка процессов противодействия вредоносной информации.

Разработанные методики и алгоритмы реализованы в рамках системы противодействия вредоносному влиянию в социальных сетях, которая является частью системы мониторинга и противодействия, разработанной в проекте РНФ [158]. Предлагаемая архитектура системы противодействия (СПД) вредоносной информации в социальных сетях включает три уровня и восемь компонентов (рис. 3.3):

1. уровень управления:
1.1. компонент менеджмента,
1.2. компонент визуализации;
2. уровень оценки содержания:
2.1. компонент анализа и оценки источников,
2.2. SQL сервер
2.3. база данных;
3. уровень реализации контрмер:
3.1. компонент выбора контрмер
3.2. компонент реализации контрмер.

Для определения функций, выполняемых компонентами системы, рассмотрим ее функциональную структуру (рис.3.4).

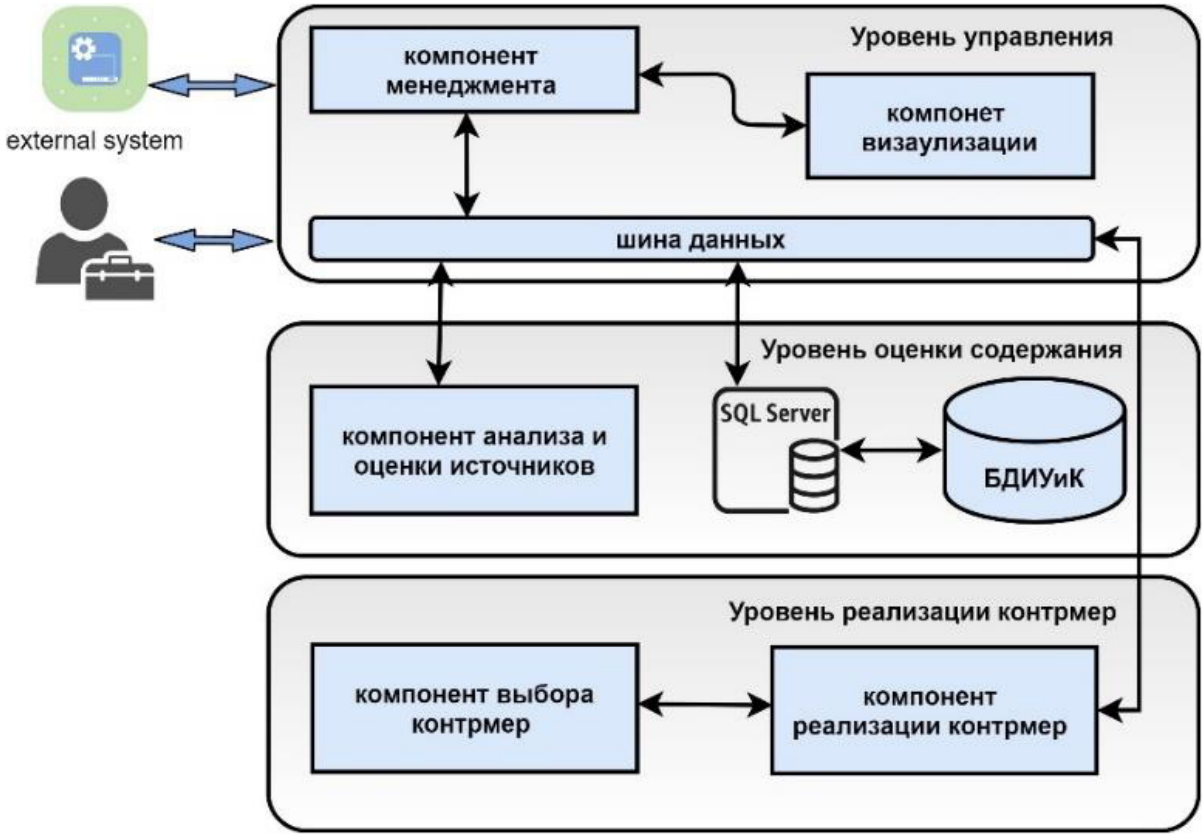


Рисунок 3.3 – Архитектура СПД вредоносной информации в социальных сетях.

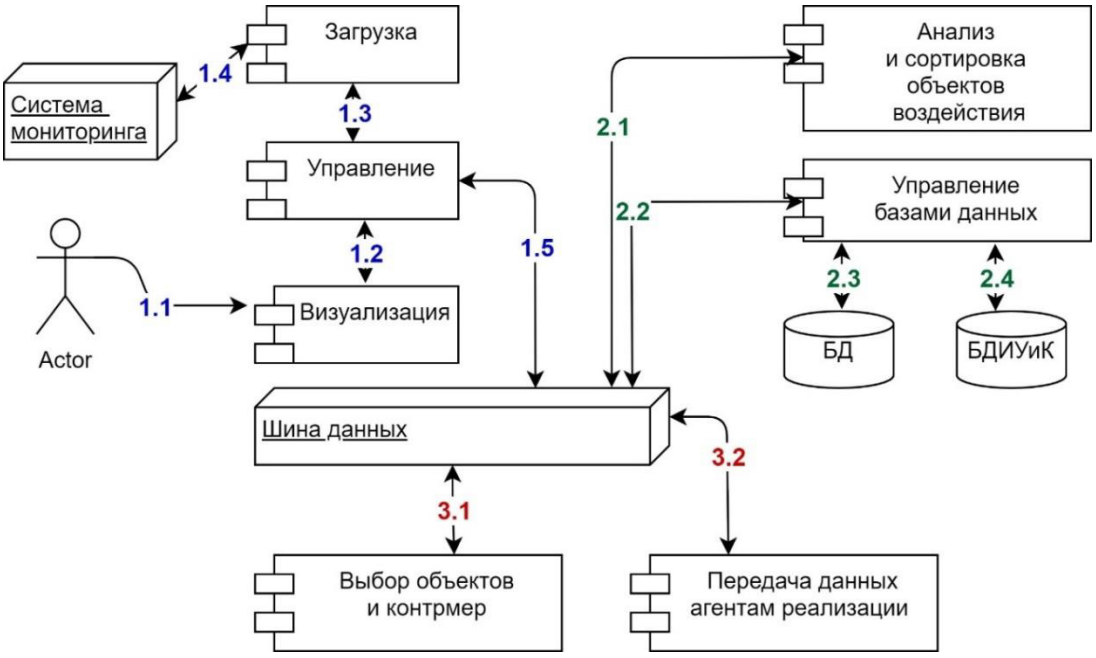


Рисунок 3.4 – Функциональная структура архитектуры системы противодействия вредоносной информации в социальных сетях.

Функциональная структура делится в соответствии с уровнями архитектуры и компонентами:

1. Уровень управления.

Функции компонента менеджера включают управление нагрузками и задачами внутри системы. Функции компонента визуализации обеспечивают связь с оператором системы противодействия. Шина данных является частью системной шины, предназначенная для передачи данных между всеми компонентами системы противодействия.

2. Уровень оценки содержания.

Компонент анализа и оценки источников включает в себя функции ранжирования по потенциалу, оценку параметров объектов воздействия, их приоритезацию. На этом же уровне находится компонент управления базами данных (сервер SQL), в функции которого входит управление базой данных информационных угроз и контрмер.

3. Уровень реализации контрмер.

В функции компонента выбора контрмер входят задачи системы поддержки принятия решения, формирование списков целей и контрмер, а через компонент реализации контрмер поддерживается связь со внешними агентами и системами реализации.

3.2.2 Программные прототипы компонентов системы противодействия

Элементы архитектуры реализованы в качестве программных прототипов: (1) программный прототип компонента анализа и оценки источников в СС, который включает алгоритм ранжирования источников, алгоритм оценки источников и алгоритм сортировки объектов воздействия; (2) программный прототип компонента выбора контрмер, который включает алгоритм ранжирования контрмер, алгоритм экспертных оценок для формирования коэффициентов; (3) программный прототип базы данных информационных угроз и контрмер (БДИУиК), который содержит

информацию о мерах противодействия вредоносной информации в СС, о типах информационных объектов, к которым контрмеры могут быть применимы, об агентах реализации, через которые могут быть реализованы меры противодействия.

3.2.2.1 Программный прототип компонента анализа и оценки источников в социальных сетях

Компонент состоит из нескольких связанных в комплекс алгоритмов:

1) алгоритм ранжирования источников по потенциалу;
2) алгоритм оценки источников по активности;
3) алгоритм сортировки объектов воздействия.

На вход в алгоритм ранжирования источников по потенциалу подается набор кортежей *<messageURL, messageType, sourceID>*, на выходе – набор кортежей *<sourceID, potentialIndex>*. Фрагмент кода алгоритма представлен в листинге на рисунке 3.5.

На вход в алгоритм оценки источников по активности подается набор кортежей *<messageURL, sourceID, likesCount, commentCount, repostCount, viewCount, subscriberCount>, на* выходе – набор кортежей *<sourceID, activityIndex, viewIndex, impactIndex>.* Фрагмент кода алгоритма представлен в листинге на рисунке 3.6.

```
#Шаг 1 Расчет суммы сообщений источника
for i in range (len(list(SourcesPotentialCalculation['messageURL']))):
    SourcesPotentialCalculation.loc[i,'potentialIndex']=0
    if SourcesPotentialCalculation['messageType'][i]== 'post':
SourcesPotentialCalculation.loc[i,'potentialIndex']=SourcesPotentialCalculation['potenti
alIndex'][i]+1
    if SourcesPotentialCalculation['messageType'][i]== 'comment':
SourcesPotentialCalculation.loc[i,'potentialIndex']=SourcesPotentialCalculation['potenti
alIndex'][i]+0.5
    if SourcesPotentialCalculation['messageType'][i]== 'reply to comment':
SourcesPotentialCalculation.loc[i,'potentialIndex']=SourcesPotentialCalculation['potenti
alIndex'][i]+0.25
## Шаг 2 Расчет потенциала источника
## Расчет первого среднего
firstAverage = 0
count = 0
for i in range (len(list(SourcesPotentialCalculation['messageURL']))):
    firstAverage = firstAverage + SourcesPotentialCalculation['potentialIndex'][i]
    count = count + 1
firstAverage = firstAverage /  count
# Расчет второго среднего
secondAverage = 0
count = 0
for i in range (len(list(SourcesPotentialCalculation['messageURL']))):
    if SourcesPotentialCalculation['potentialIndex'][i]>= firstAverage:
        secondAverage = secondAverage + SourcesPotentialCalculation['potentialIndex'][i]
        count = count + 1
secondAverage = secondAverage /  count
# Формирование результата
for i in range (len(list(SourcesPotentialCalculation['messageURL']))):
    if SourcesPotentialCalculation['potentialIndex'][i]<firstAverage:
        SourcesPotentialCalculation.loc[i,'potentialIndex']=0
    elif SourcesPotentialCalculation['potentialIndex'][i]>=firstAverage and
SourcesPotentialCalculation['potentialIndex'][i]<secondAverage:
        SourcesPotentialCalculation.loc[i,'potentialIndex']=1
    elif SourcesPotentialCalculation['potentialIndex'][i]>=secondAverage:
        SourcesPotentialCalculation.loc[i,'potentialIndex']=2
```

Рисунок 3.5 – Фрагмент кода алгоритма ранжирования источников по потенциалу

```
#Алгоритм оценки источников по активности
Шаг 1. Вычисление индекса активности
        if SourcesActivityCalculation['sourseId'][j]==subscriberCount['sourseId'][i]:
SourcesActivityCalculation.loc[j,'subscriberCount']=subscriberCount['subscriberCount'][i]
urlCounter=0
for i in range (len(list(SourcesActivityCalculation['messageURL']))):
    for j in range (len(list(SourcesActivityCalculation['messageURL']))):
        if i!=j:
            if SourcesActivityCalculation['sourseId'][i]== SourcesActivityCalculation['sourseId'][j]:
                urlCounter=urlCounter+1
urlCounter=list(SourcesActivityCalculation['sourseId'])
urlCounter =len( set(urlCounter))
activityIndex = pd.DataFrame({
    'sourseId':[],
    'activityIndex': [],  })
for i in range(len(list(SourcesActivityCalculation['sourseId']))):
    activityIndex.loc[i,'sourseId']= SourcesActivityCalculation['sourseId'][i]
activityIndex.loc[i,'activityIndex']=SourcesActivityCalculation['likesCount'][i]+SourcesActivityCalcul
ation['commentCount'][i]+SourcesActivityCalculation['repostCount'][i]
for i in range(len(list(activityIndex['sourseId']))):
    if SourcesActivityCalculation['subscriberCount'][i]!=0:
activityIndex.loc[i,'activityIndex']=activityIndex['activityIndex'][i]/SourcesActivityCalculation['subscr
iberCount'][i]
    else:
        activityIndex.loc[i,'activityIndex']=activityIndex['activityIndex'][i]/1
    activityIndex.loc[i,'activityIndex']=activityIndex['activityIndex'][i]/urlCounter
#Шаг 2 Вычисление индекса просматриваемости источника
viewIndex   = pd.DataFrame({
    'sourseId':[],
    'viewIndex': [],  })
for i in range(len(list(SourcesActivityCalculation['sourseId']))):
    viewIndex.loc[i,'sourseId']= SourcesActivityCalculation['sourseId'][i]
    viewIndex.loc[i,'viewIndex']= SourcesActivityCalculation['viewCount'][i]
for i in range(len(list(viewIndex['sourseId']))):
    if SourcesActivityCalculation['subscriberCount'][i]!=0:
viewIndex.loc[i,'viewIndex']=viewIndex['viewIndex'][i]/SourcesActivityCalculation['subscriberCount'
][i]
    else:
        viewIndex.loc[i,'viewIndex']=viewIndex['viewIndex'][i]/1
    viewIndex.loc[i,'viewIndex']=viewIndex['viewIndex'][i]/urlCounter
 # Шаг 3 Вычисление индекса влиятельности источника
 impactIndex   = pd.DataFrame({
    'sourseId':[],
    'impactIndex': [],  })
for i in range(len(list(SourcesActivityCalculation['sourseId']))):
    impactIndex.loc[i,'sourseId']= SourcesActivityCalculation['messageURL'][i]
    impactIndex.loc[i,'impactIndex']= activityIndex['activityIndex'][i]+viewIndex['viewIndex'][i]
#Формирование результата
InfluenceObjectSorting = pd.DataFrame({ ...})
```

Рисунок 3.6 − Фрагмент кода алгоритма оценки источников по активности

На вход в алгоритм сортировки объектов воздействия, входящий в программный прототип компонента анализа и оценки источников в социальных сетях, подается набор кортежей *<messageURL, sourceID, potentialIndex, activityIndex, viewIndex, impactIndex>,* на выходе формируются списки:

- target_1 [sourceID, messageURL] – высокий приоритет
- target_3 [sourceID, messageURL] – низкий приоритет
- target_2 – набор кортежей *<sourceID, messageURL, potentialIndex, activityIndex, viewIndex, impactIndex>* – объекты со средним приоритетом, для которых потребуется дополнительная оценка оператором.

