моделей для кластеризации больших объемов данных является технология распределённых вычислений MapReduce.

Модель программирования MapReduce в общем случае содержит несколько функций Map (распределитель) и Reduce (редуктор). Функция Map своими входными параметрами принимает пары key/value. Значение value каждого объекта представляет собой строку, состоящую из координат n-мерного вектора параметров. Входные данные распределяются между Mapper-ами. Начальное множество кластерных центров передаются каждому Mapper-у. Каждый центроид задается своим идентификатором в качестве ключа key и значением центроида в качестве значения value. Map функция, сравнивая значение очередного объекта со значениями кластерных центров, определяет, его принадлежность к заданным кластерам. Выходными значениями для каждого Mapper-а, пара ключ/значение для каждого объекта, где ключ – это номер ближайшего центроида, значение – вектор параметров в n-мерно векторе.

После работы функции Map, значения записываются в распределенную файловую систему. Между стадией Map и Reduce промежуточные данные сортируются и тассуются. Входными данными Reduce-ов будут выходные данные Mapper-ов.

Функция Reduce содержит все пары ключ/значение, которые получены от функции Map. Для всех пар ключ/значение, имеющих один и тот же ключ, их значение сохраняются в некотором итераторе, затем функция Reduce вычисляет среднее значение, которое имеет один и тот же ключ. Таким образом, получаем новые центроиды, и выходные данные передаются функциям Map. Этот процесс продолжается до тех пор, пока центры массы не перестанут изменяться.

Для решения задачи кластеризации с помощью парадигмы MapReduce можно выбрать платформу Hadoop, которая предназначена для создания и запуска распределенных приложений, работающая с большими объемами данных. Одним из основных компонентов платформы Hadoop является его файловая система HDFS (Hadoop distributed file system). Основной функцией распределенной файловой системы является разделение данных конкретного пользователя на блоки данных и репликация заданных блоков по локальным дискам узлов кластера.

Список источников

1. Батуркин С.А., Гостин А.М., А.В. Пруцков и др. Система внутреннего тестового контроля знаний РГРТУ: методические указания/ Рязан. гос. радиотехн. ун-т. – Рязань, 2007. – 68 с.

2. Мансурова М.Е., Шоманов А., Тулепбергенов Б., Параллельный алгоритм кластеризации для обработки гиперспектральных изображений на основе MapReduce Hadoop// Международная конференция “ИКТ: образование, наука, инновации”, Алматы, 20-21 мая 2013 г. – с. 56-61.

В.Д. Ленчук, Д.О. Маркин

СИСТЕМА МОНИТОРИНГА ОБМЕНА ЭЛЕКТРОННЫМИ СООБЩЕНИЯМИ УДАЛЕННЫМИ ПОЛЬЗОВАТЕЛЯМИ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ОСОБЕННОСТЕЙ СИСТЕМЫ ДОМЕННЫХ ИМЕН

(Академия Федеральной службы охраны Российской Федерации)

Особое место среди источников утечек конфиденциальной информации занимает электронная почта. Хоть ее доля и уменьшается, электронная почта по-прежнему имеет весомый процент среди каналов утечек, о чем свидетельствует диаграмма, представленная на рисунке 1 [1].

Рисунок 1 – Распределение утечек по каналам
за первые полугодия 2014 и 2015 годов

В процессе информационного взаимодействия при работе системы обмена электронными сообщениями существенную роль на безопасности информации оказывают факторы и угрозы, определяемые стандартами [2, 3], относящиеся к классу сетевых, то есть, реализуемых с использованием протоколов межсетевого взаимодействия. Особое значение в сетевом взаимодействии играет система адресации и маршрутизации, важным элементом которой является система доменных имен. К числу угроз, использующих систему доменных имен, относятся угрозы, основанные на модификации пакетов DNS-транзакций, которые относятся к классу угроз, направленных на создание в сети ложного маршрута. Их потенциальная опасность заключается в возможности перехвата данных, передающихся между клиентами сетевых сервисов и серверами этих сервисов. Очевидно, что использование определенных возможностей системы доменных имен, позволяющих получить доступ к информационному взаимодействию удаленных пользователей, позволит реализовать функции мониторинга, аудита и фильтрации передаваемых данных, включая и сообщения электронной почты.

Для реализации предлагаемой системы предлагается использовать следующие элементы:

1. Почтовый сервер на базе операционной системы Centos.
2. Почтовый клиент, использующий ПО Mozilla Thunderbird.
3. Точка мониторинга на базе операционной системы Debian.

Пример приведенной системы представлен на рисунке 2.

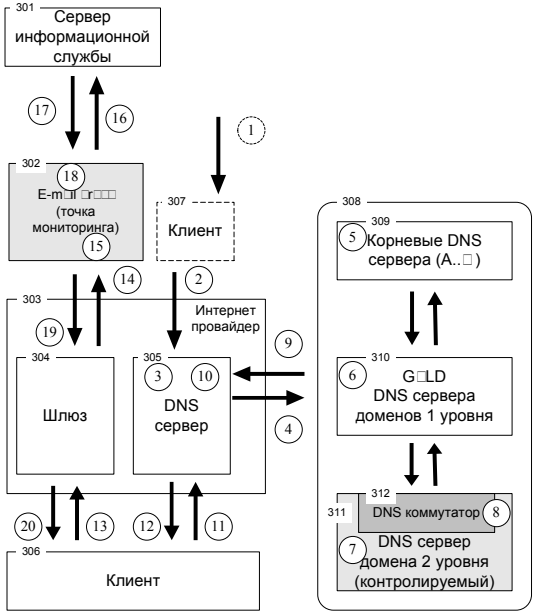

Рисунок 2 – Структурная схема, реализующая мониторинг обмена электронными сообщениями

Сущность реализации предлагаемой системы заключается в следующем: Точка мониторинга будет являться "прозрачной" для почтовых клиента и сервера, что будет достигаться с помощью преобразования сетевых адресов (NAT-Network Address Translation). То есть при обращении к серверу клиент будет отправлять запрос на точку мониторинга, на которой будет происходить контроль передаваемого трафика и отправка трафика дальше на сервер. При этом там же будет обеспечиваться подмена сетевых адресов и портов, для того чтобы точка оставалась прозрачной и для клиента, и для сервера и могла функционировать в сетях любой сложности.

Таким образом, возникает возможность осуществления контроля сетевого взаимодействия клиента и заданных информационных служб вне зависимости от их расположения и топологии компьютерной сети. Кроме этого появляется возможность мониторинга и контроля сетевого взаимодействия клиента и за-

данных информационных служб как на этапе установления сеанса соединения, так и на этапе информационного обмена, что позволяет обеспечить фильтрацию трафика на наличие сведений конфиденциального характера и сведений, составляющих государственную тайну, предотвращение несанкционированного доступа к информации.

Особенностями предлагаемой системы являются:

1. Реализация доступа к информационному обмену удаленных пользователей и удаленных информационных ресурсов.
2. Возможность получения доступа и мониторинг трафика удаленных пользователей и информационных ресурсов, использующих SSL/TLS (работающих по протоколу HTTPS).
3. "Прозрачный" режим работы точки мониторинга.
4. Отсутствие нарушений в режиме функционирования и структуре распределенной инфокоммуникационной сети (ИТКС провайдера, программно-аппаратная среда провайдера и пользователя, программно-аппаратная среда серверов информационных ресурсов).
5. Независимость от расположения пользователей и информационных ресурсов.

Литература

1. InfoWatch Исследование утечек информации за первое полугодие 2015 года // InfoWatch: Аналитика [Электронный ресурс] : сайт. – Электрон. дан. – 2003–2015. – Режим доступа: http://www.infowatch.ru/analytics/reports/16340. – Дата обращения: 07.10.2015.

2. ГОСТ Р 51275–2006. Защита информации. Объект информатизации. Факторы, воздействующие на информацию. Общие положения. – Введ. 2006.12.27. – Москва : Федеральное агентство по техническому регулированию и метрологии, 2007. – 8 с. – (Национальный стандарт Российской Федерации).

3. ГОСТ Р ИСО/МЭК 15408-1 Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 1.Введение и общая модель. – Москва. – 47 с. – (Государственный стандарт Российской Федерации).

И.В. Лёзина, Н.А. Николаева

ИДЕНТИФИКАЦИЯ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ МНОГОСЛОЙНЫМ ПЕРСЕПТРОНОМ

(Самарский национальный исследовательский университет имени академика С.П. Королёва)

Идентификация законов распределения является распространенной задачей, решение которой может быть применено в различных областях приборостроения для оценки брака на производстве, для анализа экономических процессов, для расчетов времени обслуживания заявок, для создания методик об-