DOI 10.37882/2223–2966.2021.07.15

АЛГОРИТМ СИСТЕМЫ МОНИТОРИНГА УТЕЧЕК КОНФИДЕНЦИАЛЬНЫХ ДОКУМЕНТОВ

ALGORITHM OF THE CONFIDENTIAL DOCUMENT LEAKAGE MONITORING SYSTEM

D. Zakharchenko
A. Borshevnikov
I. Chernousov
Yu. Dobrzhinskii

Summary. The article discusses the algorithm of a new system for monitoring leaks of confidential electronic documents. Provides an overview of data leakage tracking mechanisms used in DLP systems. The general scheme of the system is given, its individual elements are described. The applied leak tracking algorithms are discussed.

Keywords: information security, security policy, corporate data protection, data leaks, leak detection.

Захарченко Даниил Владимирович
Ассистент, Дальневосточный федеральный университет (г. Владивосток)
daniilZakharchenko@gmail.com
Боршевников Алексей Евгеньевич
Старший преподаватель, Дальневосточный федеральный университет (г. Владивосток)
borshevnikov.ae@dvfu.ru
Черноусов Илья
Дальневосточный федеральный университет (г. Владивосток)
chernousov.is@dvfu.ru
Добржинский Юрий Вячеславович
К.т.н, доцент, Дальневосточный федеральный университет (г. Владивосток)
dobrzhinskii.yv@dvfu.ru

Аннотация. В статье рассматривается алгоритм новой системы мониторинга утечек конфиденциальных электронных документов. Приводится обзор механизмов отслеживания утечек данных, применяемых в DLP-системах. Рассматривается общая схема системы, описываются отдельные ее элементы. Разбираются применяемые алгоритмы отслеживания утечек.

Ключевые слова: информационная безопасность, политика безопасности, защита корпоративных данных, утечки данных, обнаружение утечек.

В настоящее время все сильнее возрастает количество утечек корпоративных данных. По результатам исследования компании Infowatch за 2018 год около 42% утечек данных в мире происходили через сеть, среди которых немалую часть составляли корпоративные документы [1]. Более того, если рассматривать статистику за последние несколько лет, можно заметить тенденцию к росту доли утечек через сеть. Также стоит отметить, что, хотя общая тенденция утечек данных по вине или неосторожности внутреннего нарушителя в 2019 году снизилась до 54% от общего числа, в финансовом секторе ситуация диаметрально противоположная — число таких утечек выросло с 45% в 2018 году до 63% в 2019 [2]. В связи с этим остро встает вопрос о необходимости средств борьбы с утечками данных.

Основным решением, предотвращающим утечки данных, являются системы предотвращения утечек информации или так называемые DLP-системы (от англ. Data Leak Prevention, DLP) [3].

Прежде всего эти системы нацелены на борьбу со внутренним нарушителем, они используются, когда первоочередной задачей является защита конфиденциальной информации от внутренних атак. В общем случае для правильного внедрения DLP-систем в информационную систем от ответственного за информационную безопасность в организации требуется четкое понимание следующих тезисов:

- Какими методами и средствами внутренний нарушитель может организовать утечку критически важных данных.

- Какие информационные каналы могут быть использованы для проведения кражи информации.
- Какая информация подлежит защите, прежде всего исходя из правила «Какой урон будет нанесен моей организации при компрометации этой информации?».
- Какова разница между стоимостью внедрения DLP-системы на данные информационные ресурсы и ущербом от раскрытия этих конфиденциальных данных, что подлежит защите в первую очередь.

Основная проблема при внедрении систем подобного характера состоит в том, что комплекс предотвращения утечек информации должен уметь отличать в автоматическом режиме конфиденциальную информацию от не конфиденциальной. В зависимости от сферы деятельности организации и штата сотрудников информационная система может быть, как локальной, так и распределенной, автоматизированной или автоматической, что в свою очередь несет разный объем информации, который необходимо проанализировать и предотвратить возможную утечку в случае обнаружения нарушения. Если не выделить критические данные для защиты, а просто анализировать всю информацию внутри информационной системы, то скорее всего при рассмотрении небольших организаций столкнемся с проблемой перегрузки вычислительных мощностей, избыточной нагрузки на обеспечивающий персонал [4, 5].

DLP-система работает в связке с обеспечивающим персоналом, специалистом по информационной безопасности организации, который будет корректировать работу комплекса в зависимости от изменяющейся политики информационной безопасности компании, появления новых угроз, увеличения вычислительных мощностей компании, перехода в другую сферу деятельности, что повлечет за собой новый уровень структуры информационной системы компании, изменение числа сотрудников, появление новых возможных злоумышленников с отличными от предыдущих уровнями подготовленности.

Вся система предотвращения утечек конфиденциальных данных строится вокруг математического алгоритма и полученного на его основе программного решения, в задачи которого входит обнаружение и анализ информации, которой необходима защита. В основе большинства DLP-систем лежат две технологии: лингвистический анализ и статистический подход к выявлению утечек [6, 7].

Лингвистический метод анализа. Лингвистический метод анализа работает непосредственно с информацией в документе либо файле. Этот подход позволяет уходить от проблемы излишнего расходования вычислительных мощностей на обработку.

В общем случае лингвистическая аналитика включает в себя два важных аспекта:

- Морфологический анализ — отбор согласно абсолютно всем допустимым словоформам данных, какие следует защитить от компрометации;
- Семантический анализ — осуществление поиска критически важной информации при помощи проверки вхождения таковой в обрабатываемый файл, проверка качественных характеристик файла, анализ контекста использования данных.

Статистический метод анализа

Противоположностью лингвистического анализа с позиции точности определения утечки является статистический подход к обнаружению нарушения существующих правил разграничения доступа.

Алгоритм всей процедуры может быть описан следующим образом:

1. Происходит набор критических данных, подлежащих защите. Основным преимуществом здесь является возможность параллельной категоризации данных по степени возможного ущерба организации от его раскрытия;
2. Каждый документ делится на части, от которых вычисляется хеш-функция;
3. Каждое значение хеш-функции определенного фрагмента заносится в базу;
4. При обнаружении в подлежащих проверке частях информационной системы документа, берется фрагмент документа и сверяется с эталонным, который хранится в заранее созданной на этапе 3 базе;
5. При совпадении значений хеш-фунций от проверяемого и эталонного фрагментов система помечает этот файл как конфиденциальный, оповещает об инциденте ответственного за информационную безопасность организации и действует согласно заранее разработанной политике ИБ.

Сама же DLP система осуществляет контроль за периметром, включающем в себя как контроль сетевого трафика, так и контроль устройств, с которых сотрудник может входить в сеть и которыми он может воспользоваться на работе.

Сетевой уровень контроля

Сетевой уровень контроля представляет собой подход защиты критически важных ресурсов путем кон-

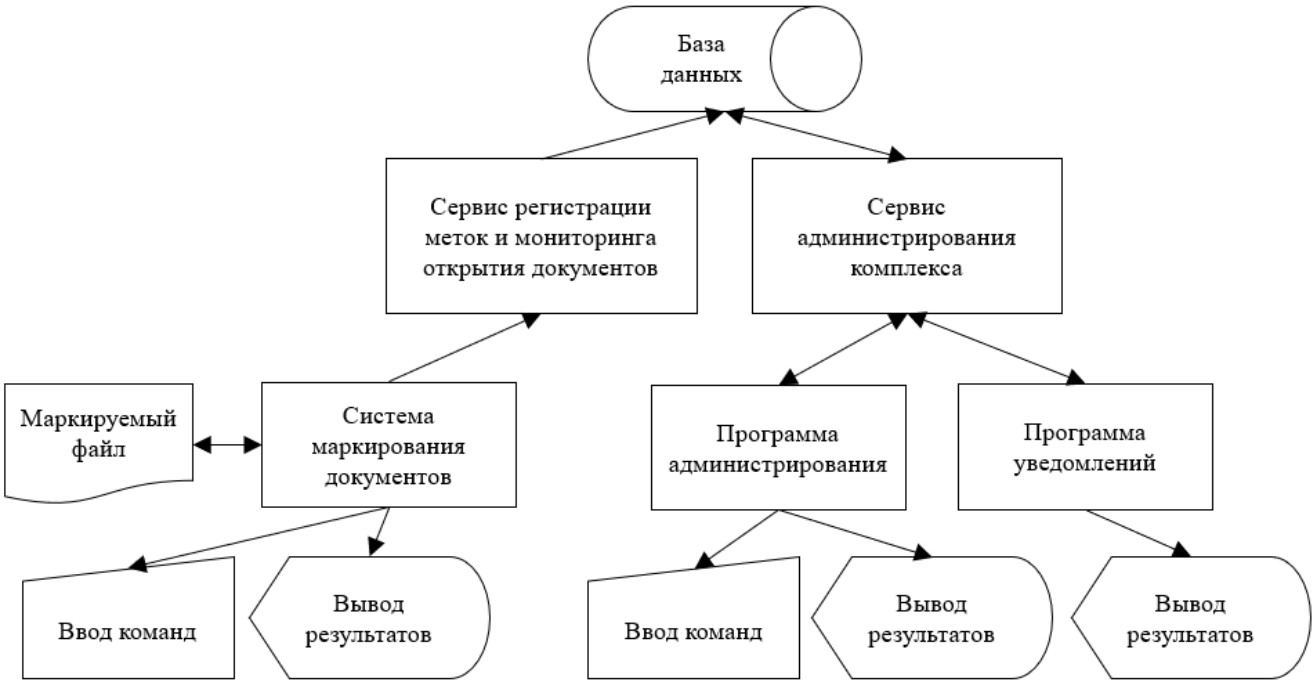

Рис. 1. Схема программного комплекса

троля трафика, пересекающего границы информационной системы.

Технически этот уровень организовывается на специальных прокси-серверах, серверах почтовой почты или специально выделенных серверах.

Хостовый уровень контроля

Хостовый уровень контроля фактически представляет собой подход полного контроля за рабочим устройством пользователя, будь то сотрудник (внутренняя угроза) организации или пользователь (внешний нарушитель). Такая система позволяет отслеживает копирование файлов на съемные носители, вспомогательную аппаратуру и средства (печатные и копировальные устройства), в процессе работы пользователя делать снимки экрана для контроля правомерности работы, как и было сказано ранее немаловажным преимуществом является возможность обработки шифрованного трафика, который поступил конечному получателю в лице хоста.

Естественно, что для качественной и всесторонней обработки данных DLP-системой необходима организация предотвращения утечек с использованием всех вышеперечисленных компонент с централизованным управлением в лице ответственного за информационную безопасность в организации [8].

Для обеспечения дополнительного контроля за инцидентами утечек конфиденциальной информации можно применять подход, основанный на фиксации события открытия документов, содержащих такую информацию, на устройствах, которые не имеют разрешения на данное действие. В случае открытия документа он отправляет информацию о данном событии, которое при получении анализируется с целью определить по полученной информации, разрешено ли открытие документа на данном устройстве. Если устройство не входит в список доверенных, то формируется отчет об инциденте, который поступает ответственному лицу. Подобный подход позволит обнаруживать случаи утечки информации даже в том случае, если DLP-система не обнаружила утечку данных.

В данной работе приводится алгоритм работы системы, которая реализует подобный подход.

Рассматриваемая система состоит из 6 компонент, представленных на рисунке 1:

1. Система маркировки.
2. Сервис регистрации меток и мониторинга открытия документов.
3. Сервис администрирования комплекса
4. Программа администрирования
5. Программа уведомлений
6. База данных

Пользователь напрямую взаимодействует с 3 системами: системой маркирования документов, программой администрирования, программой уведомлений. Первая предназначена для внедрения в документы, которые содержат в себе конфиденциальную информа-

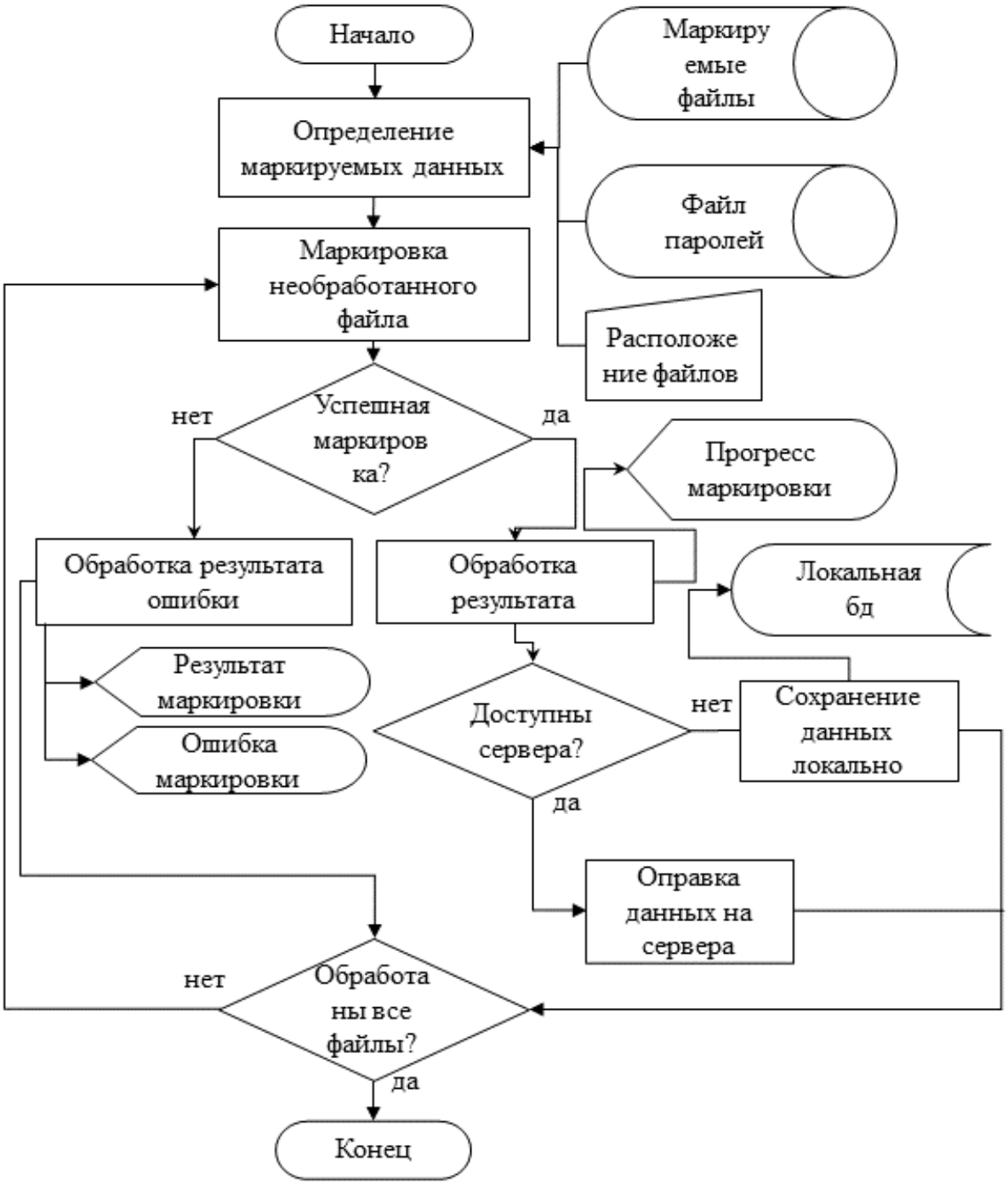

Рис. 2. Схема системы маркировки

цию, специальных механизмов, которые будут осуществлять отправку сообщений об открытии этих документов. Программа администрирования предназначена для настройки системы, а также предоставления отчетов о произошедших событиях, таких как открытие документа или проведение маркировки нового документа. Программа уведомлений предназначена для экстренного уведомления ответственного лица об обнаруженном инциденте утечки документа.

Сервис регистрации меток и мониторинга открытия документов предназначен для фиксации информации от маркировщика о новом промаркированном файле и фиксации событий открытия документов.

Сервис администрирования комплекса предназначен для настройки всего комплекса, а также сбора информации о обнаруженных утечках. Сервис администрирования состоит из следующих компонент:

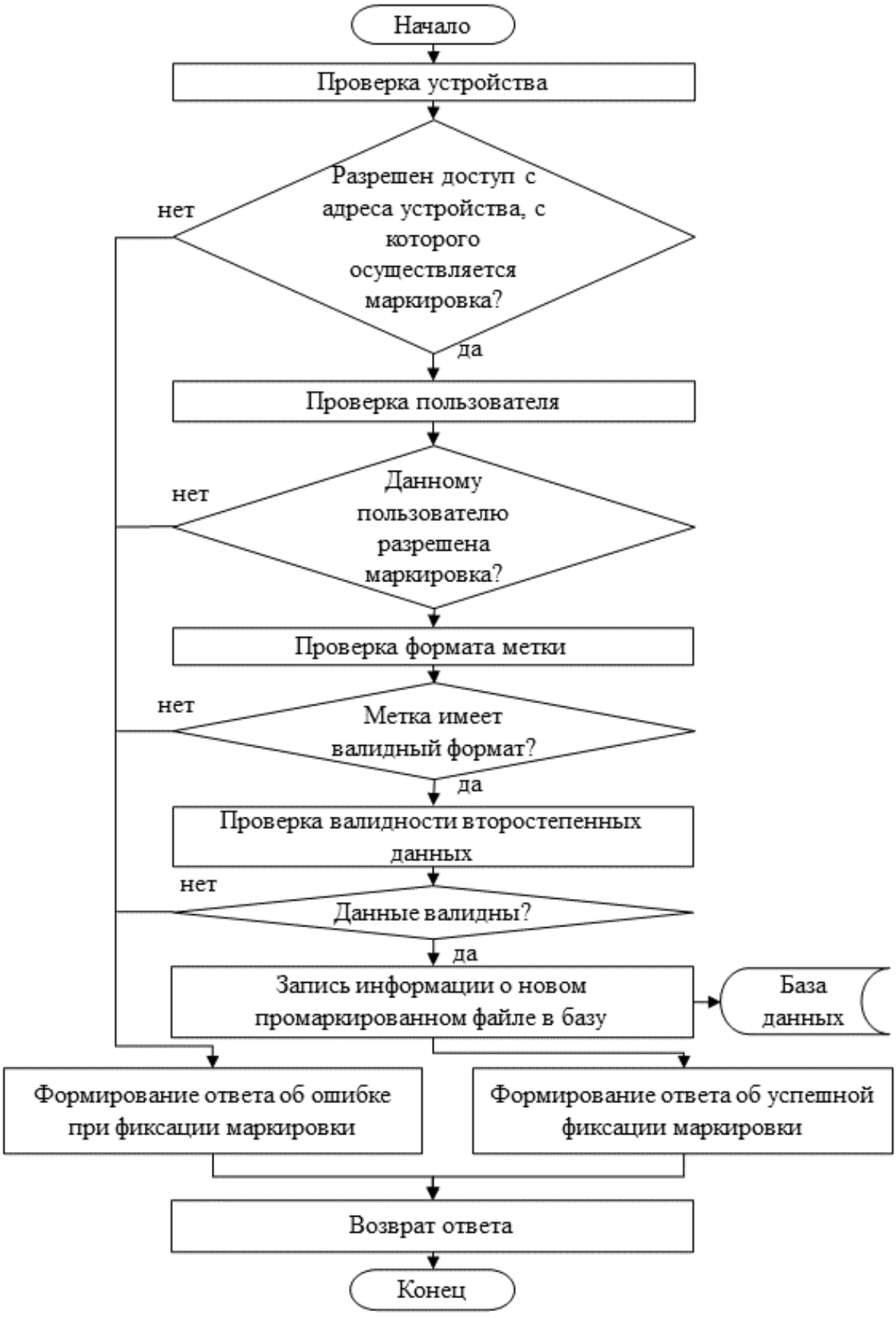

Рис. 3. Алгоритм регистрации меток

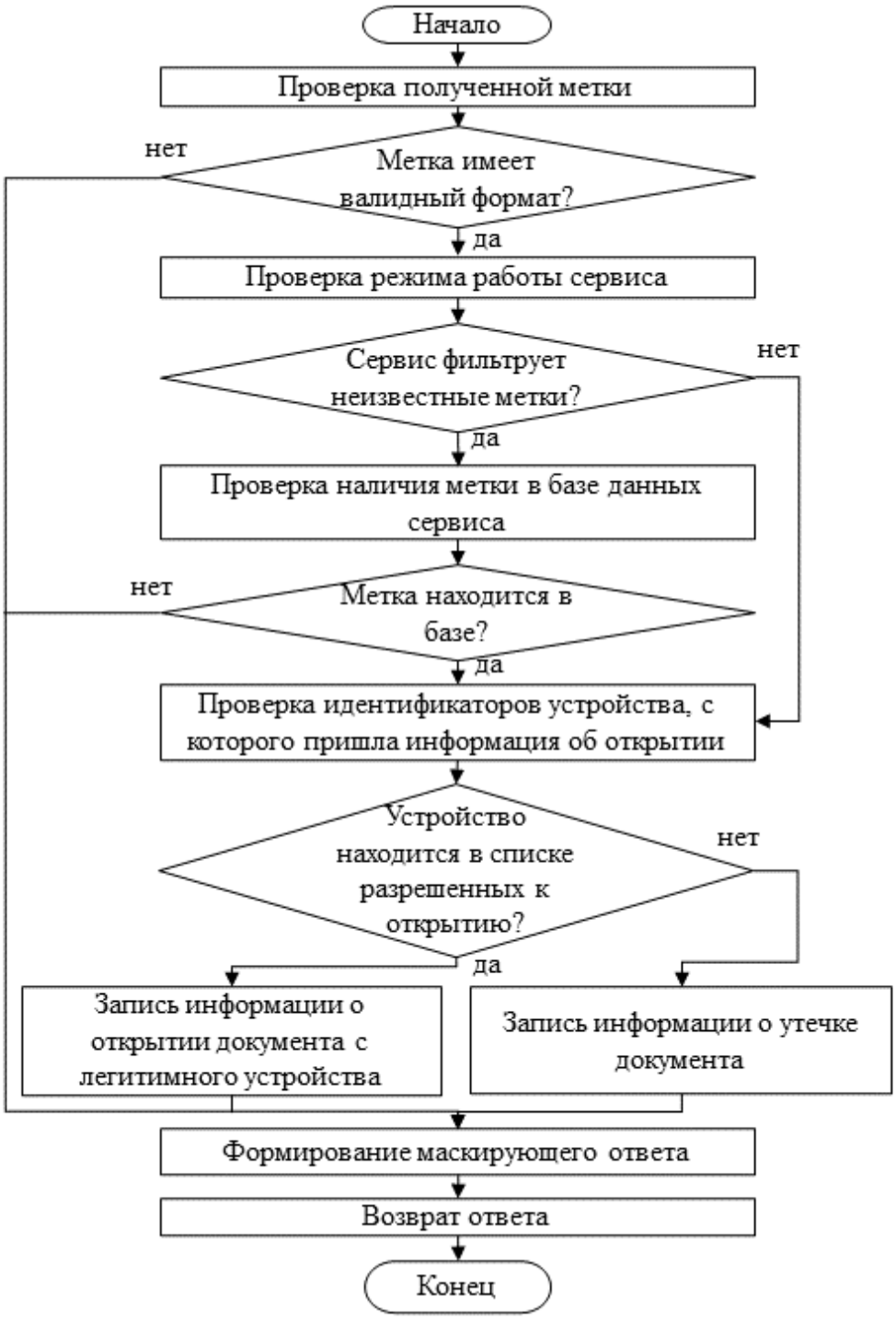

Рис. 4. Алгоритм фиксации события открытия документа.

- Регистрация пользователей, которым разрешено маркировать файлы.
- Регистрация устройств, с которых разрешено открытие документов.
- Прекращение слежения за документами.
- Отслеживание обнаруженных событий открытия.
- Отправка отчетов о промаркированных файлах.
- Отправка отчетов об открытых файлах.
- Отправка уведомлений об инцидентах утечек файлов.

База данных предназначена для хранения информации о настройках системы, промаркированных файлах, событиях открытия документов.

Элементы «Сервис администрирования комплекса», «программа администрирования», «программа

уведомлений» предназначены для настройки системы и информирования пользователя о произошедших событиях и не будут подробно рассматриваться в данной работе.

Система маркирования документов работает по алгоритму, схема которого представлена на рисунке 2.

Основная идея заключается в присвоении новым документам уникального идентификатора на основе идентификаторов пользователя и устройства, на котором он был промаркирован, а также некоторого случайного числа. Эта информация будет применяться для определения документа при инциденте утечки. Само приложение, предназначенное для маркировки, представляет собой клиентское приложение, позволяющее пользователю выбрать файл или папку с файлами для проведения маркировки.

Дополнительно пользователь указывает файл паролей от документов в случае, если данные хранятся в зашифрованном виде. Далее программа начинает поочередно промаркировать каждый документ, внедряя в него код, который будет заставлять документ требовать удаленный ресурс при каждом открытии.

Маркировка документа подразумевает под собой внедрение в документ специального идентификатора и функционала, который при открытии промаркированного документа будет отправлять сообщения сервисам комплекса. По данному идентификатору в дальнейшем возможно будет восстановить информацию о документе, устройстве, на котором оно было промаркировано, пользователе, осуществившем маркировку. Текущая реализация маркировки требует чтения документов через ограниченное число приложений для возможности отправки сообщений.

В случае успешной маркировки документа информация о нем, включая идентификаторы файла, информацию об устройстве, на котором он был промаркирован, информацию о пользователе, промаркировавшем его, группу, к которой привязан документ, информацию о метке, дату маркировки, будет переслана на сервер для последующей идентификации полученных сообщений. В случае, если на устройстве отсутствует доступ к серверу, то будет произведена запись в локальное хранилище, а программа при появлении доступа отправит сохраненные данные на сервер.

Весь прогресс маркировки отображается пользователю и в случае невозможности маркировки отдельных документов пользователь получит об этом уведомления.

Для фиксации информации о промаркированных документах используется сервис регистрации меток и мониторинга открытия документов.

Сервис представляет собой веб-приложение, обрабатывающее приходящие ему запросы по определенным алгоритмам. В нем применяются два основных алгоритма: алгоритм регистрации меток, представленный на рисунке 3, и алгоритм фиксации события открытия документа, представленный на рисунке 4. Каждый из этих алгоритмов на вход получает http запрос, и возвращает ответ запросившему.

Обработка запроса фиксации меток начинается с проверок, валидный ли пользователь отправил этот запрос. Для этого применяются специальный идентификатор пользователя, адрес устройства в сети и идентификатор зарегистрированного устройства, пересылаемые вместе с запросом.

Если данные неправильные, то пользователю будет отправлено сообщение об ошибке. Далее осуществляется проверка корректности информации, пересылаемой от маркировщика, а именно: проверяется корректность метки, применяется ли отправленная версия метки на текущем сервисе, проверяется дата. Если все данные корректны — осуществляется фиксация метки в базе данных.

При обработке запроса на фиксацию события открытия документа сервис принимает от метки сообщение о данном событии. В ответ сервис сгенерирует маскировочное сообщение и перешлет его в любом случае. Первоначально сервис осуществляет проверку на соответствие полученной метки. Если полученная метка отличается по версии от тех, которые обрабатываются сервисом — то фиксации не происходит. Далее проверяется, строгость проверки сервисом меток. Если на сервисе включен режим строгой проверки, то система ожидает, что данная метка сохранена в системе. Если режим отключен, то система будет принимать любые метки подходящей версии. Далее осуществляется проверка по идентификатору устройства, полученному вместе с запросом. Если идентификатор устройства не зафиксирован в системе, но фиксируется и заносится в базу событие утечки документа. В ином случае фиксируется событие открытия документа.

В рамках данной работы была рассмотрена возможная реализация системы мониторинга утечек конфиденциальных электронных документов.

Рассмотренная реализация, обладая определенными недостатками, уже может на практике применяться для мониторинга. Основным ограничением применяе-

мой системы является сам механизм срабатывания метки, ограниченный наличием на открывающем устройстве определенных программ и подключения к сети интернет, в случае утечки во внешнюю сеть. Данная система может применяться как вспомогательная система для отслеживания утечек конфиденциальных документов на предприятиях, а также для контроля за распространением документов внутри предприятия.

ЛИТЕРАТУРА

1. Комиссаров Е.А., Смагина И.В. Внутренние утечки информации //Образование и наука без границ: социально-гуманитарные науки. — 2019. — № . 11. — С. 44–48.
2. Исследование утечек конфиденциальной информации из организаций финансового сегмента в 2019 г. [Электронный ресурс]. — Режим доступа: https://www.infowatch.ru/resources/analytics/reports/21649, свободный (дата обращения: 01.03.2021).
3. Морозова Н.С. Проблемы современных систем предотвращения утечек данных с конечных точек сети //Безопасность информационных технологий. — 2011. — Т. 18. — № . 4. — С. 138–143.
4. Контур информационной безопасности Searchinform [Электронных источник] — Режим доступа: https://searchinform.ru/products/kib/, свободный (дата обращения: 01.03.2021).
5. Андрианов В.И., Сивков Д.И., Юркин Д.В. Методика внедрения системы предотвращения утечек информации DLP в коммерческую организацию для информационной сети с использованием больших данных //Вестник Брянского государственного технического университета. — 2020. — № . 6 (91).
6. Филяк П.Ю. и др. Применение infowatch DLP-системы для обеспечения безопасности в сети internet //Информация и безопасность. — 2018. — Т. 21. — № . 4. — С. 566–571.
7. Лопатин А.Г., Бобров Н.В. Выработка рекомендаций по внедрению DLP-системы при реализации политики информационной безопасности предприятия //Вестник Международной академии системных исследований. Информатика, экология, экономика. — 2016. — Т. 18. — № . 1. — С. 153–159
8. Андриянова Т.А., Саломатин С.Б. DLP: снижение риска утечки конфиденциальной информации Банка //Системный анализ и прикладная информатика. — 2017. — № . 3.

г. Владивосток