Программный компонент написан на языке Python, версия 3.8. В компоненте используются внешние модули из стандартных библиотек языка программирования: time, pd (pandas).

3.2.2.2 Программный прототип компонента выбора контрмер

Компонент состоит из двух независимых алгоритмов и работает в двух режимах:

1) алгоритм экспертной оценки для формирования коэффициентов;
2) алгоритм ранжирования контрмер.

В начале работы алгоритма экспертной оценки для формирования коэффициентов формируется хеш-таблица (рис. 3.7). Далее запрашиваем (через консольный ввод) данные от эксперта, суммируем результаты и вычисляем среднее значение по коэффициенту. На выходе формируется список, который сохраняется в файл в формате csv.

Входные данные для алгоритма ранжирования контрмер – это список, полученный в результате работы алгоритма экспертной оценки коэффициентов контрмер. В момент начала работы алгоритма подгружается csv файл и происходит ранжирование согласно формуле (2.9) описанной в п. 2.2.4 диссертации.

```
df = pd.DataFrame({
    'Контрмеры': [ ],
    'Значение контрмеры':[],
    'Тип':[],
    'Замещение':[],
   'Значение контрмеры Замещение':[],
    'Блок':[],
'Значение контрмеры Блок':[],
    'Зашумление':[],
'Значение контрмеры Зашумление':[],
    'Метод':[],
    'Ручной':[],
'Значение контрмеры Ручной':[],
    'Автоматизированный':[],
'Значение контрмеры Автоматизированный':[],
    'Смещанный':[],
'Значение контрмеры Смещанный':[],
    'Широта':[],
    'Группа':[],
'Значение контрмеры Группа':[],
    'Сообщение':[],
'Значение контрмеры Сообщение':[],
})
```

Рисунок 3.7 − Входные параметры алгоритма ранжирования контрмер

Таким образом, программный прототип компонента выбора контрмер, включающий алгоритм ранжирования контрмер и алгоритм экспертных оценок для формирования коэффициентов, позволяет сформировать в системе списки ранжированных контрмер и обеспечить их оценку со стороны экспертов.

Программный компонент также написан на языке Python, 3.8. В нем используются внешние модули из стандартных библиотек языка программирования: time, pd (pandas).

3.2.2.3 Программный прототип базы данных информационных угроз и контрмер (БДИУиК)

Программный прототип базы данных информационных угроз и контрмер является частью архитектуры противодействия вредоносной информации в социальных сетях. При этом, под контрмерой понимается действие или средство защиты объекта от угрозы информационной безопасности, где угроза информационной безопасности представлена вредоносной информацией. Формальный вид соответствующей применимой контрмеры $Measure$ может быть представлен следующим образом:

$$Measure =< Threat, Countermeasure >,$$

где $Threat$ – рассматриваемая угроза, *Countermeasure* – соответствующая контрмера, учитывающая атрибуты ее реализации и тип применимого информационного объекта.

Представим выделенные атрибуты и их взаимосвязи формально:

$$Threat =< Definition, Token, Keys\_words >.$$

Перечислим элементы, входящие в этот кортеж:

$Definition$ – описание угрозы в человекопонятном виде;

$Token$ – некоторые признаки угрозы информационной безопасности, позволяющие некому оператору, однозначно классифицировать угрозу;

$Keys\_words$ – семантические признаки угрозы для упрощения классификации угрозы.

Аналогичным образом опишем атрибуты контрмеры и их взаимосвязи:

$$Countermeasure =< Object, Agent, Implementation_{type}, Phase.$$

Перечислим элементы, входящие в этот кортеж:

$Object$ – информационный объект, к которому применима контрмера;

$Agent$ – агент реализации, через который может быть реализована контрмера;

$Implementation\_type$ – тип реализации контрмеры (прим.: ручной, авто и т.д.);

$Phase = Static \vee Dynamic$ – этап реализации контрмеры; контрмеры могут применяться на этапе реагирования на вредоносную информацию

(т.н. динамические контрмеры) в социальной сети, а также на этапе предотвращения ее распространения (т.н. статические контрмеры).

В приложениях Б, В представлены диаграмма базы данных информационных угроз и контрмер (Приложение Б) и ее структура (Приложение В).

Программные прототипы компонентов являются частью архитектуры противодействия вредоносной информации, часть зарегистрировано в реестре программ ЭВМ [16, 17].

3.3 Экспериментальная и теоретическая оценка методики противодействия вредоносной информации в социальных сетях

Для экспериментальной оценки был подготовлен стенд со следующими характеристиками:

1) ПК «DESKTOP-8M0KI8G»/Intel(R) Core (TM) i5-9600CPU 3.10GHz/ DDR16.00G/128SSD/1000HDD.

2) Операционная система «Edition Windows 10 Education Version 20H2», установлена 6/11/2020, сборка 19042.685/Experience Windows Feature Experience Pack 120.2212.551.0.

3) 2 шт. монитора DELL

4) Дополнительное ПО:

- Microsoft Office Standard 2019;
- Pycharm 2020.14 (Community Edition), сборка #PC-201.8743.11, версия встроенной VM: OpenJDK 64-Bit Server VM by JetBrains, язык Python 3.8;
- Редактор Notepad++;
- MySQL Workbench 8.0;
- Microsoft SQL Server 2014.

5) Внешние сервисы:

- Google Chrome (сервисы Google Формы, Google Документы);
- Компонент сбора данных из социальной сети [160];

- Компонент сбора комментариев к постам в СС [161];

6) Разработанные программные прототипы и компоненты:

- Программный прототип компонента анализа и оценки источников в социальных сетях;
- Программный прототип компонента выбора контрмер [16]
- Программный прототип базы данных информационных угроз и контрмер (БДИУиК) [17]

Экспериментальная оценка проводилась в несколько этапов. В начале оценивались разработанные программные прототипы и компоненты, затем экспериментально оценивалось время работы оператора при противодействии вредоносной информации без использования методики, и время работы оператора с использованием методики. Ресурсопотребление оценивалось на основе полученных на двух предыдущих этапах данных.

3.3.1 Экспериментальная оценка комплекса алгоритмов анализа источника и ранжирования контрмер

В качестве информационной угрозы был выбран инцидент, связанный с аварией в Москве 8 июня 2019 года. Определены хронологические рамки исследования с 8 июня по 30 октября 2019г. Проведен полнотекстовый поиск по следующим ключевым словам: («ДТП Михаил Ефремов», «Михаил Ефремов Авария», «Сергей Захаров Михаил Ефремов», «Актер Михаил Ефремов», «Либерал Михаил Ефремов», «Оппозиционер Михаил Ефремов»).

Было собрано из СС 15132 сообщения, среди которых посты, комментарии, ответы к комментариям. Для каждого сообщения были собраны сведения о количестве лайков, комментариев, репостов, просмотров, получены сведения с названием источника (рис. 3.8). Для сбора данных использовались [160], [161]. Данные были получены в формате csv и преобразованы в книгу Excel.

	sourseId	messageURL	messageType	likesCoun	comment	repostCou	viewCoun	sul
1								
2	#necro_tv	-34410764_3418879	post	183	160	4	25926	
3	#Белковский	-87516069_3550	comment	1	0	0	37	
4	#ГОВОРИТ ЧЕЛЯБИНСІ	-102627320_11036	reply to comme	1	0	0	227	
5	#ГОВОРИТ ЧЕЛЯБИНСІ	-102627320_11067	post	1	1	0	288	
6	#НОД Кадуй	-196176869_1425	reply to comme	2	0	0	30	
7	#Петербург КультМир	-185485741_4687	reply to comme	0	0	0	136	
8	#ПрограммаСулакшин	-174833102_17472	comment	3	1	1	298	
9	#ТУТ #Омск	-198370307_27	comment	0	0	0	36	

Рисунок 3.8 Пример экспериментального набора данных

Далее большой набор данных был разделен на 10 малых наборов по 1000 сообщений в каждом. Каждый малый набор был проанализирован и ранжирован с использованием программного прототип компонента анализа и оценки источников в социальных сетях, получены следующие результаты и характеристики (таблица 3.2, таблица 3.3).

Таблица 3.2. − Результаты анализа и сортировки объектов воздействия

Набор данных	**Target1 (Source)**	**Target2 Объекты для оператора**	**Target3 (MessageURL)**
1	11	96	570
2	14	87	553
3	10	81	598
4	9	58	588
5	5	82	617
6	4	55	627
7	2	12	661
8	2	21	631
9	1	32	673
10	13	105	568

В столбце Target1 объектом воздействия рекомендуется Source и показано количество источников с высоким приоритетом для противодействия, которым принадлежит 334 сообщения из 1000 для 1 набора данных, 360 сообщений из 1000 для второго и тд.

В столбце Target2 показано количество сообщений, для которых приоритет средний и требуется дополнительная оценка оператором.

В столбце Target3 объектом воздействия рекомендуется MessageURL и показано количество таких сообщений с низким приоритетом для каждого набора.

Таким образом последовательность работы оператора согласно полученным результатам следующая (для 1го набора данных): 1) оператору необходимо согласовать 11 объектов воздействия (источников) с высоким приоритетом на предмет противодействия им; 2) оператору необходимо провести анализ 96 объектов воздействия с учетом всех характеристик (количество комментариев, лайков, просмотров, репостов, индекса активности, индекса просматриваемости, потенциала, индекса влиятельности); 3) проверить 570 объектов воздействия в последнюю очередь, в связи с их низким приоритетом для противодействия.

Экспериментальная оценка комплекса алгоритмов показала работоспособность подхода к анализу и сортировке объектов воздействия.

Таблица 3.3. − Результаты экспериментальной оценки характеристик работы программного прототипа компонента анализа и оценки источников в социальных сетях

Набор данных	Time/сек для Алгоритма	Дополнительная нагрузка на CPU	Дополнительная нагрузка на память
1	42.5323	25%	512Mb
2	40.8572	22%	512Mb
3	41.0728	28%	128Mb
4	41.1996	24%	300Mb
5	41.7115	29%	100Mb
6	41.1129	22%	128Mb
7	40.6353	28%	212Mb
8	40.6813	22%	300Mb
9	42.4953	21%	410Mb
10	41.0861	28%	512Mb

Далее проводилась экспериментальная оценка алгоритма ранжирования контрмер.

В начале был составлен список контрмер, зависимых от агентов реализации. Затем были приглашены к участию в эксперименте 10 экспертов, которым направлялась анкета для голосования, оформленная в сервисе Google Формы.

В исследовании приняли участие следующие эксперты:

1. Государственные служащие, специалисты-эксперты по надзору и контролю в СМИ (3 эксперта).

2. Сотрудники репутационного агентства (2 руководителя, 1 менеджер среднего звена, итого 3 эксперта).

3. Сотрудники Высшей школы журналистики и массовых коммуникаций Санкт-Петербургского государственного университета (1 доктор политических наук, 2 кандидата политических наук, 1 аспирант, итого 4 эксперта).

На первом этапе голосования эксперты оценивали возможность использования контрмеры для противодействия вредоносной информации в социальных сетях.

Затем экспертам для следующего голосования была отправлена сводная таблица, в которой для каждой уточняемой величины эксперты выставляли оценки сложности от 1 до 10.

При этом экспертами оценивались следующие величины:

w_i – вес класса способа противодействия.

$lc_{i,j}$ – уровень сложности класса (level of complexity).

cw_x – начальная сложность реализации контрмеры.

Полученные оценки от экспертов были усреднены и данные переданы в алгоритм ранжирования контрмер.

На выходе были получены следующие результаты (таблица 3.4, первые 14 строк из 35).

Таблица 3.4 − Результат экспертной оценки контрмер и их последующего ранжирования

	Метод воздействия			Тип воздействия			**Сложность**
	1			2			
	Позит-й	Нег-й.	Смеш.	Авто	Автом-й	Ручной	
Контрмера	2	1	3	1	2	3	
Уведомление ЕАИС о сообщении	1	0	0	0	0	1	4
Уведомление ЕАИС об источнике	1	0	0	0	0	1	4
Блокировка сообщения в браузере	0	1	0	1	0	0	6
Блокировка источника в браузере	0	1	0	1	0	0	6
Блокировка сообщения через антивирус	0	1	0	1	0	0	6
Блокировка источника через антивирус	0	1	0	1	0	0	6
Блокировка сообщения через систему РК	0	1	0	1	0	0	6
Блокировка источника через систему РК	0	1	0	1	0	0	6
Блокировка сообщения через ОС	0	1	0	1	0	0	6
Блокировка источника через ОС	0	1	0	1	0	0	6
Блокировка сообщения через ЕАИС	0	1	0	0	1	0	6
Блокировка источника через ЕАИС	0	1	0	0	1	0	6
Блокировка сообщения через социальную сеть	0	1	0	1	1	0	8
Блокировка источника через социальную сеть	0	1	0	1	1	0	8

В результате проведенного эксперимента были ранжированы контрмеры с учетом коэффициентов и уровней сложности для каждой меры противодействия.

3.3.2 Экспериментальная оценка методики противодействия вредоносной информации в социальных сетях

Рассмотрим основные нефункциональные требования к методике, разделенные на следующие группы: (1) оперативность; (2) обоснованность; (3) ресурсопотребление.

3.3.2.1 Оценка оперативности

Основными шагами методики являются:

1. На стадии настройки противодействия вредоносной информации: (а) настройка системы запросов; (б) ранжирование контрмер.

2. На стадии эксплуатации противодействия вредоносной информации: (а) запрос на сбор данных; (б) ранжирование и сортировка объектов воздействия; (в) противодействия вредоносной информации.

В общем случае время выполнения методики противодействия вредоносной информации в социальных сетях будет складываться из продолжительности операций рассматриваемых стадий и шагов [142] (3.5):

$$T^{M} = T_1^{HC} + T_2^{HC} + T_1^{ЭК} + T_2^{ЭК} + T_3^{ЭК}, \qquad (3.5)$$

где T_i – время выполнения i-го шага, $i = 1:5$.

Время выполнения шагов методики противодействия вредоносной информации в социальных сетях рассматривается как случайная величина, вероятность которой подчиняется нормальному закону распределения. В существующей литературе, для оценки времени выполнения наиболее часто применяется закон бета-распределения в интервале $[t_{\min}, t_{\max}]$ с плотностью распределения [142, 162]:

$$f(t) = \begin{cases} \frac{(t-t_{min})^{\alpha-1}(t_{max}-t)^{\beta-1}}{(t_{max}-t_{min})^{\alpha+\beta-1}B(\alpha,\beta)}, t_{min} \leq t \leq t_{max} \\ 0, t_{max} \leq t \leq t_{min} \end{cases}, \qquad (3.6)$$

где t_{min} и t_{max} – минимальное и максимальное время выполнения, t – величина, определяющая время выполнения, $B(\alpha,\beta)$ –функция Эйлера, $\alpha > 0$, $\beta > 0$ – параметры бета-распределения.

Ожидаемое время выполнения методики и его дисперсия рассчитываются с помощью двухоценочной методики [142].

Вероятность того, что время выполнения шага в целом будет не выше допустимого значения $T^{additional}$, вычисляется по формуле (3.7):

$$P_{op}\left(T \le T^{additional}\right) = \Phi(Z), \tag{3.7}$$

где $\Phi(Z)$ – значение функции Лапласа при (3.8):

$$Z = \frac{T^{additional} - \sum_{i=1}^{n} T_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2(T_i)}}. \tag{3.8}$$

Для формирования исходных данных были проведены исследования и эксперименты, выяснилось, что самым затратным процессом с точки зрения оперативности является время работы оператора на 1,2 шагах на стадии настройки и 1, 4 шагах на стадии эксплуатации. Временные затраты на другие процессы не превышают нескольких секунд.

Экспертам были последовательно поставлены задачи:

1. Оценить время, необходимое оператору на определение информационных угроз и их признаков так, чтобы на выходе эксперт заполнял таблицу с 3-мя информационными угрозами и их признаками.

2. Оценить время, затрачиваемое оператором на выбор доступных агентов реализации и контрмер так, чтобы на выходе эксперт заполнял таблицу с 5 контрмерами и доступными агентами реализации.

3. Оценить время, необходимое оператору на запуск запроса на сбор и анализ информации.

4. Оценить время, затрачиваемое оператором на выбор объектов воздействия и корректировку пар, цель-контрмера.

На основе проведенных исследований и экспериментов были получены основные временные показатели работы оператора для стадий и шагов

методики противодействия вредоносной информации в социальных сетях. Полученные значения приведены в таблице 3.5.

Таблица 3.5. − Временные показатели работы оператора с использованием методики противодействия вредоносной информации в социальных сетях

Шаг	T_i^{min} **, мин**	T_i^{max} **, мин**	$T_i = \frac{3T_i^{min} + 2T_i^{max}}{5}$	$\sigma^2(T_i) = 0.4\left(T_i^{max} - T_i^{min}\right)^2$
$T_1^{НС}$	*47*	*63,1*	*53,44*	*103,6840*
$T_2^{НС}$	*10,2*	*14,5*	*12*	*7,3960*
$T_1^{ЭК}$	*1*	*1,34*	*1,14*	*0,0460*
$T_3^{ЭК}$	*1,58*	*8,01*	*4,15*	*16,5370*
Итого по методике, мин			***70,73***	***127,6630***

Для сравнения временных показателей методики с процессом противодействия вредоносной информации без использования методики были также проведены исследования. Экспертам предлагалось оценить время на формирование запроса к системе мониторинга (шаг 1), время на принятия решения о выборе объекта воздействия (шаг 2), время на принятие решения о выборе контрмер (шаг 3). Получено экспериментальное значение для $T^{additional} = 102$ минутам.

Значение функции Лапласа $\Phi(Z)$ при $T^{additional} = 102$ минуты для методики (3.9):

$$\left(\frac{T^{additional} - \sum_{i=1}^{4} T_i}{\sqrt{\sum_{i=1}^{4} \sigma_i^2(T_i)}}\right) = \left(\frac{102 - 70{,}73}{\sqrt{127{,}663}}\right) \approx 2{,}767. \qquad (3.9)$$

Таким образом, по значениям функции Лапласа, заданных в табличном виде, вероятность выполнения методики за заданное время составляет $P_{op}\left(T_m \le T^{additional}\right) = 0.9942$, что соответствует предъявляемым требованиям ($P_{op}^{additional} = 0{,}99$) к *оперативности*.

Одновременно с этим исследования показали, что сократилось общее время работы оператора противодействия вредоносной информации с 102,08 минут до 70,73.

3.3.2.2 Оценка ресурсопотребления

Оценка ресурсопотребления в диссертации проводилась по ряду частных показателей, характерных для 2-го шага стадии эксплуатации методики противодействия вредоносной информации:

1. Использование центрального процессорного устройства [142] (3.10):

$$R_{CP} = \frac{Q_{CP}^{M}}{Q_{CP}^{GEN}}, \quad (3.10)$$

где Q_{CP}^{M} – время центрального процессора, потраченное на выполнение методики, Q_{CP}^{GEN} – общее доступное процессорное время;

2. Использование оперативной памяти [142] (3.11):

$$R_{DDR} = \frac{Q_{DDR}^{M}}{Q_{DDR}^{GEN}}, \quad (3.11)$$

где Q_{DDR}^{M} – объем оперативной памяти, использованный в процессе выполнения методики, Q_{DDR}^{GEN} – общий объем оперативной памяти.

(3) Время работы оператора [142] (3.12):

$$R_{expert} = \frac{Q_{expert}^{M}}{Q_{expert}^{GEN}}, \quad (3.12)$$

где Q_{expert}^{M}– время работы оператора, потраченное на противодействие вредоносной информации в социальных сетях по методике, Q_{expert}^{GEN}– общее время работы оператора.

Для исходных данных $R_{CP} = \frac{Q_{CP}^{M}}{Q_{CP}^{GEN}}$, $R_{DDR} = \frac{Q_{DDR}^{M}}{Q_{DDR}^{GEN}}$ берутся значения измерений времени работы алгоритма ранжирования источников и сортировки объектов воздействия (пункт 3.3.1).

Оценка ресурсопотребления соответствует заданной в требованиях, если все вышеперечисленные показатели соответствуют условию $r \leq R^{additional}$ [142]. Так как в ходе экспериментов используется отдельный компьютер для выполнения задачи анализа и сортировки, то $R^{additional} = 0.75$ (25% ресурсов выделяется на общие задачи операционной системы и сопутствующих

программ, 25 % времени работы оператора выделяется в течение одного рабочего дня длиной в 8 часов на прочие обязанности).

В процессе проведения экспериментов были получены следующие показатели:

1. В время работы программного прототипа ранжирования и оценки источников, было использовано на 50% только одно ядро центрального процессорного устройства (то есть показатель $R_{CP} = 0{,}125$).

2. Из-за особенностей экспериментального стенда и операционной системы (Windows 10, версия 20H2, 64 бит), в процессе выполнения методики использовалось не более 512 Мб оперативной памяти (то есть показатель $R_{DDR} = 0{,}09125$).

3. Экспериментальное время работы оператора по методике составило 73 минуты, то есть показатель R_{expert} = 0,152083)

Полученные значения показателей свидетельствуют о том, что $P_{\text{res}}\left(r \le R^{additional}\right)$=1, а следовательно требование определенное в первой главе диссертации $P_{\text{res}}\left(r \le R^{additional}\right) \ge P_{\text{res}}^{additional}$, где $P_{\text{res}}^{additional} = 0{,}99$ выполняется. Это позволяет сделать вывод, что оценка *ресурсопотребления* соответствует предъявляемым требованиям.

3.3.3 Теоретическая оценка методики противодействия вредоносной информации в социальных сетях

В данном подразделе диссертации приведены результаты оценки методики противодействия вредоносной информации с учетом требований к обоснованности.

3.3.3.1 Теоретическая оценка обоснованности

Было проведено исследование аналогов и определен набор параметров, которые учитываются в них при выборе объекта воздействия и контрмеры. Для каждого параметра вводится обозначение буквой русского алфавита,

предназначенное для удобства сведения показателя в таблицу, используемую для оценки.

В качестве основных аналогов выбраны следующие решения.

Патент RU2651252, полученный Лабораторией Касперского на способ ограничения пользователю к подозрительным объектам социальной сети. В основе решения лежит идея построения социального графа для профиля пользователя, источника и сообщения, и анализа их взаимосвязи [126].

Zerofox Inc «Brand Protection» [133]. Решение, ранее описанное в пунктах 1.2.3 и 3.2 диссертации. Выбор этого аналога для сравнения обоснован тем, что данной решение создает списки подозрительных источников и сообщений и в дальнейшем сверяет новые объекты социальной сети со списками. После проверки нового информационного объекта доступ может ограничен на стороне пользователя.

Методика, предложенная от имени компании Ithreat Cyber Group Inc «Устройства и методы повышения веб-безопасности и сдерживания киберзапугивания» [153]. Методика является частью системы родительского контроля. Система позволяет ограничивать доступ к вредоносной информации.

Методика от компании Creopoint, которая позволяет собирать информацию о бренде, за счет времени работ оператора оценивать истинность или ложность информации и затем выбирать объекты для сдерживания распространения дезинформации [138].

И способ противодействия вредоносной информации принятый на основании 149 ФЗ [28] с добавлением в систему «Единый реестр доменных имен, указателей страниц сайтов в сети «Интернет» и сетевых адресов, позволяющих идентифицировать сайты в сети «Интернет», содержащие информацию, распространение которой в Российской Федерации запрещено» ЕАИС [132].

На основании проведенного исследования был сформирован общий список параметров учитываемых при выборе объектов воздействия и контрмер для методики и ее аналогов:

1. А – возможность учета количества комментариев к сообщению.
2. Б – возможность учета количества лайков к сообщению.
3. В – возможность учета количества просмотров сообщения.
4. Г – возможность учета количества репостов сообщения.
5. Д – возможность учета индекса активности.
6. Ж – возможность учета индекса просматриваемости.
7. З – возможность учета индекса влиятельности.
8. И – возможность учета количества сообщений источника.
9. К – возможность учета приоритета объекта воздействия.
10. Л – возможность учета коэффициента сложности реализации контрмеры.
11. М – возможность учета метода реализации контрмеры.
12. Н – возможность учета типа контрмеры.

Составлена сравнительная таблица 3.6.

Таблица 3.6. − Сравнительная таблица параметров, учитываемых при выборе объекта воздействия и контрмеры для методики и аналогов

Сравнительная оценка обоснованности методики противодействия вредоносной информации в СС	Учитываемые параметры N_{param}												Оценка
	А	Б	В	Г	Д	Ж	З	И	К	Л	М	Н	
RU2651252, Лаборатория Касперского	0	0	1	1	0	0	0	1	0	0	0	0	3
Zerofox Inc «Brand Protection»	1	1	1	1	0	0	0	1	0	0	0	0	5
Ithreat Cyber Group Inc	1	1	1	1	0	0	0	0	0	0	0	0	4
Creopoint Inc	1	1	1	1	1	1	0	1	1	0		0	8
ЕАИС Роскомнадзора	1	1	1	1	0	0	0	0	1	0	1	1	7
Разработанная методика	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**12**

С учетом требований к обоснованности, сформулированных в пункте 1.3 диссертации, целевой функцией методики противодействия вредоносной

информации в социальных сетях с учетом требований к обоснованности является максимизация количества учитываемых параметров $N_{param} \rightarrow max$, для выбираемых объектов воздействия и контрмер в ходе противодействия вредоносной информации в социальных сетях.

В сравнении с аналогами количество учитываемых параметров при использовании методики больше, такое что $N_{param}^{M} > maxN_{param}^{s}$, где N_{param}^{M} – количество учитываемых параметров для методики, $maxN_{param}^{s}$ – максимальное количество учитываемых параметров для аналогов. При этом $N_{param}^{M} = 12$, $maxN_{param}^{s} = 8$.

В результате теоретической оценки можно сделать вывод об увеличении количества учитываемых параметров, а значит и об росте обоснованности принятия решения о противодействии.

3.3.4 Сравнительная оценка функциональных характеристик с аналогами

Для сравнения функциональных характеристик методики с аналогами были взяты те, же аналоги, что и в пункте 3.3.4 диссертации.

Согласно требованиям, определенным в пункте 1.3 диссертации функциональные требования, представляют собой перечень функций, которые должна выполнять система при использовании методики противодействия вредоносной информации в социальных сетях.

Для каждого функционального требования вводится обозначение буквой русского алфавита, предназначенное для удобства сведения показателя в таблицу, используемую для оценки.

1. А – возможность формирования задачи на сбор сообщений, дополнительных данных и анализ сообщений для системы мониторинга.

2. Б – возможность настройки доступных мер противодействия в системе.

3. В– возможность анализа источников сообщений в полученном наборе данных.

4. Г – возможность ранжирования и сортировки объектов воздействия в полученном наборе данных.

5. Д – возможность ранжирования и сортировки доступных контрмер из базы контрмер для каждого набора данных.

6. Ж – возможность выбора цели воздействия для противодействия.

7. З – генерация отчетов о полученных результатах в виде, адаптированном для эксперта по информационной безопасности

8. И – генерация отчетов о работе системы в виде, адаптированном для администратора системы.

9. К – учет специфики режима работы системы (настройка, эксплуатация).

В процессе сравнительной оценки функциональных характеристик использовалась балльную оценку методик: «–» – 0 баллов, «+/–» – 0.5 балла, «+» – 1 балл.

Результаты сравнения функциональных характеристик методики с аналогами внесены в таблицу 3.7.

Таблица 3.7. – Сравнение функциональных характеристик методики с аналогами

Методика противодействия вредоносной информации в СС	**Учитываемые параметры**						**Оценка**
	А	**Б**	**В**	**Г**	**Д**	**Ж**	
RU2651252, Лаборатория Касперского	–	–	+	–	+/–	+/–	2
Zerofox Inc «Brand Protection»	+	+	+	–	+/–	+/–	4
Ithreat Cyber Group Inc	–	+	–	–	+/–	–	1.5
Creopoint Inc	+	+	+	+/–	+/–	+/–	4.5
ЕАИС Роскомнадзора	+	+	–	–	+/–	+/–	3
Разработанная методика	**+**	**+**	**+**	**+**	**+**	**+**	**6**

Анализ результатов сравнения существующих аналогов противодействия вредоносной информации в социальных сетях и предложенной методики позволяет сделать следующие выводы. Во-первых, ни одна из методик, кроме предложенной, одновременно не удовлетворяет

всем функциональным требованиям. Во-вторых, все методики в той или иной степени позволяют ранжировать контрмеры. В-третьих, параметры сообщений, источников, контрмер учитываются только в предложенной методике и в решении от компании Creopoint Inc. В-четвертых, отставание ближайших аналогов от предложенной методики составляет от 1,5 балла до 4-х. Таким образом, предложенная методика выигрывает у ближайших аналогов.

3.4 Предложения по практическому использованию результатов исследования

Разработанные в диссертационной работе модели, алгоритмы, методика и архитектура могут быть использованы для следующих решений.

Во-первых, для повышения информационной безопасности государства, общества и личности в социальных сетях за счет обоснованного выбора объектов воздействия для мер противодействия. Если оператор действующей системы противодействия в государстве будет получать в первую очередь информацию об объектах воздействия с высоким приоритетом, на который влияет количество просмотров, активность аудитории, общее количество сообщений с вредоносной информацией на странице в социальной сети, тогда он сможет принимать решение о противодействии таким объектам незамедлительно. А противодействие сообщениям, которые никто не видит и не читает в социальных сетях будет осуществляться в последнюю очередь. Это позволит перераспределить внимание оператора и повысит качество принимает решений за счет новых, учитываемых параметров. В экстренных ситуациях, связанных с противодействием экстремизму и терроризму, система может быть настроена таким образом, чтобы противодействие запускалось в автоматическом, а не автоматизированном режиме.

Потенциально перспективным направлением использования результатов исследования может быть анализ источников вредоносных сообщений в социальных сетях, содержащих призывы к суициду детей

и подростков. Это позволит на уровне муниципальных и городских администраций выявлять самых активные страницы в социальных сетях и сосредоточиваться на защите участников и подписчиков таких сообществ от необдуманных поступков. Этот же подход возможно было бы использовать на уровне администрации города и районов для поиска самых активных источников распространения наркотиков через социальные сети.

Во-вторых, предложенные модели, алгоритмы, методика и архитектура могут использоваться в госкорпорациях и коммерческих организациях для защиты репутации и бренда. Список информационных угроз для такой задачи может включать негативные отзывы, а последующий анализ источников и их сортировка позволят сфокусировать усилия организации на нивелирование негативного имиджа. Также модели, алгоритмы и методика могут использоваться компаниями для защиты от утечки конфиденциальной информации в социальных сетях или от распространения информации с нарушением авторского права.

В-третьих, предложенные алгоритмы и методика противодействия могут быть использованы для совершенствования следующих решений: систем родительского контроля, антивирусов. Анализ источников и сортировка объектов воздействия позволит выделять объекты воздействия и сохранять сведения о них в системах для последующей проверки новых объектов на предмет связи с ними. Предполагается, что это позволить принимать решение об ограничении доступа пользователю к объекту без анализа контента. Также предложенные алгоритмы могут стать частью крупной SIEM- системы, осуществляющей анализ источников в социальных сетях и на основании полученных данных формирующей ограничительные списки для сотрудников организации.

3.5 Вывод по главе 3

1. Предложена методика противодействия вредоносной информации в социальных сетях, которая основывается на использовании моделей,

алгоритмов и отличается от аналогов тем, что обеспечивает анализ информации; формирование списков объектов воздействия для противодействия им, их последующую сортировку; предоставление оператору выбранного и альтернативных вариантов с обоснованием выбора. А также обеспечивает ранжирование контрмер доступных в системе для противодействия вредоносной информации в социальных сетях.

2. Предложены архитектура и программные прототипы компонентов системы, которые отличаются от существующих тем, что ориентированы на ранжирование и выбор доступных контрмер. Архитектура содержит оригинальные компоненты анализа и оценки источника вредоносной информации, базу данных с информацией о мерах противодействия вредоносной информации в СС, информацию об агентах реализации, через которые контрмеры будут реализованы. В силу этого архитектура позволяет формировать наборы исходных данных для исследований и разработок в области противодействия вредоносной информации, а также для исследований и разработок решений для систем поддержки принятия решения.

3. Проведена теоретическая и экспериментальная оценка разработанной методики и прототипов. В качестве критериев достижения цели исследования были выбраны такие свойства как оперативность, ресурсопотребление и обоснованность. Результаты оценки подтвердили достижение заявленных требований.

4. Предложены варианты применения разработанных в ходе исследования моделей, алгоритмов, методики и архитектуры противодействия вредоносной информации в социальных сетях.

ЗАКЛЮЧЕНИЕ

В диссертационной работе в целях повышения эффективности противодействия вредоносной информации в социальных сетях решена задача разработки модельно-методического аппарата для повышения обоснованности принимаемого решения о противодействии за счет увеличения количества учитываемых параметров при выборе объекта воздействия и контрмеры. Что в свою очередь достигается за счет анализа источников вредоносной информации и ранжирования контрмер. Получены следующие научные результаты, составляющие **итоги** исследования:

1) проведен анализ существующих моделей вредоносной информации и информационного обмена;

2) проведен анализ существующих алгоритмов оценки источников в СС, существующих систем мониторинга и методик противодействия вредоносной информации в СС;

3) разработан комплекс моделей социальной сети, источника и вредоносной информации;

4) разработан комплекс алгоритмов анализа источников и ранжирования контрмер;

5) разработана методика противодействия вредоносной информации в социальных сетях;

6) разработана архитектура и программные прототипы компонентов системы противодействия вредоносной информации. Проведена экспериментальная и теоретическая оценка предложенных моделей, алгоритмов, методики и архитектуры.

Все результаты, выносимые на защиту, являются новыми. Разработан комплекс моделей социальной сети, источника и вредоносной информации, отличающийся от существующих наличием новых элементов, атрибутов, связей между ними, и характеризующий объекты в социальных сетях. Также предложен комплекс алгоритмов анализа источников вредоносной

информации и ранжирования контрмер, который отличается от аналогов учетом связей и зависимых атрибутов объектов в социальной сети. В качестве результата работы алгоритмы анализа источников формируют сортированный список объектов воздействия. В комплекс входит алгоритм ранжирования контрмер, учитывающий коэффициенты и уровни сложности для каждой контрмеры. Разработана методика противодействия вредоносной информации в социальных сетях, ориентированная на автоматический и автоматизированный выбор объектов воздействия и мер противодействия вредоносной информации из списка ранжированных контрмер и поддержку принятия решения о противодействии. Предложена архитектура и программные прототипы компонентов системы противодействия вредоносной информации, которая ранжирует контрмер, содержит оригинальные компоненты анализа и оценки источника вредоносной информации, базу данных с информацией о мерах противодействия, информацию об агентах реализации, через которые контрмеры будут реализованы.

Сформулированы **рекомендации** по применению результатов работы для информационной безопасности государства, общества и личности в социальных сетях, а также для защиты интересов организации. Так, например, методика может использоваться в государственных и городских ситуационных центрах для противодействия экстремизму и терроризму, предотвращению распространения фейковых новостей, информации о способах суицида или призывов к нему. Развитие алгоритмов позволит их использовать в научных исследованиях, также алгоритмы, методика и архитектура могут быть использованы для усовершенствования систем родительского контроля, систем управления репутацией бренда.

В качестве **перспектив дальнейшей разработки тематики** можно выделить следующие. Прежде всего – расширение учитываемых атрибутов в алгоритмах. Например, могут одновременно обнаруживаться признаки искусственной активности, анализироваться характристики авторов

сообщений, дискретные признаки появления сообщений (время публикации, скорость публикации сообщений от одного автора или связанных сообщений на разных страницах). А также – усовершенствование методики для анализа провокаций и вбросов большого количества сообщений. Возможным направлением исследованием является разработка единого центра противодействия с открытыми интерфейсами взаимодействия с агентами реализации и с социальными сетями для защиты интересов общества и личности в информационной сфере государства.

Положения, выносимые на защиту, **соотнесены с пунктом 3 паспорта специальности** 05.13.19 – «Методы и системы защиты информации, информационная безопасность»: «Методы, модели и средства выявления, идентификации и классификации угроз нарушения информационной безопасности объектов различного вида и класса» и **соотнесены с пунктом 5 паспорта специальности** 05.13.19 – «Методы и системы защиты информации, информационная безопасность»: «Методы и средства (комплексы средств) информационного противодействия угрозам нарушения информационной безопасности в открытых компьютерных сетях, включая Интернет».

СПИСОК ЛИТЕРАТУРЫ

1. Виткова Л.А. Модель вредоносной информации и ее распространителя в социальных сетях / Л.А. Виткова, Д.В. Сахаров, Д.Р. Голузина // Защита информации. Инсайд. – Спб., 2020. –№3 (93). – С. 66-72.

2. Гамидов Т.О. Разработка моделей и алгоритмов анализа данных для исследования хода инцидентов и кризисов в социальных сетях / Т.О. Гамидов, Л.А. Виткова, М.М. Ковцур // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. – СПб., 2020. – № 2. – С. 3-10.

3. Виткова Л.А. Выбор мер противодействия вредоносной информации в социальных сетях / Л.А. Виткова, А.А. Чечулин, Д.В. Сахаров // Вестник Воронежского института ФСИН России. – Воронеж, 2020. – Т. 3. – С. 20-29.

4. Виткова Л.А. Архитектура системы выявления и противодействия нежелательной информации в социальных сетях. / Л.А. Виткова, И.Б. Саенко // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. – СПб., 2020. – № 3. – С. 33-39.

5. Виткова Л.А. Методика анализа аудитории канала распространения информации в социальных сетях. // Известия высших учебных заведений. Технология легкой промышленности. – СПб, 2018. – Т. 42, № 4. – С. 5-10.

6. Проноза А.А. Методика выявления канала распространения информации в социальных сетях / А.А. Проноза, Л.А. Виткова, А.А. Чечулин, И. В. Котенко, Д.В. Сахаров // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. – СПб., 2018. – Т. 14, № 4. – С.362-377

7. Kotenko I.V. The intelligent system for detection and counteraction of malicious and inappropriate information on the Internet / I.V. Kotenko, L.A.

Vitkova, I.B. Saenko, O.N. Tushkanova, A.A. Branitsky// AI Communications, 2020. – Vol 33(1). – C. 1-13. – ISSN 0921-7126

8. Vitkova L.A. Selection of countermeasures against propagation of harmful information via Internet / L. A. Vitkova, A. P. Pronichev, E. V. Doynikova, I. B. Saenko // IOP Conference Series: Materials Science and Engineering, 2021 Vol 1032, – 1032 012017. – ISSN 1757-8981

9. Vitkova, L.A. The technology of intelligent analytical processing of digital network objects for detection and counteraction of inappropriate information / L.A. Vitkova, I.B. Saenko, A.A. Chechulin, I.B. Parashchuk // The 1st International Conference on Computer Technology Innovations dedicated to the 100th anniversary of the Gorky House of Scientists of Russian Academy of Science (ICCTI – 2020). Official conference proceedings, 2020. – P 13-19. – ISBN 978-5-9676-1216-9

10. Vitkova L.A. Approach to Identification and Analysis of Information Sources in Social Networks / L. A. Vitkova, M. V. Kolomeets // Proceedings of the 13th International Symposium on Intelligent Distributed Computing (IDC 2019), October 7-9, 2019, Saint-Petersburg, Russia. 2020. P. 285-293. – ISSN 1860-949X.

11. Vitkova L.A. An Approach to Creating an Intelligent System for Detecting and Countering Inappropriate Information on the Internet / L.A. Vitkova, I.B. Saenko, O.N. Tushkanova // Proceedings of the 13th International Symposium on Intelligent Distributed Computing (IDC 2019), October 7-9, 2019, Saint-Petersburg, Russia. 2020. – P. 244-254. – ISSN 1860-949X.

12. Vitkova, L.A. Hybrid Approach for Bots Detection in Social Networks Based on Topological, Textual and Statistical Features / L.A. Vitkova, Kotenko I.V., M.V. Kolomeets, O.N. Tushkanova, A.A. Chechulin // Advances in Intelligent Systems and Computing 1156 AISC, 2019, P. 412-421

13. Pronoza A.A. Visual analysis of information dissemination channels in social network for protection against inappropriate content / A.A. Pronoza, L.A. Vitkova, A.A. Chechulin, I.V. Kotenko // 3rd International Scientific Conference on Intelligent Information Technologies for Industry, IITI 2018. Sochi, Russian

Federation, 17-21 September 2018. Advances in Intelligent Systems and Computing. Vol. 875, 2019. P. 95-105.

14. Kotenko I.V. Monitoring and counteraction to malicious influences in the information space of social networks / I.V. Kotenko, I.B. Saenko, A.A. Chechulin, V.A. Desnitsky, L.A. Vitkova, A.A. Pronoza // The 10th Social Informatics conference (SocInfo2018). September 25–28, 2018, Saint Petersburg, Russia. Proceedings, Part II. Lecture Notes in Computer Science, Vol.11186, Springer 2018, P.1 59-167. – ISBN 978-3-030-01158-1.

15. Виткова Л.А. Компонент сегментации пользователей по их активности в социальных сетях / Л.А. Виткова, А.А. Чечулин, И.В. Котенко – Свидетельство о государственной регистрации программы для ЭВМ № 2019664733. Зарегистрировано в Реестре программ для ЭВМ 13.11.2019.

16. Федорченко Е.В. Компонент выбора мер противодействия нежелательной, сомнительной и вредоносной информации / Е.В. Федорченко, Л.А. Виткова, А.П. Проничев, И.Б. Саенко. – Свидетельство о государственной регистрации программы для ЭВМ № 2020665591. Зарегистрировано в Реестре программ для ЭВМ 27.11.2020.

17. Виткова Л.А. База данных для учета нежелательной информации совместно с мерами противодействия / Л.А. Виткова, Е.О. Березина, А.П. Проничев, И.Б. Саенко, И.В. Котенко – Свидетельство о государственной регистрации программы для ЭВМ № 2020622557. Зарегистрировано в Реестре программ для ЭВМ 08.12.2020.

18. Акофф Р.Л. Планирование будущего корпорации / М.: Сирин, 2002 – 256 с.

19. ГОСТ Р 53647.9-2013 Менеджмент непрерывности бизнеса. Управление организацией в условиях кризиса, М.: Стандартинформ, 2014. – 70 с.

20. Информация. Что такое информация / Большой Энциклопедический словарь (БЭС) // Словопедия: Словари. – URL: http://www.slovopedia.com/2/200/228368.html (дата обращения: 26.01.2021)

21. Доктрина Информационной безопасности. Указ Президента РФ от 5 декабря 2016 г. N 646 «Об утверждении Доктрины информационной безопаснсти РФ». – URL: https://demo.garant.ru/#/document/71556224/paragraph/1/doclist/1042/showentries/0/highlight/доктрина информационной безопасности:0 (дата обращения: 26.01.2021).

22. Морозов И.Л. Политический экстремизм: учебное пособие / И. Л. Морозов; Федеральное агентство по образованию и науке, Фил. Гос. образовательного учреждения высш. проф. образования «Московский энергетический ин-т (технический ун-т)» в г. Волжском, Каф. «Социально-гуманитарные науки». – Волжский: Филиал ГОУ ВПО "МЭИ(ТУ)", 2008. - 122 с.

23. Некрасова Н.В. Информационный аспект экстремизма и терроризма и деструктивные тенденции в СМИ / Н.В.Некрасова // Вестник Российского университета дружбы народов. Серия: Социология. 2013. – № 1. – URL: https://cyberleninka.ru/article/n/informatsionnyy-aspekt-ekstremizma-i-terrorizma-i-destruktivnye-tendentsii-v-smi (дата обращения: 26.01.2021).

24. Макаренко С.И., Чукляев И.И. Терминологический базис в области информационного противоборства / С.И Макаренко, И.И. Чукляев // Вопросы кибербезопасности. 2014. – № 1 (2). – С. 13–21.

25. Минькович Т.В. Информационные технологии : понятийно - терминологический аспект / Т.В. Минькович // ОТО. 2012. – Т. 2. – С. 371–389.

26. Приказ Роскомнадзора N 84, МВД России N 292, Роспотребнадзора N 351, ФНС России ММВ-7-2/461@ от 18.05.2017 «Об утверждении Критериев оценки материалов и (или) информации, необходимых для принятия решений Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций, Министерством внутренних дел Российской Федерации, Федеральной службой по надзору в сфере защиты прав потребителей и благополучия человека, федеральной

налоговой службой о включении доменных имен и (или) указателей страниц сайтов в информационно-телекоммуникационной сети «Интернет», а также сетевых адресов, позволяющих идентифицировать сайты в сети «Интернет», содержащие запрещенную информацию, в единую автоматизированную информационную систему «Единый реестр доменных имен, указателей страни…». – URL: http://www.consultant.ru/document/cons_doc_LAW_218948/ (дата обращения: 16.08.2019)

27. Экстремистские материалы // Министерство юстиции Российской Федерации: офиц. Сайт – URL: https://minjust.gov.ru/ru/extremist-materials/ (дата обращения 27.01.2021)

28. Федеральный закон от 27 июля 2006 г. N 149-ФЗ «Об информации, информационных технологиях и о защите информации». – URL: http://ivo.garant.ru/#/document/12148555/paragraph/3471:0 (дата обращения: 26.01.2021). Доступ из системы ГАРАНТ

29. Федеральный закон от 29 декабря 2010 г. N 436-ФЗ «О защите детей от информации, причиняющей вред их здоровью и развитию». – URL: http://ivo.garant.ru/#/document/12181695/paragraph/1:0 (дата обращения: 26.01.2021). Доступ из системы ГАРАНТ.

30. Международный пакт о гражданских и политических правах (Нью-Йорк, 16 декабря 1966 г.). – URL: https://demo.garant.ru/#/document/2540295/paragraph/270/doclist/1036/showentries/0/highlight/О гражданских и политических правах:2 (дата обарщения: 26.01.2021). Доступ из системы ГАРАНТ.

31. Рекомендация Комитета министров Совета Европы N R (89) 7 государствам-членам «О принципах распространения видеограмм, содержащих насилие, жестокость или порнографию» (принята Комитетом министров 22.04.1989 на 425-ом заседании Представителей министров) – URL: https://base.garant.ru/2562858/ (дата обращения: 27.01.2021). Доступ из системы ГАРАНТ.

32. Международная конвенция о ликвидации всех форм расовой дискриминации /Конвенции и соглашения// Декларации, конвенции, соглашения и другие правовые материалы – URL: https://www.un.org/ru/documents/decl_conv/conventions/raceconv.shtml (дата обращения: 27.01.2021).

33. Ferrara E., Varol O., Davis C., Menczer F., Flammini A. The Rise of Social Bots // Communications of the ACM. 2016. V. 59, № 7. P. 96–104.

34 Губанов Д. А. Программа для расчета влияния и влиятельности пользователей социальных сетей на основе акциональной модели. Свидетельство о государственной регистрации программы для ЭВМ № 2019665357 РФ. Зарегистрировано в Реестре программ для ЭВМ 22.11.2019.

35. Varol O., Ferrara E., Davis A.C., Menczer F., Flammini A. Online Human Bot Interaction // Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017). 2017. P. 280–289.

36. Marwick A., Lewis R. Media manipulation and disinformation online // Data & Society Research Institute. 2017. P 1–104.

37. Classmatess: сайт. – URL: https://www.Classmates.com/. (дата обращения: 27.01.2021).

38. Livejournal: сайт. – URL: https://www.livejournal.com/. (дата обращения: 27.01.2021).

39. MySpace: сайт. – URL: https://myspace.com/. (дата обращения: 27.01.2021).

40. Facebook: сайт. – URL: Facebook.com. (дата обращения: 27.01.2021).

41. Twitter: сайт. – URL: twitter.com. (дата обращения: 27.01.2021).

42. Google Академия: сайт. – URL: https://scholar.google.ru/. (дата обращения: 27.01.2021).

43. Wasserman S., Galaskiewicz J. Advances in social network analysis: Research in the social and behavioral sciences. 1994. 300 p.

44. Cook K.S., Whitmeyer J.M. Two Approaches to Social Structure: Exchange Theory and Network Analysis // Annual Review of Sociology. Annual Reviews, 1992. Vol. 18, № 1. P. 109–127.

45. Otte E., Rousseau R. Social network analysis: A powerful strategy, also for the information sciences // Journal of Information Science. 2002. Vol. 28, № 6.

46. Dang, L., Chen, Z., Lee, J., Tsou, M. H., Ye X. Simulating the spatial diffusion of memes on social media networks // International Journal of Geographical Information Science. 2019. Vol. 33, № 8. P. 1545–1568.

47. Grandjean M. Analisi e visualizzazioni delle reti in storia. L'esempio della cooperazione intellettuale della Società delle Nazioni // Memoria e ricercar. 2017. Vol. 55, № 2. P. 371–393.

48. Brennecke J., Rank O. The firm's knowledge network and the transfer of advice among corporate inventors–A multilevel network study //Research Policy. – 2017. – Т. 46. – №. 4. – С. 768-783.

49. Communication theory // Oxford Reference. – URL: https://www.oxfordreference.com/view/10.1093/oi/authority.20110810104639648 (дата обращения: 28.01.2021).

50. Гавра. Д. П. Основы теории коммуникации: учебное пособие / Д.П. Гавра. – СПб.: Питер, 2011. – 288 с.

51. C.E. Shannon. A Mathematical Theory of Communication // The Bell System Technical Journal. 1948. Vol. XXVII, № 3. P. 379–423.

52. Westley B.H., MacLean M.S. A Conceptual Model for Communications Research // Journal. Q. SAGE Publications, 1957. Vol. 34, № 1. P. 31–38.

53. Shelke S., Attar V. Source detection of rumor in social network – A review // Online Social Networks and Media. Elsevier B.V., 2019. Vol. 9. P. 30–42.

54. Luo W., Tay W.P., Leng M. How to Identify an Infection Source with Limited Observations. – URL: // ieeexplore.ieee.org. (дата обращения: 28.01.2021).

55. Wang Zh.; Zhang W; Wei Tan Ch. On inferring rumor source for SIS model under multiple observations // International Conference on Digital Signal Processing, DSP (2015). 2015. P. 1543-8675.

56. Kimura M., Motoda H., Saito K. Discovering Influential Nodes for SIS Models in Social Networks // Springer. 2009. Vol. 5808 LNAI. P. 302–316.

57. Meel P., Vishwakarma K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities // Expert Systems with Applications. 2020. Vol. 153. P. 112-986.

58. Song L.P., Jin Z., Sun G.Q. Modeling and analyzing of botnet interactions // Physica A: Statistical Mechanics and its Applications. 2010. Vol. 390, № 2. P. 347–358. doi:10.1016/j.physa.2010.10.001.

59. Kumari A., Singh S.N. Online influence maximization using rapid continuous time independent cascade model // Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering. 2017. P. 356–361.

60. Cheng J., Adamic L.A., Dow P.A., Kleinberg J., Leskovec J. Can cascades be predicted? // WWW 2014 - Proceedings of the 23rd International Conference on World Wide Web. 2014. – URL: https://dl.acm.org/doi/proceedings/10.1145/2566486 (дата обращения: 28.01.2021).

61. Yu X., Chu T. Learning the structure of influence diffusion in the independent cascade model // Chinese Control Conference, CCC. 2017. P. 5647-5651. doi:10.23919/chicc.2017.8028255

62. Kempe D., Kleinberg J., Tardos É. Maximizing the spread of influence through a social network // Journal of Chemical Theory and Computation. 2003. Vol. 11. P. 105-147. doi: 10.4086/toc.2015.v011a004

63. Wojtasik K. How and Why Do Terrorist Organizations Use the Internet? // Polish Political Science Yearbook. 2018. Vol. 46, № 2. P. 105–117.

64. Shao X., Ni X. Measurement of cyber communication behavior based on lasswell's paradigm // ACM International Conference Proceeding Series. New York, New York, USA: Association for Computing Machinery, 2019. P. 1–5.

65. Lipschiltz H. J. A Social Network Analysis of Ferguson and Its Progeny // Africana Race and Communication: A Social Study of Film, Communication, and Social Media / ed. Jr. J.L.C. Lexington Books, 2017. P. 19–37.

66. Fiske J. Introduction to Communication Studies // Google Книги. – URL: https://books.google.ru/books/about/Introduction_to_Communication_Studies.html?id=J3XzYCuDLNYC&redir_esc=y (дата обращения: 15.03.2021).

67. Yin L., Haoyang Zh., Tian B.; Yong D. An evidential link prediction method and link predictability based on Shannon entropy // Physica A: Statistical Mechanics and its Applications, 2017. Vol. 482. P. 699–712.

68. Sander W. Measures of Information // Handbook of Measure Theory. Elsevier, 2002. P. 1523–1565.

69. Koohikamali M., Sidorova A. Information re-sharing on social network sites in the age of fake news // Informing Sci. 2017. Vol. 20. P. 215–235.

70. Mirsarraf M., Shairi H., Ahmadpanah A. Social semiotic aspects of instagram social network // Proceedings – 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2017. Institute of Electrical and Electronics Engineers Inc., 2017. P. 460–465.

71. Newcomb T.M. Social psychology: the study of human interaction. 1903 – Free Download, Borrow, and Streaming : Internet Archive. – URL: https://archive.org/details/socialpsychology0000newc_z7j5 (дата обращения: 15.03.2021).

72. Guo Y.H. Liu L., Wu Y., Hardy J. Interest-aware content discovery in peer-to-peer social networks // ACM Trans. Internet Technol. Association for Computing Machinery, 2018. Vol. 18, № 3. P. 1–21.

73. Small M.L., Perry B.L., Pescosolido B., SmithIntroduction E.(N).: The Past and Future of Ego-Centric Network Analysis. – URL:

https://scholar.harvard.edu/files/mariosmall/files/small_etal_pastandpresentegonetworks.pdf (дата обращения: 15.03.2021).

74. Охапкин В.П., Охапкина Е.П., Исхакова А.О., Исхаков А.Ю. Деструктивное информационно-психологическое воздействие в социальных сетях / В.П. Охапкин, Е.П. Охапкина, А.О. Исхакова, А.Ю. Исхаков // Моделирование, оптимизация и информационные технологии. – Т. 8, № 1.– С. 1–14.

75. Андреева Г.М. Социальная психология: учебник для студентов высших учебных заведений, обучающихся по направлению с специальности ‘Психология’. 5-е изд., / Г.М. Андреева. – М.: Аспект Пресс, 2009. – 362 с.

76. Mican D., Sitar-Tăut D.A., Mihuţ I.S. User behavior on online social networks: Relationships among social activities and satisfaction // Symmetry (Basel). 2020. Vol. 12, № 10. P.2-16. doi:10.3390/sym12101656

77. Haug M. Mass Communication on Social Media // Proceedings of the 2020 on Computers and People Research Conference. New York, NY, USA: Association for Computing Machinery (ACM), 2020. P. 169–169.

78. Minnseok Ch., Chong Un P., Jiyoung W. Epidemic Modeling of Sentiment Diffusion on Web Forums // Advanced Science Letters. 2017. Vol. 23, № 10. P. 10477-10480(4).

79. Chen T., Shi J., Yang J., Cong G., Li G. Modeling Public Opinion Polarization in Group Behavior by Integrating SIRS-Based Information Diffusion Process // Complexity. 2020. P. 1076-2787 doi:10.1155/2020/4791527

80. Jager W. Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Based Simulation Model of Attitude Change // Computational & Mathematical Organization Theory. 2004. Vol. 10. P. 295–303

81. Albert R., Barabási A.L. Statistical mechanics of complex networks // Reviews of modern physics, 2002. Vol. 74, № 1. P. 47–97.

82. Губанов Д.А., Новиков Д.А., Чхатршвили А.Г. Модели репутации и информационного управления в социальных сетях/ Д.А. Губанов, Д.А.

Новиков, А.Г. Чхартшвили // Управление большими системами: сборник трудов. 2009. – Т. 26, № 1. –С. 209–234.

83. Rogers E.M. Diffusion of Innovations, 5th Edition // Free Press. Amazon.com: Books [Electronic resource] // Free Press. 2003.

84. Forest J. Influence warfare: how terrorists and governments fight to shape perceptions in a war for ideas // Choice Reviews Online. 2009. P. 392.

85. Bass F.M. A new product growth for model consumer durables // Management Science. 2004. Vol. 50, № 12 SUPPL. P. 1825–1832.

86. Susarla A., Oh J.H., Tan Y. Social networks and the diffusion of user-generated content: Evidence from youtube // Information Systems Research. 2012. Vol. 23, № 1. P. 23–41.

87. Granovetter M. Threshold Models of Collective Behavior // Am. J. Sociol. 1978. Vol. 83, № 6. P. 1420–1443.

88. Ran Y., Deng X., Wang X., Jia T. A generalized linear threshold model for an improved description of the spreading dynamics // Chaos: An Interdisciplinary Journal of Nonlinear Science, – 2020. – Т. 30. – №. 8. – С. 083127.

89. Михайлов А. П., Маревцева Н. А. Модели информационной борьбы //Математическое моделирование. – 2011. – Т. 23. – №. 10. – С. 19-32.

90. Михайлов А.П., Измоденова К.В. Об оптимальном управлении процессом распространения информации // Математическое моделирование. – 2005. –Т. 17, № 5. – С. 67–76.

91. Марцева Н.А. Простейшие математические модели информационного противоборства // Математическое моделирование социальных процессов. – 2010. – Т 11. – С. 59–72.

92. Peng S., Zhou Y., Cao L., Yu S., Niu J., Jua W. Influence analysis in social networks: A survey //Journal of Network and Computer Applications. – 2018. – Т. 106. – С. 17-32.

93. Choi J., Shin J., Yi Y. Information source localization with protector diffusion in networks // Journal Communications Networks. 2019. Vol. 21, № 2. P. 136–147.

94. Chai Y., Wang Y., Zhu L. Information Sources Estimation in Time-Varying Networks // IEEE Trans. Inf. Forensics Secur. Institute of Electrical and Electronics Engineers Inc., 2021. P. 2621–2636.

95. George B., Shekhar S. Time Aggregated Graphs // Encyclopedia of Database Systems. Springer New York, 2016. P. 1–2.

96. Askarizadeh M., Tork Ladani B. Soft rumor control in social networks: Modeling and analysis // Engineering Applications of Artificial Intelligence, 2021. Vol. 100. P. 1–12.

97. Hosni A.I.E., Li K. Minimizing the influence of rumors during breaking news events in online social networks // Knowledge-Based System. 2020. Vol. 193. P. 105-452.

98. Bastos M.T., Mercea D. The Brexit Botnet and User-Generated Hyperpartisan News // Elsevier. SAGE Publications Inc., 2019. Vol. 37, № 1. P. 38–54.

99. Shin J., Jian L., Driscoll K., Bar F. The diffusion of misinformation on social media: Temporal pattern, message, and source // Computers in Human Behavior. 2018. Vol. 83. P. 278–287.

100. Alizadeh M., Shapiro J.N., Buntain C., Tucker J.A. Content-based features predict social media influence operations // Science Advances. 2020. Vol. 6, № 30. P. eabb5824.

101. Yuan D. Sun. H. Reverse Intervention for Dealing with Malicious Information in Online Social Networks // Computing and Informatics - formerly Computers and Artificial Intelligence. 2020. Vol. 39, № 1. P. 156–173.

102. Расторгуев С.П. Литвиненко М.В. Информационные войны в сети Интернет / под ред. Михайловского А.Б. – М.: АНО «Центр стратегических оценок и прогнозов», 2014. – 128 с.

103. Губанов Д.А., Петров И.В., Чхартишвили А.Г. Многомерная модель динамики мнений в социальных сетях: индексы поляризации // Проблемы управления. 2020. – Т. 3. – С. 26–33.

104. Kotenko I., Chechulin A., Komashinsky D. Categorisation of web pages for protection against inappropriate content in the internet // International Journal of Internet Protocol Technology. 2017. Vol. 10, № 1. P. 61–71.

105. Kotenko I., Checulin A., Shorov A., Komashinsky D. Analysis and evaluation of web pages classification techniques for inappropriate content blocking // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2014. Vol. 8557 LNAI.

106. Kotenko I., Saenko I., Chechulin A., Desnitsky V., Vitkova L., Pronoza A. Monitoring and counteraction to malicious influences in the information space of social networks // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, 2018. Vol. 11186 LNCS. P. 159–167.

107. Руководящий документ. Утвержден Приказом Ростехрегулирования от 06.04.2005 № 77-ст. «Рекомендации по стандартизации Р 50.1.053-2005. Информационные технологии. Основные термины и определения в области технической защиты информации». 2005.

108. ГОСТ Р 50922-2006 Защита информации. Основные термины и определения», М.: Стандартинформ, 2006.

109. Iqbuzz. Сервис мониторинга социальных медиа и онлайн-СМИ: сайт. – URL: https://www.iqbuzz.pro (дата обращения: 22.03.2021).

110. YouScan – платформа для аналитики соцмедиа: сайт. – URL: https://youscan.io/(дата обращения: 22.03.2021).

111. ДиалогНаука. Система «Лавина Пульс» для раннего предупреждения об утечках конфиденциальной информации: сайт. – URL: https://www.dialognauka.ru/solutions/sistema_lavina/ (дата обращения: 22.03.2021).

112. Brand Analytics. Система мониторинга и анализа социальных медиа и СМИ: сайт. – URL: https://br-analytics.ru (дата обращения: 22.03.2021).

113. ПрессИндекс. Мониторинг СМИ и социальных сетей в режиме реального времени: сайт. – URL: https://pressindex.ru/ (дата обращения: 22.03.2021).

114. 15 of the Best Social Media Monitoring Tools to Save You Time. Newberry C. – URL: https://blog.hootsuite.com/social-media-monitoring-tools/ (дата обращения: 22.03.2021).

115. ООО «БалтИнфоКом». Многопользовательская система анализа и визуализации данных в графовом виде «Октопус». Свидетельство о государственной регистрации программы для ЭВМ № 2016615916 от 15.04.2016г.

116. SemanticForce. Система мониторинга и анализа онлайн-медиа в режиме реального времени: сайт. – URL: https://www.semanticforce.net/ru/ (дата обращения: 22.03.2021)

117. Adblock Plus: сайт. – URL: https://adblockplus.org/en/about (дата обарщения: 22.03.2021).

118. Вопросы о расширении Антишок – URL: https://yandex.ru/support/browser/faq/faq-antishock.html (дата обращения: 22.03.2021).

119. The Interactive Advertising Bureau (IAB) Russia: сайт. – URL: https://iabrus.ru/ (дата обарщения: 22.03.2021).

120. Расширение Родительский контроль – Блокировка порно сайтов. – URL: https://chrome.google.com/webstore/detail/parental-control-adult-bl/peocghcbolghcodidjgkndgahnlaecfl?hl=ru (дата обращения: 22.03.2021).

121. Антикремлебот – Подсветка ботов. – URL: https://github.com/civsocit/gosvon (дата обращения: 22.03.2021).

122. MetaBot – подсветка ботов в YouTube. – URL: https://github.com/YTObserver/YT-ACC-DB/ (дата обращения: 22.03.2021).

123. Расширение Site blocker. – URL: https://chrome.google.com/webstore/detail/site-blocker/offfjidagceabmodhpcngpemnnlojnhn (дата обращения: 22.03.2021).

124. Яндекс Радар. Браузеры в Росии. – URL: https://radar.yandex.ru/browsers (дата обращения: 22.03.2021).

125. Балау Э.И, Байкова П.Д, Ефимов Н.В. Модель информационной безопасности детей. Настройка браузеров в целях родительского контроля / Э.П. Баллау, П.Д.Байкова, Н.В. Ефимов // Фундаментальные и прикладные исследования молодых ученых. 2020. – С. 345–351

126. *Патент № RU2651252C1,* Российская Федерация, МПК G06F 21/55 (2013.01), G06F 17/30 (2006.01). Способ ограничения доступа пользователю к подозрительным объектам социальной сети: No 2017115052: заявл. 28.04.2017: опубл. 18.04.2018 / Ларкина А.Н., Тушканов В.Н.; заявитель АО «Лаборатория Касперского». – С. 1-22

127. Минин А.Я. Информационная безопасность в образовании: обучающихся и обучающих. / А.Я. Минин // Наука и школа. – 2017. – Т. 1.– С. 29–35.

128. Kozlov F., Yuen I., Kowalczyk J., Bernhardt D., Freeman D. Evaluating Changes to Fake Account Verification Systems // 23rd International Symposium on Research in Attacks, Intrusions and Defenses ({RAID} 2020). 2020. P. 135–148.Статистика социальных сетей в России 2020. – URL: https://livedune.ru/blog/statistika_socsetej_v_rossii (дата обращения: 22.03.2021).

129. Announcing the winners of Facebook’s request for proposals on misinformation and polarization - Facebook Research. – URL: https://research.fb.com/blog/2020/08/announcing-the-winners-of-facebooks-request-for-proposals-on-misinformation-and-polarization/ (дата обращения: 22.03.2021).

130. *Патент №* US8244848B1, США. Integrated social network environment. /S. Narayanan, A. Li, Ch. Eugene, L. Namita, G. Peter, X. Deng. 2010.

131. FAQ по архитектуре и работе ВКонтакте. – URL: https://habr.com/ru/company/oleg-bunin/blog/449254/ (дата обращения: 22.03.2021).

132. Единый реестр доменных имен, указателей страниц сайтов в сети «Интернет» и сетевых адресов, позволяющих идентифицировать сайты в сети «Интернет», содержащие информацию, распространение которой в Российской Федерации запрещено (ЕАИС) / Роскомнадзор: официальный сайт. – URL: https://eais.rkn.gov.ru/ (дата обращения: 22.03.2021).

133. *Патент №* US9191411B2, США. Protecting against suspect social entities /Foster J.C., Cullison Ch.B., Francis R., B.E. 2014. Заявитель: Zerofox Inc

134. *Патент №* US20200014790A1, США. Devices and methods for improving web safety and deterrence of cyberbullying. / Day R.W., Wise E., Sigler S. R.J.P. 2020. Заявитель: WEBSAFETY Inc

135. *Патент №* US9961115B2, США. Cloud-based analytics to mitigate abuse from internet trolls / Dalton M.D. L.J.S. 2018. Заявитель: International Business Machines Corp (IBM)

136. *Патент №* US9659185B2, США. Method for detecting spammers and fake profiles in social networks /Elovici Y., Fire M. G. Katz. 2017. Заявитель: BG Negev Technologies and Applications Ltd.

137. Остапенко А.Г., Соколова Е.С., Ещенко А.В., Остапенко А.А. Чапурина Т.Ю. Основы метрологии контентов для мониторинга социальных сетей на предмет обеспечения информационной безопасности (часть 1) / А.Г.Остапенко, Е.С. Семенова, А.В. Ещенко, Т.Ю. Чапурина // Информационная безопасность. – 2019. – Т. 22, № 2. – С. 170–180.

138. *Патент №* US10747837B2, США. Containing disinformation spread using customizable intelligence channels / Goldenstein J-C, Searing J. E., FINN E. J. 2019. Заявитель: Creopoint Inc.

139. Pondhe N. J., Jadhav H.B. A System to Filter Unwanted Messages on Social Networking Site // JARIIE. 2017. Vol. 3, № 1. P. 356–359.

140. Паращук И.Б., Башкирцев А.С. К вопросу обоснования систем показателей качества процессов принятия решения и поддержки принятия решения в интересах управления информационными сетями / И.Б. Паращук, А.С. Башкирцев // Информация и космос. – 2016. – Т. 2. – С. 65–71.

141. Ушаков И. А. Обнаружение инсайдеров в компьютерных сетях на основе комбинирования экспертных правил, методов машинного обучения и обработки больших данных: диссертация ... кандидата технических наук: 05.13.19 / Ушаков Игорь Александрович; [Место защиты: ФГБУН Санкт-Петербургский институт информатики и автоматизации Российской академии наук]. – Санкт-Петербург, 2020. – 215 с. Методы и системы защиты информации, информационная безопасность. Хранение: OD 61 20-5/725.

142. Чечулин А. А. Построение и анализ деревьев атак на компьютерные сети с учетом требования оперативности: диссертация ... кандидата технических наук: 05.13.19 / Чечулин Андрей Алексеевич; [Место защиты: С.-Петерб. ин-т информатики и автоматизации РАН]. - Санкт-Петербург, 2013. - 152 с. Методы и системы защиты информации, информационная безопасность. Хранение: OD 61 14-5/933;

143. ГОСТ Р 20886-85. Организация данных в системах обработки данных. Термины и определения (с Изменениями N 1, 2), М.: Стандартинформ, 1986.

144. Taylor R.W., Frank R.L. CODASYL Data-Base Management Systems // ACM Comput. Surv. 1976. Vol. 8, № 1. P. 67–103.

145. Codd E.F. The Relational Model for Database Management: Version 2 // Database. 1990. 538 p.

146. Виткова Л.А. Модель и алгоритмы защиты от вредоносной информации в социальных сетях// В сборнике: Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2020). IX Международная научно-техническая и научно-методическая конференция: сборник научных статей. Санкт-Петербург, 2020. С. 235-240.

147. Parinov A.V., Sokokova E.S., Urasov V.G., Tolstykh N.N., Filatov V.V. Destructive Content In Multinetwork Socio-Informative Space: Formalization Of The Procedure Of Detection // International Journal of Pure and Applied Mathematics. 2018. Vol. 119, № 15. P. 2587–2591.

148. Pronoza A., Vitkova L., Chehulin A., Kotenko I. Visual analysis of information dissemination channels in social network for protection against inappropriate content // Advances in Intelligent Systems and Computing. 2019. Vol. 875. P. 95–105.

149. Vitkova L., Saenko I., Tushkanova O. An Approach to Creating an Intelligent System for Detecting and Countering Inappropriate Information on the Internet // Studies in Computational Intelligence. 2020. Vol. 868. P. 244–254.

150. Виткова Л.А., Кураева А.М., Проноза А.А., Чечулин А.А. Анализ методов выявления и оценки страниц лидеров мнений в социальных сетях // В сборнике: Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2019). сборник научных статей VIII Международной научно-технической и научно-методической конференцияи: в 4 т. 2019. С. 233-237.

151. Виткова Л.А. Место и роль мониторинга и противодействия нежелательной информации в социальных сетях //В сборнике: Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2019). сборник научных статей VIII Международной научно-технической и научно-методической конференцияи: в 4 т. 2019. С. 209-212.

152. Денисов Е.И., Андреянов Я.В., Виткова Л.А., Сахаров Д.В. Информационное воздействие социальных сетей //В сборнике: Региональная информатика «РИ-2018». материалы конференции. 2018. С. 569-570.

153. *Патент №* US10084742B2, США. Social media threat monitor / M.A. Lewis, J.R. Bedser, J. Pinyan. 2018. Заявитель: Ithreat Cyber Group Inc.

154. Юсупов Р.М. Заболотских В.П. Концептуальные и научно-методологические основы информатизации / Р.М. Юсупов, В.П. Зболотских. – М.: Наука, 2009. – 541 с.

155. Vitkova L. A. et al. Selection of countermeasures against propagation of harmful information via Internet //IOP Conference Series: Materials Science and Engineering. – IOP Publishing, 2021. – Т. 1032. – №. 1. – С. 012017.

156. CREOpoint Filtering social media and containing disinformation / Creopoint Inc. – URL: https://www.mycreopoint.com/ (дата обращения: 06.04.2021).

157. *Патент №* WO2020163508A1, WIPO (PCT). Propagation de signalement de désinformation à l'aide de canaux d'intelligence personnalisables/ Goldenstein J-C, Searing J. E., Finn E.J. 2020. Заявитель: Creopoint Inc.

158. Чечулин А.А. Мониторинг и противодействие вредоносному влиянию в информационном пространстве социальных сетей. – URL: https://rscf.ru/prjcard_int?18-71-10094 (дата обращения: 07.04.2021).

160. Проноза А.А. Чечулин А.А. Компонент анализа данных о сообществах в социальной сети ВКонтакте. Свидетельство № 2019667056, 2019.

161. Левшун Д.С., Чечулин А.А. Компонент сбора комментариев в социальной сети ВКонтакте. Свидетельство о регистрации программы для ЭВМ 2019663976., 29.10.2019.

162. Рунеев А.Ю., Котенко И.В. Основы теории управления в системах военного назначения. Часть 2: учебное пособие. / А.Ю. Рунев, И.В. Котенко. – СПб.: ВУС, 2000. – 158 с.

ПРИЛОЖЕНИЕ А. Список контрмер

Таблица А.1. − Список контрмер на шаге 2 на стадии настройки методики противодействия вредоносной информации

Метрика сложности для реализации контрмеры	начальная сложность	Классы	КС					
			Тип воздействия			Метод воздействия		
		Вес класса	Вес класса			Вес класса		
		Подкласс	позитивный	негативный	нейтральный	авто	автомат-й	ручной
		Сложность реализации	Уровень сложности	Уровень сложности	Уровень сложности	Уровень сложности	Уровень сложности	Уровень сложности
c_1	cw_1	Блокировка источника через оператора связи	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*
...	...	Блокировка источника через социальную сеть	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*
cn	cw_n	Блокировка источника через black_list	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*
		Блокировка сообщения через оператора связи	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*
		Блокировка сообщения через социальную сеть	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*
		Блокировка сообщения через black_list	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*	*{0,1}*

ПРИЛОЖЕНИЕ Б. Диаграмма базы данных информационных угроз и контрмер

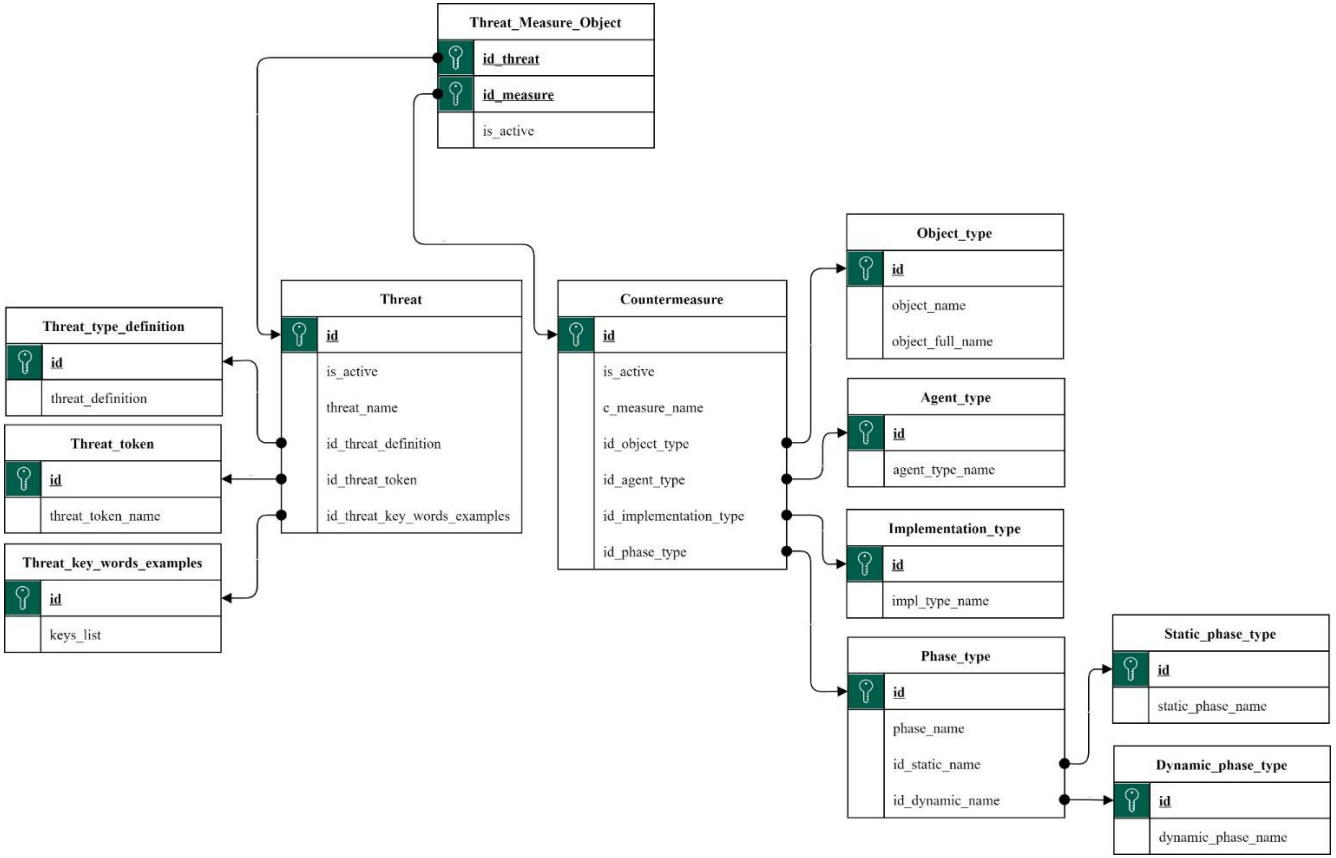


Рисунок Б.1 – Диаграмма базы данных информационных угроз и контрмер

ПРИЛОЖЕНИЕ В. Структура базы данных информационных угроз и контрмер

Структура базы данных информационных угроз и контрмер:

1) Threat_Measure_Object – содержит список угроз и соответствующих им контрмерам;
2) Threat – содержит список угроз в соответствии с атрибутами вредоносной информации;
3) Threat_type_definition – содержит список с описаниями типов угроз;
4) Threat_token – содержит список признаков типов угроз;
5) Threat_key_words_examples – содержит список ключевых слов в соответствии с типами угроз;
6) Countermeasure – содержит список контрмер в соответствии с атрибутами их применения;
7) Object_type – содержит список объектов, к которым применимы контрмеры;
8) Agent_type – содержит список неких агентов реализации, агентов, через которые реализуется контрмера;
9) Implementation_type – содержит список типов реализации контрмер;
10) Phase_type – содержит список категории этапов реализации контрмер;
11) Static_phase_type – содержит список статических этапов реализации контрмер;
12) Dynamic_phase_type – содержит список динамических этапов реализации контрмер.

Ниже представлены обозначения и описания полей, указанных выше реляционных таблиц базы данных контрмер.

Таблица Threat_Measure_Object:

id_threat – уникальный, целочисленный идентификатор угрозы (внешний ключ таблицы Threat);

id_measure – уникальный, целочисленный идентификатор контрмеры (внешний ключ таблицы Countermeasure);

is_active – битовое поле, идентифицирующее применимость к текущей угрозе соответствующей контрмеры (1=true, применима; 0=false, не применима).

Таблица Threat:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор угрозы (первичный ключ);

is_active – битовое поле, идентифицирующее возможность реализации угрозы (1=true, реализуема; рассматривались только реализуемые угрозы);

threat_name – ненулевое уникальное текстовое поле, идентифицирующее угрозу в Th-формате (Th1, Th2…) с кратким описанием угрозы;

id_threat_definition – уникальный, целочисленный идентификатор описания угрозы (внешний ключ таблицы Threat_type_definition);

id_threat_token – уникальный, целочисленный идентификатор признака угрозы (внешний ключ таблицы Threat_token);

id_threat_key_words_examples – уникальный, целочисленный идентификатор ключевых слов, характеризующих угрозу (внешний ключ таблицы Threat_key_words_examples).

Таблица Threat_type_definition:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор подробного описания угрозы (первичный ключ);

threat_name – ненулевое уникальное текстовое поле, содержащее подробное описание угрозы.

Таблица Threat_token:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор признака угрозы (первичный ключ);

threat_token_name – ненулевое, уникальное текстовое поле, содержащее признак угрозы.

Таблица Threat_key_words_examples:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор ключевых слов, характеризующих угрозу (первичный ключ);

keys_list – ненулевое, уникальное текстовое поле, содержащее список ключевых слов, характеризующих угрозу.

Таблица Countermeasure:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор контрмеры (первичный ключ);

is_active – битовое поле, идентифицирующее возможность реализации контрмеры (1=true, реализуема; рассматривались только реализуемые угрозы);

c_measure_name – ненулевое уникальное текстовое поле, идентифицирующее контрмеру в С-формате (С1, С2…) с кратким описанием контрмеры;

id_object_type – уникальный, целочисленный идентификатор типа объекта, к которому применима контрмера (внешний ключ таблицы Object_type);

id_agent_type – уникальный, целочисленный идентификатор агента реализации контрмеры (внешний ключ таблицы Agent_type);

id_implementation_type – уникальный, целочисленный идентификатор типа реализации контрмеры (внешний ключ таблицы Implementation_type);

id_phase_type – уникальный, целочисленный идентификатор категории этапа реализации контрмеры (внешний ключ таблицы Phase_type).

Таблица Object_type:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор объекта, к которому применима контрмера (первичный ключ);

object_name – ненулевое, уникальное текстовое поле, содержащее описание сетевого информационного объекта в виде аббревиатуры;

object_full_name – ненулевое, уникальное текстовое поле, содержащее подробное описание объекта.

Таблица Agent_type:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор агента реализации контрмеры (первичный ключ);

agent_type_name – ненулевое, уникальное текстовое поле, содержащее описание агента, через который реализуется контрмера.

Таблица Implementation_type:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор типа реализации контрмеры (первичный ключ);

impl_type_name – ненулевое, уникальное текстовое поле, содержащее описание типа реализации контрмеры.

Таблица Phase_type:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор категории этапа реализации контрмеры (первичный ключ);

phase_name – ненулевое уникальное текстовое поле, идентифицирующее категорию этапа реализации контрмеры;

id_static_name – уникальный, целочисленный идентификатор статического этапа реализации контрмеры (внешний ключ таблицы Static_phase_type);

id_dynamic_name – уникальный, целочисленный идентификатор динамического этапа реализации контрмеры (внешний ключ таблицы Dynamic_phase_type).

Таблица Static_phase_type:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор статического этапа реализации контрмеры (первичный ключ);

static_phase_name – ненулевое уникальное текстовое поле, содержащее описание статического этапа реализации контрмеры.

Таблица Dynamic_phase_type:

id – ненулевой, уникальный, целочисленный, автоинкрементируемый идентификатор динамического этапа реализации контрмеры (первичный ключ);

dynamic_phase_name – ненулевое уникальное текстовое поле, содержащее описание динамического этапа реализации контрмеры.

ПРИЛОЖЕНИЕ Г. Список публикаций соискателя по теме диссертации

Публикации в рецензируемых журналах из списка ВАК:

1. Виткова Л.А. Модель вредоносной информации и ее распространителя в социальных сетях / Л.А. Виткова, Д.В. Сахаров, Д.Р. Голузина // Защита информации. Инсайд. – Спб., 2020. –№3 (93). – С. 66-72.

2. Гамидов Т.О. Разработка моделей и алгоритмов анализа данных для исследования хода инцидентов и кризисов в социальных сетях / Т.О. Гамидов, Л.А. Виткова, М.М. Ковцур // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. – СПб., 2020. – № 2. – С. 3-10.

3. Виткова Л.А. Выбор мер противодействия вредоносной информации в социальных сетях / Л.А. Виткова, А.А. Чечулин, Д.В. Сахаров // Вестник Воронежского института ФСИН России. – Воронеж, 2020. – Т. 3. – С. 20-29.

4. Виткова Л.А. Архитектура системы выявления и противодействия нежелательной информации в социальных сетях. / Л.А. Виткова, И.Б. Саенко // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. – СПб., 2020. – № 3. – С. 33-39.

5. Виткова Л.А. Методика анализа аудитории канала распространения информации в социальных сетях. // Известия высших учебных заведений. Технология легкой промышленности. – СПб, 2018. – Т. 42, № 4. – С. 5-10.

6. Проноза А.А. Методика выявления канала распространения информации в социальных сетях / А.А. Проноза, Л.А. Виткова, А.А. Чечулин, И. В. Котенко, Д.В. Сахаров // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. – СПб., 2018. – Т. 14, № 4. – С.362-377

Публикации в зарубежных изданиях из баз данных WOS и Scopus:

7. Kotenko I.V. The intelligent system for detection and counteraction of malicious and inappropriate information on the Internet / I.V. Kotenko, L.A. Vitkova, I.B. Saenko, O.N. Tushkanova, A.A. Branitsky// AI Communications, 2020. – Vol 33(1). – C. 1-13. – ISSN 0921-7126

8. Vitkova L.A. Selection of countermeasures against propagation of harmful information via Internet / L. A. Vitkova, A. P. Pronichev, E. V. Doynikova, I. B. Saenko // IOP Conference Series: Materials Science and Engineering, 2021 Vol 1032, – 1032 012017. – ISSN 1757-8981

9. Vitkova, L.A. The technology of intelligent analytical processing of digital network objects for detection and counteraction of inappropriate information / L.A. Vitkova, I.B. Saenko, A.A. Chechulin, I.B. Parashchuk // The 1st International Conference on Computer Technology Innovations dedicated to the 100th anniversary of the Gorky House of Scientists of Russian Academy of Science (ICCTI – 2020). Official conference proceedings, 2020. – P 13-19. – ISBN 978-5-9676-1216-9

10. Vitkova L.A. Approach to Identification and Analysis of Information Sources in Social Networks / L. A. Vitkova, M. V. Kolomeets // Proceedings of the 13th International Symposium on Intelligent Distributed Computing (IDC 2019), October 7-9, 2019, Saint-Petersburg, Russia. 2020. P. 285-293. – ISSN 1860-949X.

11. Vitkova L.A. An Approach to Creating an Intelligent System for Detecting and Countering Inappropriate Information on the Internet / L.A. Vitkova, I.B. Saenko, O.N. Tushkanova // Proceedings of the 13th International Symposium on Intelligent Distributed Computing (IDC 2019), October 7-9, 2019, Saint-Petersburg, Russia. 2020. – P. 244-254. – ISSN 1860-949X.

12. Vitkova, L.A. Hybrid Approach for Bots Detection in Social Networks Based on Topological, Textual and Statistical Features / L.A. Vitkova, Kotenko I.V., M.V. Kolomeets, O.N. Tushkanova, A.A. Chechulin // Advances in Intelligent Systems and Computing 1156 AISC, 2019, P. 412-421

13. Pronoza A.A. Visual analysis of information dissemination channels in social network for protection against inappropriate content / A.A. Pronoza, L.A.

Vitkova, A.A. Chechulin, I.V. Kotenko // 3rd International Scientific Conference on Intelligent Information Technologies for Industry, IITI 2018. Sochi, Russian Federation, 17-21 September 2018. Advances in Intelligent Systems and Computing. Vol. 875, 2019. P. 95-105.

14. Kotenko I.V. Monitoring and counteraction to malicious influences in the information space of social networks / I.V. Kotenko, I.B. Saenko, A.A. Chechulin, V.A. Desnitsky, L.A. Vitkova, A.A. Pronoza // The 10th Social Informatics conference (SocInfo2018). September 25–28, 2018, Saint Petersburg, Russia. Proceedings, Part II. Lecture Notes in Computer Science, Vol.11186, Springer 2018, P.1 59-167. – ISBN 978-3-030-01158-1.

Публикации в сборниках трудов конференций включенных в РИНЦ:

15. Виткова Л.А. Методология выявления искусственной мобилизации протестной активности в соцсетях / Л.А. Виткова, К.А. Науменко // Тезисы докладов научного семинара «Фундаментальные проблемы управления производственными процессами в условиях перехода к Индустрии 4.0» в рамках МНТК «Автоматизация», 2020. – С. 212-214

16. Виткова Л.А. Модель и алгоритмы защиты от вредоносной информации в социальных сетях // IX МНТиНМК «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (АПИНО-2020)». 26-27 февраля 2020 г. Сборник научных статей. 2020. – Т. 1. – С. 235-240.

17. Валиева К.А. Методика обнаружения вредоносной информации в информационном пространстве социальных сетей / К.А. Валиева, Л.А. Виткова, Е.В. Смирнов // IX МНТиНМК «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (АПИНО-2020)». 26-27 февраля 2020 г. Сборник научных статей. 2020. – Т. 1. – С. 206-211.

18. Виткова Л.А. Противодействие распространению нежелательной информации в информационном пространстве социальных сетей / Л. А. Виткова, М.А. Справцева // IX МНТиНМК «Актуальные проблемы

инфотелекоммуникаций в науке и образовании» (АПИНО-2020)». 26-27 февраля 2020 г. Сборник научных статей. 2020. – Т. 1. – С. 258-261.

19. Виткова Л.А. О моделировании процессов выявления и противодействия террористической и экстремистской активности в интернете и социальных сетях / Л.А. Виткова, Е.В. Дойникова, А.П. Проничев // Сборник научных статей XVII Санкт-Петербургской международной конференции «Региональная информатика (РИ-2020)». СПб: СПОИСУ, 2020. – С. 117-118.

20. Виткова Л.А. Распределенный сбор и обработка данных в системах мониторинга информационного пространства социальных сетей / Л.А. Виткова, И.В. Котенко, А.В. Хинензон // VIII МНТиНМК «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (АПИНО 2019). 2019. – Т. 1. – С. 228-232

Свидетельства о государственной регистрации программ для ЭВМ:

21. Виткова Л.А. Компонент сегментации пользователей по их активности в социальных сетях / Л.А. Виткова, А.А. Чечулин, И.В. Котенко – Свидетельство о государственной регистрации программы для ЭВМ № 2019664733. Зарегистрировано в Реестре программ для ЭВМ 13.11.2019.

22. Федорченко Е.В. Компонент выбора мер противодействия нежелательной, сомнительной и вредоносной информации / Е.В. Федорченко, Л.А. Виткова, А.П. Проничев, И.Б. Саенко. – Свидетельство о государственной регистрации программы для ЭВМ № 2020665591. Зарегистрировано в Реестре программ для ЭВМ 27.11.2020.

23. Виткова Л.А. База данных для учета нежелательной информации совместно с мерами противодействия / Л.А. Виткова, Е.О. Березина, А.П. Проничев, И.Б. Саенко, И.В. Котенко – Свидетельство о государственной регистрации программы для ЭВМ № 2020622557. Зарегистрировано в Реестре программ для ЭВМ 08.12.2020.

ПРИЛОЖЕНИЕ Д. Копии актов внедрения

МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ «САНКТ-ПЕТЕРБУРГСКИЙ ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР РОССИЙСКОЙ АКАДЕМИИ НАУК» (СПб ФИЦ РАН)

14 линия В.О., д. 39, Санкт-Петербург, 199178
Телефон: (812) 328-34-11, факс: (812) 328-44-50, E-mail: info@spcras.ru, https://spcras.ru/
ОКПО 04683303, ОГРН 1027800514411, ИНН/КПП 7801003920/780101001

УТВЕРЖДАЮ
Директор СПб ФИЦ РАН
Профессор РАН

_____ А.Л. Ронжин
12 апреля 2021 г.

Акт внедрения результатов диссертационного исследования Витковой Лидии Андреевны «Модели, алгоритмы и методика противодействия вредоносной информации в социальных сетях», представленного на соискание ученой степени кандидата наук по научной специальности 05.13.19 – Методы и системы защиты информации, информационная безопасность (технические науки)

Комиссия в составе: председателя – заведующего лабораторией проблем компьютерной безопасности, доктора технических наук, профессора Котенко Игоря Витальевича; ведущего научного сотрудника лаборатории проблем компьютерной безопасности, доктора технических наук, профессора Саенко Игоря Борисовича; ведущего научного сотрудника лаборатории проблем компьютерной безопасности, кандидата технических наук, доцента Чечулина Андрея Алексеевича, составила настоящий акт в том, что результаты диссертационного исследования Витковой Лидии Андреевны «Модели, алгоритмы и методика противодействия вредоносной информации в социальных сетях» были внедрены при выполнении научно-исследовательских работ в лаборатории проблем компьютерной безопасности (Грант Российского научного фонда № 18-71-10094 «Мониторинг и противодействие вредоносному влиянию в информационном пространстве социальных сетей», 2018-2021; Грант Российского научного фонда № 18-11-00302 «Интеллектуальная обработка цифрового сетевого контента для эффективного обнаружения и противодействия нежелательной, сомнительной и вредоносной информации», 2018-2020). С применением разработанных Витковой Л.А. моделей, алгоритмов, методики и архитектуры решались общие задачи противодействия вредоносной информации в социальных сетях и в сети Интернет, включая:

Результат «комплекс моделей социальной сети, источника и вредоносной информации» использовался при решении задачи разработки общего подхода и требований, предъявляемых к компонентам выработки и выбора мер противодействия нежелательной, сомнительной и вредоносной информации в проекте: Грант Российского научного фонда № 18-11-00302 «Интеллектуальная обработка цифрового сетевого контента для эффективного обнаружения и противодействия нежелательной, сомнительной и вредоносной информации».

Результат «методика противодействия вредоносной информации» использовался при решении задачи разработки пошаговой методики оценки сложности для реализации мер противодействия, которая позволяет ранжировать меры по степени сложности их выполнения в проекте: Грант Российского научного фонда № 18-71-10094 «Мониторинг и противодействие вредоносному влиянию в информационном пространстве социальных сетей».

Результат «комплекс алгоритмов анализа источников и ранжирования контрмер» и результат «архитектура и программные прототипы компонентов системы противодействия вредоносной информации в социальных сетях» использовались при решении задач разработки алгоритмов, архитектуры и программных прототипов компонентов выявления цели для реализации мер противодействия вредоносному влиянию в социальных сетях в проекте: Грант Российского научного фонда № 18-71-10094 «Мониторинг и противодействие вредоносному влиянию в информационном пространстве социальных сетей».

Комиссия отмечает теоретическую, практическую значимость и новизну полученных в работе результатов.

Председатель комиссии:

Заведующий лаборатории
проблем компьютерной безопасности,
доктор технических наук,
профессор — Котенко Игорь Витальевич

Члены комиссии:

Ведущий
научный сотрудник лаборатории
проблем компьютерной безопасности,
доктор технических наук
профессор — Саенко Игорь Борисович

Ведущий
научный сотрудник лаборатории
проблем компьютерной безопасности,
кандидат технических наук
доцент — Чечулин Андрей Алексеевич

МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ ИМ. ПРОФ. М.А. БОНЧ-БРУЕВИЧА» (СПбГУТ)

Санкт-Петербург

УТВЕРЖДАЮ
и.о. ректора СПбГУТ, д.т.н.,
проф. _____ Г. М. Машков
_____ 2021 г.

Акт

об использовании результатов диссертационной работы
Витковой Лидии Андреевны
«Модели, алгоритмы и методика противодействия вредоносной информации в социальных сетях» в учебном процессе университета

Настоящий Акт составлен в том, что результаты диссертационной работы Витковой Лидии Андреевны, а именно:

- комплекс моделей социальной сети, источника и вредоносной информации;
- комплекс алгоритмов анализа источников и ранжирования контрмер;
- методика противодействия вредоносной информации;
- архитектура и программные прототипы компонентов системы противодействия вредоносной информации в социальных сетях

используются кафедрой защищенных систем связи федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» в учебном процессе по направлению подготовки магистрантов первого года обучения 10.04.01

«Информационная безопасность» в дисциплине «Технологии обеспечения информационной безопасности больших данных» (рабочая программа дисциплины, регистрационный № 20.05/333-Д) при чтении курсов лекций, проведении практических занятий и лабораторных работ.

Председатель комиссии:

заведующий кафедрой ЗСС,
к.т.н., доцент — Красов Андрей Владимирович

Члены комиссии:

доцент кафедры ЗСС,
к.т.н., доцент — Кушнир Дмитрий Викторович

доцент кафедры ЗСС,
к.т.н., доцент — Волкогонов Владимир Никитич

старший преподаватель
кафедры ЗСС — Гельфанд Артем Максимович

14.04.2021

МИНОБРНАУКИ РОССИИ
федеральное государственное
бюджетное образовательное учреждение
высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПРОМЫШЛЕННЫХ ТЕХНОЛОГИЙ
И ДИЗАЙНА»
(СПбГУПТД)

Б. Морская ул., д. 18, Санкт-Петербург, 191186
Тел. (812) 315-75-25 Факс (812) 571-95-84
E-mail: rector@sutd.ru http://www.sutd.ru
ОКПО 02068605, ОГРН 1027809192102,
ИНН/КПП 7808042283/784001001

12.04.2021г. № 38-03-25/03-37

на № _____ от _____

УТВЕРЖДАЮ

Проректор по научной работе

СПбГУПТД

д.т.н., проф. Макаров А.Г.

«12.04» 2021г.

АКТ

о внедрении научных результатов диссертационной работы
Витковой Лидии Андреевны
«Модели, алгоритмы и методика противодействия вредоносной информации в социальных сетях»

Комиссия в составе: зам. зав. кафедрой интеллектуальных систем и защиты информации (ИСЗИ) к.т.н., доц. Вагнер В.И., к.т.н. Штеренберг С.И., составили настоящий акт о том, что научные результаты, Витковой Л.А., полученные ей в ходе диссертационного исследования на тему «Модели, алгоритмы и методика противодействия вредоносной информации в социальных сетях», используются на кафедре ИСЗИ СПбГУПТД при подготовке лекционно-практических занятий, а именно:

- комплекс моделей социальной сети, источника и вредоносной информации, комплекс алгоритмов анализа источников и ранжирования контрмер и методика противодействия вредоносной информации для направления подготовки бакалавров 10.03.01 – «Информационная безопасность» по дисциплине «Комплексная защита на предприятии»
- архитектура и программные прототипы компонентов системы противодействия вредоносной информации в социальных сетях для направления подготовки бакалавров 10.03.01 – «Информационная

безопасность» по дисциплине «Технологии и методы программирования»

Комиссия считает, что внедрение указанных научных результатов Витковой Л.А. в образовательный процесс СПбГУПТД позволило повысить качество подготовки бакалавров по направлению 10.03.01 Информационная безопасность.

Комиссия отмечает практическую значимость и новизну полученных в работе результатов.

Председатель комиссии:

зам. зав. кафедрой Интеллектуальных систем и защиты информации, к.т.н., доцент	Вагнер Виктория Игоревна

Члены комиссии:

Доцент кафедры Интеллектуальных систем и защиты информации, к.т.н.	Штеренберг Станислав Игоревич

GLORYSTORY

репутационное агентство

ООО «Жасмин»
Инн7810880035 КПП 781001001

тел.: +7 812 655 0558
www.glorystory.ru

УТВЕРЖДАЮ
Генеральный директор
ООО «Жасмин»
Мартьянова Ж.А.
«3» апреля 2021г.

АКТ

о внедрении результатов диссертационной работы
Витковой Лидии Андреевны
«Модели, алгоритмы и методика противодействия вредоносной информации в социальных сетях»

Комиссия в составе:
Мартьяновой Ж.А. – председателя комиссии, генерального директора;
Намятовой К.А. – член комиссии, управляющий партнер;
Крупени К.А. – член комиссии, управляющий партнер;

составила настоящий акт о том, что результаты диссертационной работы Витковой Лидии Андреевны, а именно:

- комплекс моделей социальной сети, источника и вредоносной информации;
- комплекс алгоритмов анализа источников и ранжирования контрмер;
- методика противодействия вредоносной информации;
- архитектура и программные прототипы компонентов системы противодействия вредоносной информации в социальных сетях.

используются при анализе репутационного поля в социальных сетях и при подготовке репутационных портретов для заказчиков PA GloryStory. Также результаты диссертационной работы и программные прототипы компонентов архитектуры используются для анализа источников информации и сортировки объектов противодействия.

Комиссия отмечает практическую значимость и новизну полученных в работе результатов.

Председатель комиссии: Мартьянова Ж.А.

Члены комиссии: Намятова К.А.

ПРИЛОЖЕНИЕ Ж. Копии зарегистрированных свидетельств на результаты интеллектуальной собственности

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019663984

Компонент устранения неопределенности оценки и категоризации смыслового наполнения информационных объектов на основе использования методов обработки неполных, противоречивых и нечетких знаний

Правообладатель: ***Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (RU)***

Авторы: ***Виткова Лидия Андреевна (RU),***
Паращук Игорь Борисович (RU)

Заявка № **2019662768**
Дата поступления **18 октября 2019 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***29 октября 2019 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020665591

Компонент выбора мер противодействия нежелательной, сомнительной и вредоносной информации

Правообладатель: ***Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук" (RU)***

Авторы: ***Федорченко Елена Владимировна (RU), Виткова Лидия Андреевна (RU), Проничев Алексей Петрович (RU), Саенко Игорь Борисович (RU)***

Заявка № **2020664742**
Дата поступления **21 ноября 2020 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***27 ноября 2020 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2020622557

База данных для учета нежелательной информации совместно с мерами противодействия

Правообладатель: ***Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук" (RU)***

Авторы: ***Виткова Лидия Андреевна (RU), Березина Елизавета Олеговна (RU), Проничев Алексей Петрович (RU), Саенко Игорь Борисович (RU), Котенко Игорь Витальевич (RU)***

Заявка № **2020622323**

Дата поступления **24 ноября 2020 г.**

Дата государственной регистрации в Реестре баз данных ***08 декабря 2020 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев