DOI 10.37882/2223–2966.2022.03.18

РАЗРАБОТКА МЕТОДИКИ ВНЕДРЕНИЯ И ВЫЯВЛЕНИЯ ЭФФЕКТИВНОСТИ SIEM-СИСТЕМЫ В СРЕДЕ ДОВЕРЕННОЙ ЗОНЫ[1]

DEVELOPMENT OF A METHODOLOGY FOR IMPLEMENTING AND IDENTIFYING THE EFFECTIVENESS OF A SIEM SYSTEM IN A TRUSTED ZONE ENVIRONMENT

A. Krasov

Summary. Classification of steganographic methods of information transformation is described in the basic model of threats to the security of personal data during their processing in personal data information systems, approved by the FSTEC of the Russian Federation in 2008. This technique will be useful for familiarization of all managers and heads of security services of organizations that plan to install a SIEM system. The article presents the results of work on the Grant-I B5/2020 project, proposals for improving the basic model.

Keywords: threat model, SIEM, trusted zone, machine learning, IPS/IDS.

Красов Андрей Владимирович
К.т.н., доцент, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича
krasov@inbox.ru

Аннотация. Классификация стеганографических методов преобразования информации описана в базовой модели угроз безопасности персональных данных при их обработке в информационных системах персональных данных, утверждённой ФСТЭК РФ в 2008 году. Данная методика будет полезна для ознакомления всех руководителей и начальников служб безопасности организациях, которые планируют установить SIEM-систему. В статье приводятся результаты работы по проекту Грант-ИБ 5/2020, предложения по совершенствованию базовой модели.

Ключевые слова: модель угроз, SIEM, доверенная зона, машинное обучение, IPS/IDS.

Используемый для запуска методики специализированный стеганографический агент в целом подчиняется следующим обязательным правилам при создании:

1. Циклическое усовершенствование программы или ее алгоритма. Программа модифицирует себя, решает некоторую эталонную задачу и оценивает результат. На основе результата принимается решение о новом цикле модификации;
2. В случае агента такой подход может использоваться для модификации своего кода при переходе на новую машину и обхода таким образом антивирусов. При этом можно как менять существующий код, так и добавлять новый, который не несет смысловой нагрузки;
3. Программа может попытаться найти в ИС исходный код нового решения какой-то из своих функций, загрузить его, соответствующим образом изменить свой код с учетом новой реализации данной функции и попробовать откомпилировать, и проверить работу новой копии. Т.о. можно использовать агента, который со временем будет сам находить лучшие решения проблем и использовать их в своей работе.
4. На основе этого для агента строится плановое выполнение мониторинга (рис. 1) для поддержания работоспособности стеганографической системы защиты ПО в организации.

Поставленная цель построения доверенной среды достигается тем, что в одном из выбранных способов [2,3] формируется массив для запоминания фрагментированных пакетов сообщения и массивы для запоминания параметров, выделенных из запомненных пакетов сообщений, принимают очередной пакет сообщения из канала связи, запоминают его, анализируют приоритетный пакет на обнаружение факта наличия или отсутствия компьютерной атаки, и при отсутствии компьютерной атаки, передают очередной пакет сообщения в информационно-вычислительную сеть, а в случае обнаружения компьютерной атаки, принимают решение о запрете передачи ЦВЗ в ОС Linux и удаляют ранее запомненные значения Hid-кода (скрытый код) сообщения из массивов, дополнительно в качестве выделен-

[1] Исследование выполнено при финансовой поддержке Минобрнауки России (грант ИБ) в рамках научного проекта № 5/2020.

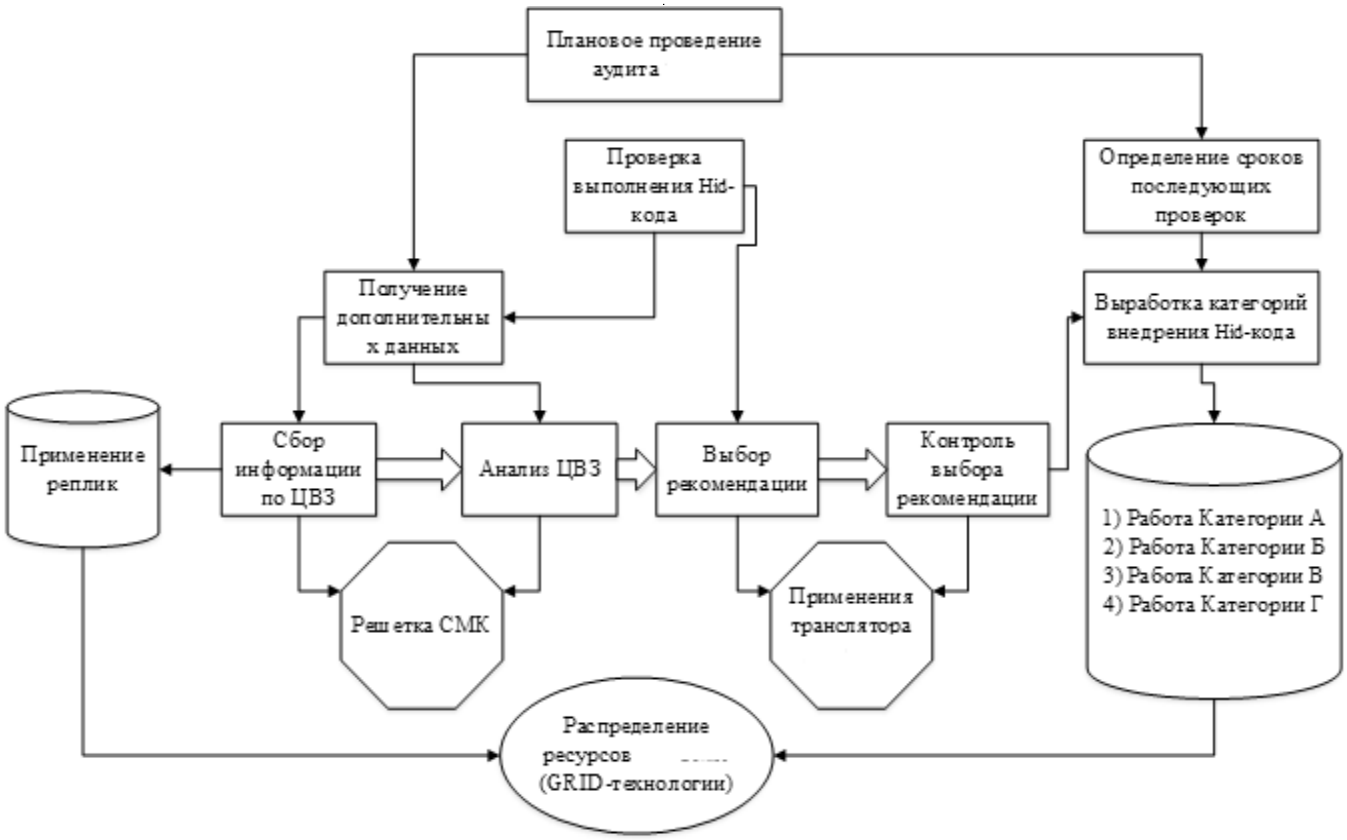


Рис. 1. Мониторинг ОС Linux среды работы агента

ных полей из запомненных пакетов сообщений используют поля данных: «Время жизни агента» {T}, «Опции» {O}, «IP адрес назначения» {D}, «IP адрес источника» {I}, которые запоминают в сформированных для них массивах. Реализация заявленной методики поясняется алгоритмом (рис. 2).

Подготовленное поле «Время жизни агента» определяет максимальное время существования дейтаграммы в сети. Поле «Опции» является необязательным и имеет переменную длину. Поддержка опций должна реализоваться во всех модулях IP (узлах и маршрутизаторах).

Далее подходит очередь для реализации конечного стеговложения информации. Основной элемент маркирования — образы ОС Linux. На данный момент выбран стандарт ISO-образ — это неформальный термин для обозначения образа оптического диска, содержащего файловую систему стандарта ISO 9660. Для маркирования задействованы все исполняемые и библиотечные файлы, связанные с ОС Linux. Для этого позднее будут разобраны концепции таблиц символов и строк файловой системы.

В языке низкоуровневого программирования Ассемблер понятие «эквивалентные инструкции» подразумевает одиночные инструкции или последовательности инструкций, выполняющие одну и ту же операцию, и имеющие одинаковую длину. Если количество эквивалентных инструкций равно N, то при замене одной из этих инструкций на их эквивалент можно вложить $log\_2\ N$ битов скрытого сообщения. Имея вышеперечисленные параметры, необходимо построить соответствующий алгоритм порядка действий, при котором будет показан общий путь стеговложения в защищаемой ОС Linux(см. рис. 3).

Классификация данных по типу и категориям

С анализом информации внутри компании отлично справляется такое средство мониторинга информации как SIEM-система. Данная система способна обрабатывать информацию с различных источников: межсетевые экраны, IPS/IDS, средства антивирусной защиты, сканеры уязвимости, системы контроля целостности, операционные системы и базы данных. На основе анализа данных из этих источников выявляются отклонения от нормального функционирования, заданного критериями безопасности, и в случае обнаружения происходит оповещение администратора безопасности [1].

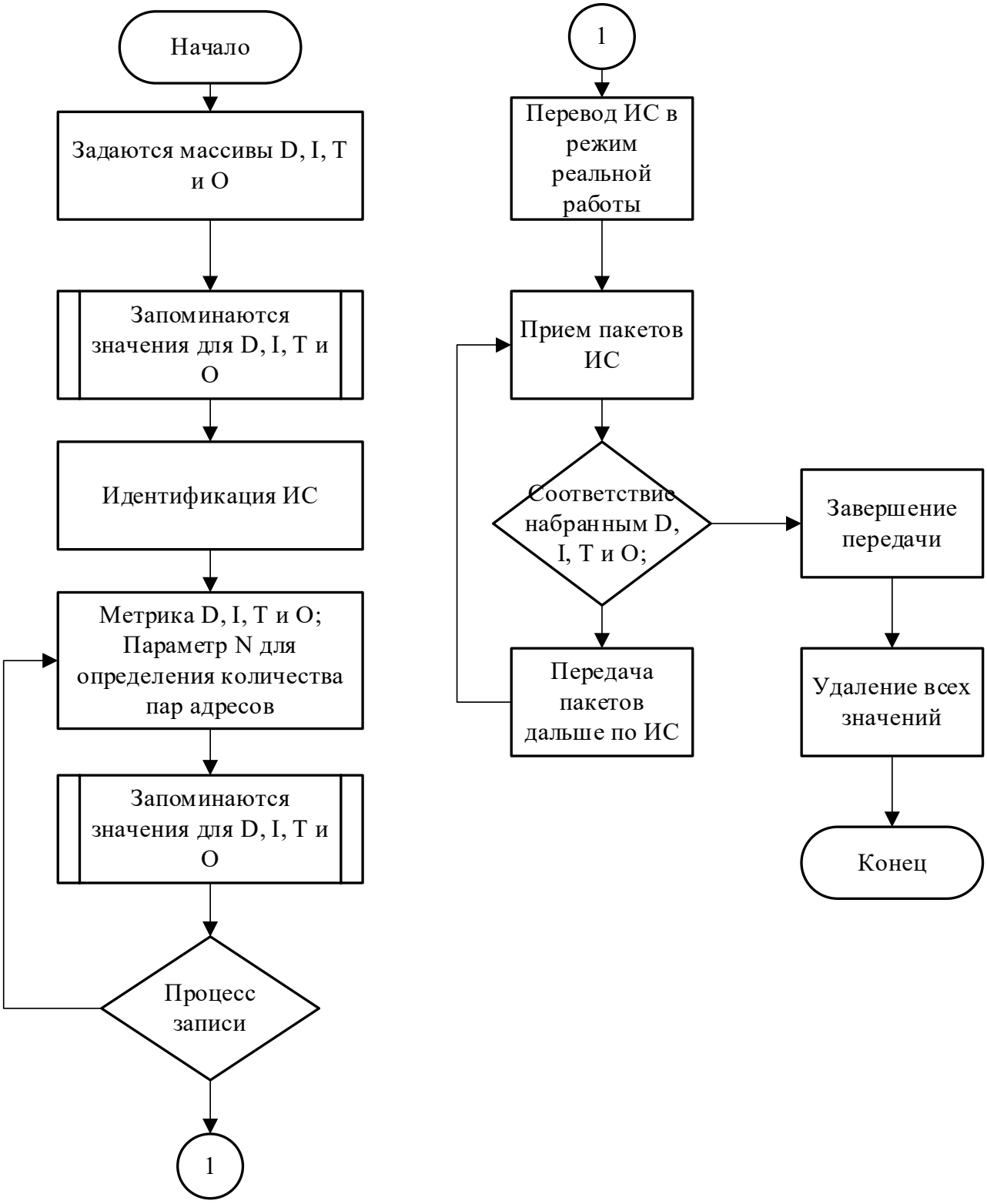
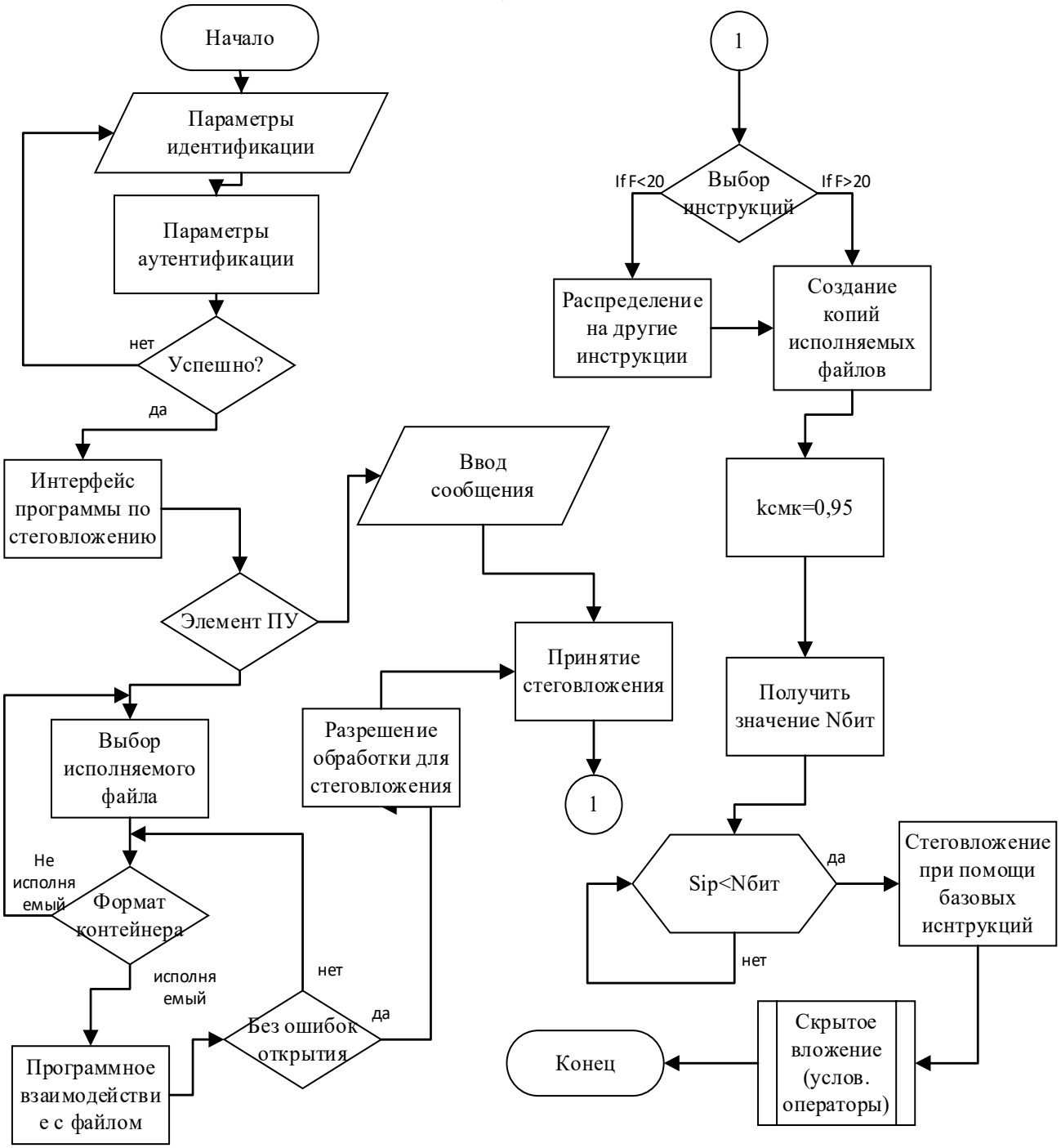


Рис. 2. Блок-схема алгоритма способа реализации протоколирования в агенте, функционирующем в ОС Linux



а)

Рис. 3. Блок-схема.
а) работы ПО для внедрения стеговложения в файлы ОС Linux,

б)

Рис. 3. Блок-схема.
б) алгоритма понижения рисков обнаружения стеговложения при использовании генерации агента.

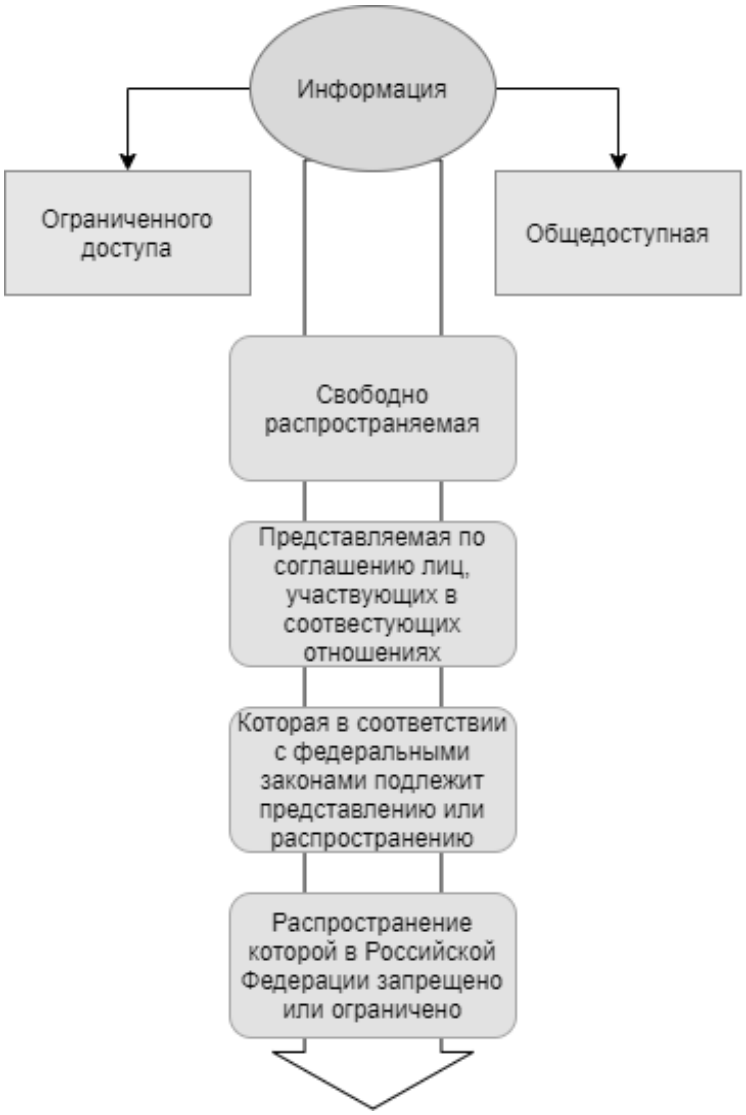


Рис. 4. Классификация информации

При построении эффективной SIEM-системы необходимо определить и классифицировать информацию, которая будет обрабатываться данной системой, и ее значимость для компании. Для этого необходимо рассмотреть подробнее понятие информация и какой она бывает в отношении утечек и злоумышленников [2,3,6].

Для корректного определения обрабатываемых и хранящихся данных, особенно если речь идет о защите информации, необходимо обратиться к федеральному закону 149-ФЗ «Об информации, информационных технологиях и о защите информации» [4]. Данный федеральный закон разделяет информацию, в зависимости от степени доступа к ней, на общедоступную информацию, а также на информацию, доступ к которой ограничен федеральными законами. По порядку ее представления или распространения информацию можно разделить на: информацию, свободно распространяемую;

1. информацию, которая в соответствии с федеральными законами подлежит предоставлению или распространению;
2. информацию, для которой запрещено или ограничено распространение на территории Российской Федерации.

Данные, которые следует считать информацией ограниченного доступа, можно определить надлежащим федеральным законом (к примеру, случаи с ПДн, банковской или адвокатской тайной), или же данная задача решается обладателем данных, который независимо систематизирует их в соответствии с описанным свойствам (коммерческая тайна или секреты производства) [5].

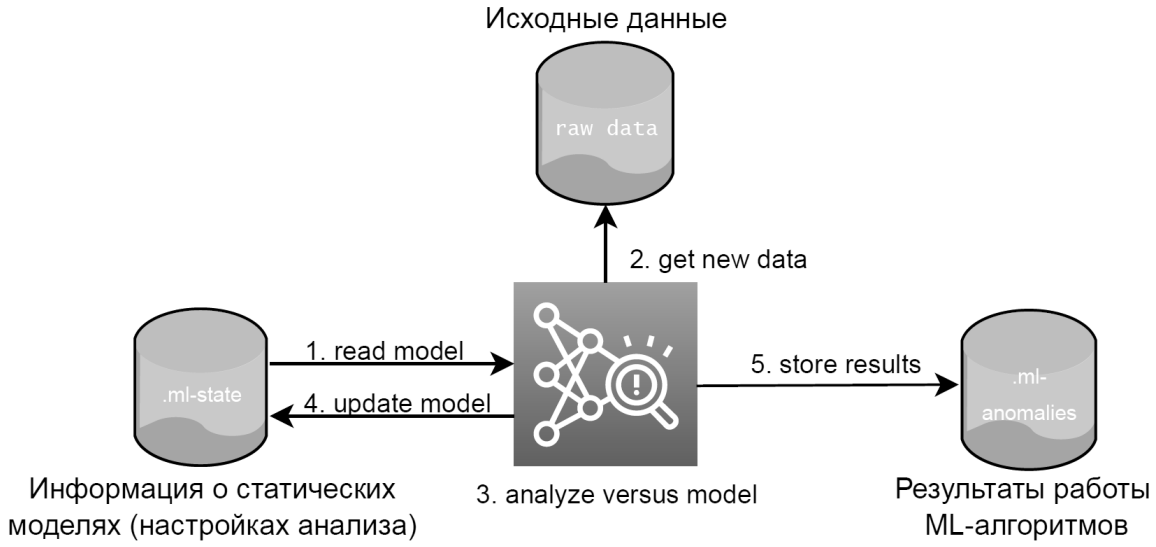


Рис. 5. Машинное обучение в Elasticsearch

Вне зависимости от подходящего метода соотношения имеющихся данных к информации ограниченного доступа, следует ясно осознавать условия и ограничения, предъявляемые к распространению таких данных, а также всю ответственность за вероятные утечки. Развертывание SIEM-системы в компании сможет помочь не только осуществить необходимые действия по защите информации, но также избежать неконтролируемое распространение конфиденциальных данных.

Также крайне важно то, что практически все SIEM-системы помогают реализовать ретроспективный анализ и организовать полноценное расследование в случае, если инцидент действительно случился. Принимаемые в настоящее время нормативные документы ФСБ России, ФСТЭК России, Банка России (в особенности по вопросам предоставления защищенности критически важных ИС) как правило учитывают обязательное внедрение процессного подхода к построению систем защиты информации и мониторинга событий. Значительная часть заинтересованности при этом уделяется обрабатыванию инцидентов и разбору последствий от реализованных утечек. Данная политика в результате обязана послужить причиной неотвратимости наказания, минимизации ущерба и снижению рисков повторения инцидентов [7–9].

Возможность оценки рисков с помощью машинного обучения

Машинное обучение в SIEM-системах может предоставить некие «рекомендации» для действий в будущем, основанные на опыте предыдущих событий, которые взяты с разных компонентов информационной инфраструктуры.

Машинное обучение открывает для компании весьма большие возможности. Как и любая другая технология, машинное обучение требует грамотного управления. Чтобы получить корректно работающую систему, необходимо обучить систему с помощью алгоритмов. Алгоритмы необходимо обрабатывать на входных данных и анализировать. Иначе говоря, необходимо предоставить системе качественную выборку данных и объяснить, как ее обработать, а затем убедиться в том, что система выдвигает правильные гипотезы по отношению к тем или иным явлениям. На рисунке 5 отражен алгоритм машинного обучения:

Алгоритмы машинного обучения делятся на две категории — «с учителем» и «без учителя». Для алгоритмов «без учителя» не нужна первоначальная выборка данных. Для алгоритмов «с учителем» систему нужно обучать с помощью различных методов. В Elastic Stack используется алгоритм из категории «с учителем» [19]. Процесс обучение и анализа эффективности работы модели строится на основе вышеприведенной схемы, состоящей из четырех этапов (Рисунок 6):

1. Определить задачу для ML. Необходимо понять, что модель должна выявлять, установить тип обучение.
2. Выбрать и преобразовать исходные данные. Необходимо подготовить набор данных, на котором модель будет обучаться.
3. Обучить и оценить модель.
4. Использовать обученную модель и прогнозировать.

В Elastic Stack основными возможностями машинного обучения являются:

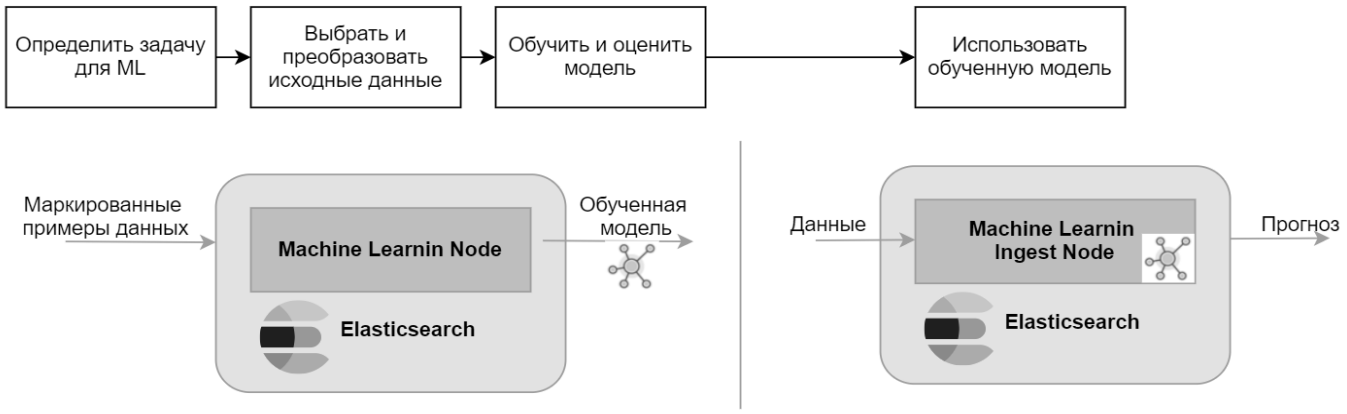


Рис. 6. Общий подход при работе с машинным обучением

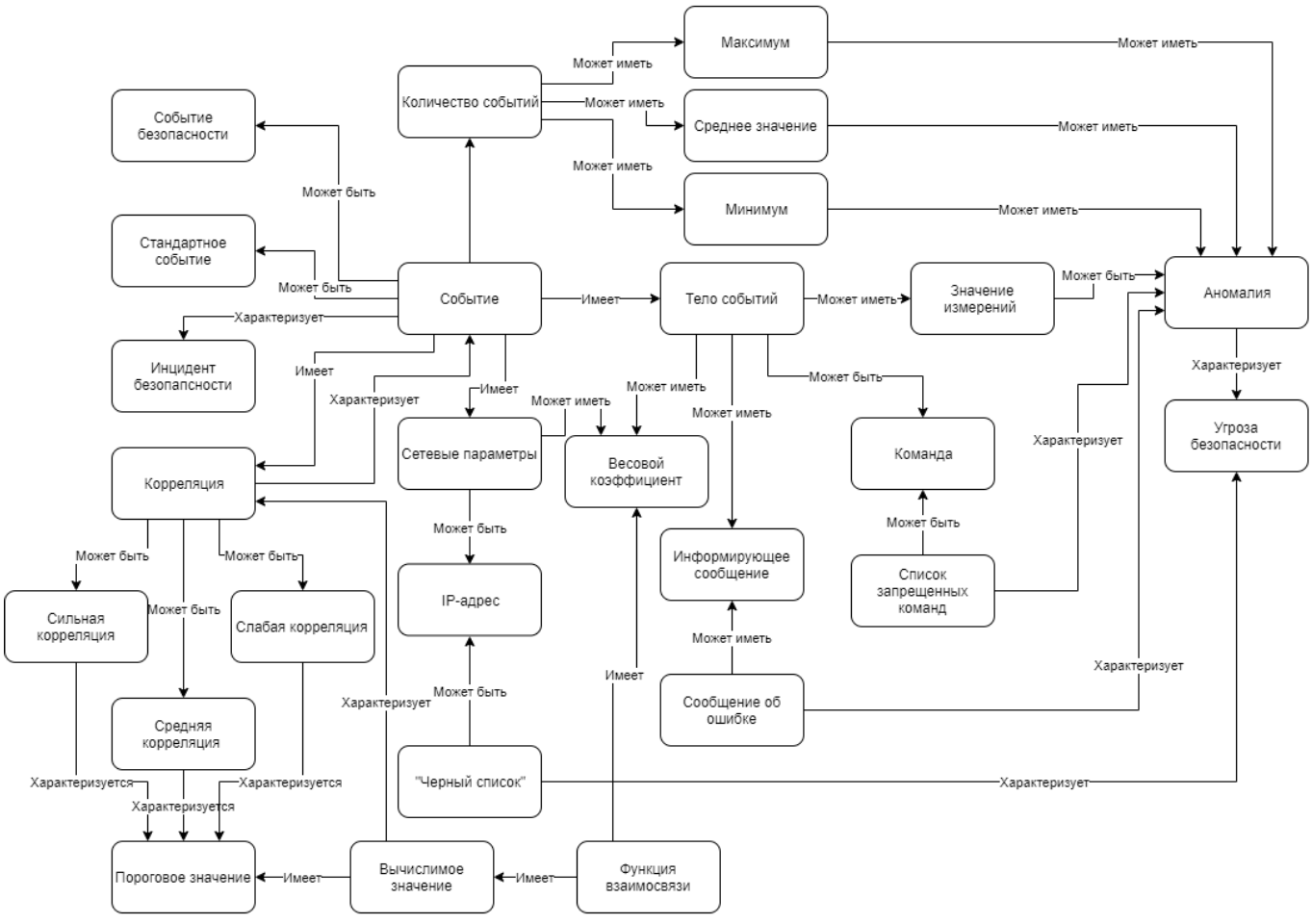


Рис. 7. Модель выявления угроз безопасности в SIEM-системе

- выявление аномалий (в режиме онлайн и с отправкой оповещений);
- прогнозирование (максимум на 8 недель вперед) [19].

Алгоритм машинного обучения изучает данные из индексов Elasticsearch. При этом управлять заданиями для анализа можно как через веб-интерфейс Kibana, так и через API.

Изучив возможности стека Elastic Stack, можно сделать вывод, что с помощью встроенных инструментов данной SIEM-системы можно разработать методику оценки рисков.

Предлагаемая методика

Преимуществом SIEM-систем является то, что они способны выявлять и в дальнейшем проводит расследование инцидентов безопасности. Для этого они используют механизм корреляции событий. Механизм корреляции заключается в том, что система ищет общие атрибуты сбытый и связывает события в значимые кластеры. Центральными концептами являются «событие» (как основная единица работы SIEM) и «угроза безопасности». В системе также целесообразна корреляция событий на основе правил, заключающаяся в создании «шаблонов» событий, при соответствии которым события маркируются как небезопасные и SIEM-система генерирует уведомление об инциденте безопасности. Для обнаружения аномалий в количестве событий и в значениях измерений, передаваемых в теле событий, целесообразен статистический подход [12].

Следовательно, для выявления аномалий и угроз безопасности к SIEM-системе формируются следующие требования:

- корреляция событий на основе правил;
- корреляция событий на основе статистических методов.

В работе [3,6,10] была описана и отражена в виде схемы модель выявления угроз безопасности (Рисунок 7).

В области управления рисками активно распространяются модели машинного обучения, но препятствуют этому чрезмерные трудозатраты и материальные издержки, связанные с внедрением и поддержкой. Сложнее всего — подготовить для модели выборку для обучения. Ниже приведены требования, которым должна отвечать обучающая выборка:

- достаточный объем выборки (если отсутствуют универсальные критерии достаточности);
- историчность (не менее года);
- однозначность классификации наблюдений на основе специально разработанных правил, которые наиболее применимы к выборке и соответствуют целям применения модели;
- однородность статистики [12].

Благодаря SIEM-системе, можно не просто выявить аномальные события, но и спрогнозировать их появление в будущем. Ниже будет описан алгоритм выявления аномалий и их прогнозирование [9].

В процессе анализа важным аспектом является выявление тенденций и закономерностей. При использовании anomaly-detection заметные тенденции данных определяются автоматически после нескольких циклов, с применением тенденции линейного роста и цикличных гармоник [14]. Чем больше данных будет подано в систему, тем точнее будет прогнозирование. По мере увеличение данных функция распределения вероятности станет более полной и гибкой.

В Elastic Stack аномалия — отклонение значения контролируемой функции от рассчитанных нормальных границ. Чем менее вероятно такое отклонение, тем более высокий уровень критичности получит аномалия. Движок машинного обучения использует сочетание разных алгоритмов: кластеризация, различные типы декомпозиции временных рядов, байесовское иерархическое моделирование и корреляционной анализ.

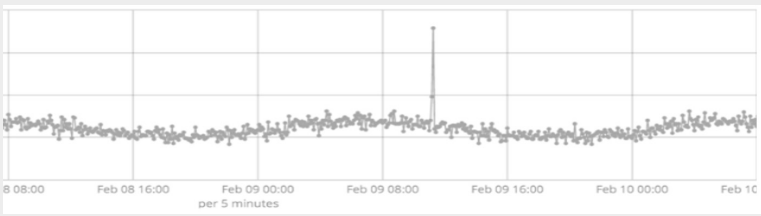
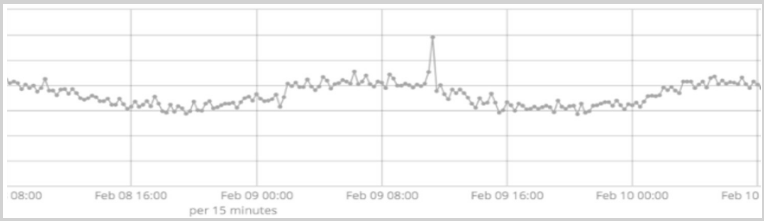
В Elastic Stack возможен анализ аномалий трех видов: анализ одной метрики (Single Metric), анализ нескольких метрик (Multi Metric), анализ корреляции метрик (Population-анализ).

В первых двух видах каждая метрика анализируется в изолированной среде, то есть поведение иных метрик не учитывается. Чтобы включить в расчёты корреляцию различных метрик, используется Population-анализ. Анализ изменений одной единственной метрики — самый простой способ. После создания, алгоритм сразу же начинает искать аномалии.

Параметр Aggregation устанавливает подход к поиску аномалий [14]. Например, при значении «Min» аномальными будут считаться значения ниже типичных. Параметр Bucket span отвечает за гранулярность промежутков событий на таймлайне, по которым будет вестись анализ. Можно довериться автоматическому значению или выбрать вручную. При слишком низкой гранулярности можно пропустить аномалию. В таблице 1 указаны примеры значений параметра Bucket span и графики, выводимые при этих значениях.

Для эффективного анализа очень важно правильно выставить длительность собранных данных. В ходе анализа алгоритм старается выявить повторяющиеся

Таблица 1. Значение Bucket span и соответствующие графики

Значение параметра Bucket span	Пример графика
5 минут	
15 минут	
60 минут	

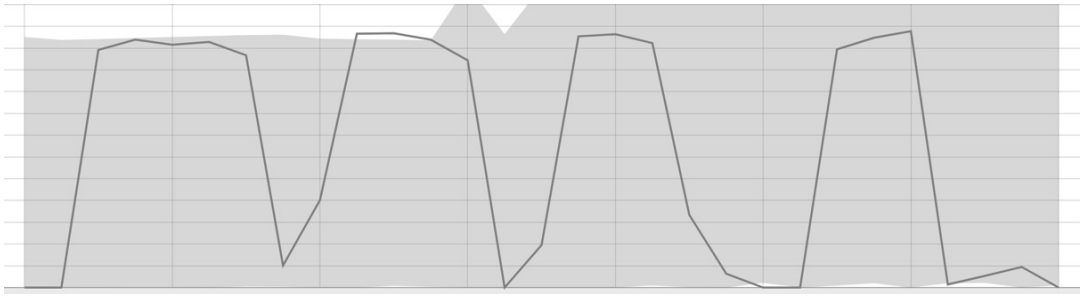


Рис. 8. Пример графика с базовыми линиями при небольшом отрезке данных

промежутки и рассчитать доверительные промежутки. На основании этого он выявляет аномалии — нетипичные отклонения от обычного поведения метрики. Например, на рисунке 8 отображены базовые линии при небольшом отрезке данных:

После нескольких циклов алгоритм сокращает вероятность отклонения от нормы. На рисунке 9 отображены линии после того, как алгоритм распознал закономерности.

После запуска задания машинного обучения алгоритм определяет аномальные отклонения от нормы и ранжирует их по вероятности аномалии.

Таким образом, методика оценки рисков с использованием SIEM-системы может состоять из следующих этапов:

1. Определение материальных и нематериальных активов.
2. Подготовка SIEM-системы, настройка необходимых индексов в SIEM-системе.
3. Настройка интеграции SIEM-системы с компонентами информационной инфраструктуры.
4. Настройка заданий для обнаружения аномалий, выбор необходимых метрик для отслеживания, установка пороговых значений.
5. Мониторинг событий, аномалий, инцидентов информационной безопасности в SIEM-системе, исследование причин возникновений, построение возможных тактик реализации угроз.

Анализ вероятности возникновения и уровень критичности аномалии и инцидента информационной безопасности в SIEM-системе.

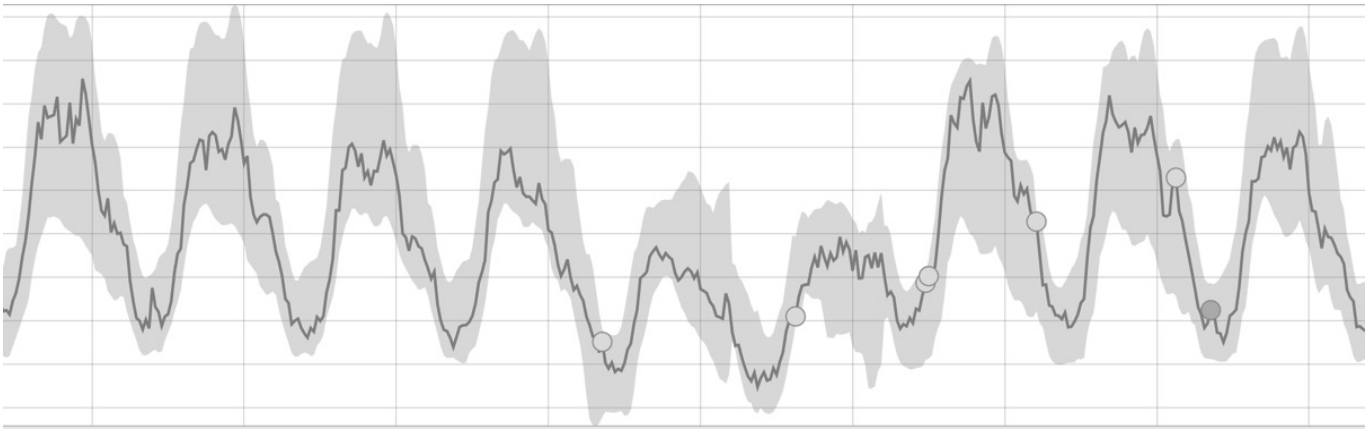


Рис. 9. Пример графика спустя несколько циклов

Заключение

В последствии была предложена методика оценки рисков с использованием анализа событий в SIEM-системе. На практическом примере в системе Elastic Stack было продемонстрировано, как система способна выявлять аномальную активность, рассчитывать уровень критичности событий и производить прогноз событий в будущем. Благодаря данной системе можно получить более детальное описание и характеристики событий, что в последствии можно использовать в процессе оценки рисков.

ЛИТЕРАТУРА

1. Базовая модель угроз безопасности персональных данных при их обработке в информационных системах персональных данных (выписка). ФСТЭК России, 2008 год
2. ГОСТ Р 53113.1–2008 «Защита информационных технологий и автоматизированных систем от угроз безопасности, реализуемых с использованием скрытых каналов».
3. Штеренберг, С.И. Разработка методики построения доверенной среды на основе скрытого программного агента. Часть 1. исследование / С.И. Штеренберг, А.В. Красов // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки.— 2021.— № 2.— С. 14–20.— DOI 10.46418/2079–8199_2021_2_2.
4. Свидетельство о государственной регистрации программы для ЭВМ № 2020617876 Российская Федерация. Модель угроз и нарушителя: № 2020616749: заявл. 29.06.2020: опубл. 15.07.2020 / А.В. Красов, А.А. Миняев, А.И. Пешков; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ).
5. Красов А.В., А.М., Гельфанд А.М., Коржик В.И. [и др.]. Построение доверенной вычислительной среды. СПб: Индивидуальный предприниматель Петрив Роман Богданович, 2019.— 108 с.— ISBN978–5–6043143–2–6.
6. Штеренберг С.И., Красов А.В., Разработка методики построения доверенной среды на основе скрытого программного агента. Часть 2. Тестирование и оценка эффективности // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2021. № 3. С. 3–8.
7. Шелухин О.И., Канаев С.Д. Стеганография. Алгоритмы и программная реализация. М.: Горячая линия — Телеком, 2017,— 592 с.
8. Буйневич М.В., Израилов К.Е. Антропоморфический подход к описанию взаимодействия уязвимостей в программном коде. Часть 1. Типы взаимодействий // Защита информации. Инсайд. 2019. № 5 (89). С. 78–85.
9. Ушаков И.А., Обнаружение инсайдеров в корпоративной компьютерной сети на основе технологий анализа Больших данных // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2019. № 4. С. 38–43.
10. Штеренберг С.И., Данилова Ю.С., Разработка методики внедрения и выявления эффективности siem-системы // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2020. № 3. С. 40–45.
11. Сагдеев А.К., Штеренберг И.Г., Штеренберг С.И., Виноградова О.М., Разработка блока обнаружения и коррекции ошибок для устройства диагностирования каналов передачи цифровой информации // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2020. № 1. С. 15–24.
12. Миняев А.А., Красов А.В., Сахаров Д.В., Метод оценки эффективности системы защиты информации территориально-распределенных информационных систем персональных данных // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 2: Искусствоведение. Филологические науки. 2020. № 1. С. 29.

13. Е.А. Дешевых, И.А. Ушаков, А.А. Чечулин, Интеграция SIEM-систем с системами корреляции событий безопасности, основанных на технологии больших данных // Информационные технологии в управлении (ИТУ-2016) Материалы 9-й конференции по проблемам управления. Председатель президиума мультиконференции В.Г. Пешехонов. — 2016. — С. 684–687.
14. Штеренберг С.И. Обнаружение вторжений в распределенных информационных системах на основе методов скрытого мониторинга и анализа больших данных / диссертация. канд. техн. наук, г. Санкт-Петербург, 2018. — 182 с.

Журнал «Современная наука: актуальные проблемы теории и практики»

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича