Министерство науки и высшего образования Российской Федерации

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Научно-образовательный центр «Высшая ИТ школа»

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА**

СНИЖЕНИЕ КОГНИТИВНОЙ НАГРУЗКИ НА ИНЖЕНЕРОВ SOC ЧЕРЕЗ АВТОМАТИЗАЦИЮ ПРОЦЕССОВ ИНТЕГРАЦИИ ИСТОЧНИКОВ ДАННЫХ.

Шамов Егор Сергеевич

Направление подготовки 09.03.04 Программная инженерия

Направленность (профиль) «Программная инженерия»

Научный руководитель

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ З.А. Полни Меня,

Заполни мои регалии и место работы

*подпись*

«\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_ 20 \_\_\_ г.

Научный консультант

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.С. Зоркин,

ведущий инженер, ООО "ТЦР"

*подпись*

«\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_ 20 \_\_\_ г.

Автор работы

студент группы № 972103

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Е. С. Шамов

*подпись*

«\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_ 20 \_\_\_ г.

Оглавление

[1 Введение в предметную область 6](#_Toc195825875)

[1.1 Устройство SIEM 6](#_Toc195825876)

[1.1.1 Компоненты SIEM 6](#_Toc195825877)

[1.1.2 Процессы SIEM 6](#_Toc195825878)

[2 Анализ существующих процессов 7](#_Toc195825879)

[2.1 Процесс настройки сбора событий 7](#_Toc195825880)

[2.1.1 Создание коллектора 7](#_Toc195825881)

[2.1.2 Прямая интеграция с Kafka 7](#_Toc195825882)

[2.2 Процесс настройки нормализации событий 7](#_Toc195825883)

[2.2.1 Создание парсера 7](#_Toc195825884)

[2.2.2 Написание правил нормализации 7](#_Toc195825885)

[3 Автоматизация процесса 8](#_Toc195825886)

[3.1 Автоматизация поиска свободного порта 8](#_Toc195825887)

[3.1.1 Варианты решения проблемы 8](#_Toc195825888)

[3.1.2 Обоснование выбранного алгоритма 8](#_Toc195825889)

[3.2 Автоматизация создания топика 8](#_Toc195825890)

[3.2.1 Создание топика в собственном кластере 8](#_Toc195825891)

[3.2.2 Создание топика в кластере KaaS решения 8](#_Toc195825892)

[3.3 Автоматизация создания коллектора 8](#_Toc195825893)

[3.3.1 Требуемые параметры для запуска 8](#_Toc195825894)

[3.3.2 Генерация конфигурационного файла 8](#_Toc195825895)

[3.3.3 Развертывание сервиса 8](#_Toc195825896)

[3.3.4 Предоставление информации о развернутом экземпляре пользователю 8](#_Toc195825897)

[3.4 Автоматизация создания парсера 8](#_Toc195825898)

[3.4.1 Генерация конфигурационных файлов из шаблона 8](#_Toc195825899)

[3.4.2 Механизм развертывания парсера 8](#_Toc195825900)

Глоссарий

Введение

Security Operations Center (SOC) – это специализированное подразделение, которое осуществляет мониторинг и реагирование на инциденты информационной безопасности в режиме реального времени.

Вся работа SOC построена на анализе данных с различных компонентов наблюдаемых систем. Внутри SOC любое событие из какой-либо наблюдаемой системы мы называем логом. Лог (или событие) — это небольшая запись, которую система или приложение делает при наступлении каких-то значимых ситуаций. Например: вход пользователя в систему, ошибка в приложении, инициация нового сетевого соединения и т.д. В контексте информационной безопасности (далее ИБ) такие логи помогают понять, что происходило в системе и выявить подозрительную активность или инциденты.

Для сбора, хранения и управления логами в SOC используется система управления информацией и событиями безопасности (англ. Security Information and Event Management System, SIEM). Помимо перечисленного, в задачи SIEM входит обогащение логов дополнительной информацией из других источников, а также последующий корреляционный анализ, который позволяет выявлять сложные цепочки взаимосвязанных событий, указывающих на потенциальные инциденты ИБ.

Поскольку SIEM-система обрабатывает десятки, а иногда и сотни тысяч событий в секунду, в её архитектуре нередко используется промежуточный слой — шина или буфер обмена данными, например Apache Kafka. Такая шина позволяет эффективно распределять и балансировать потоки логов, обеспечивая равномерную доставку событий между компонентами системы. Также Kafka играет ключевую роль в обеспечении отказоустойчивости, масштабируемости и надежности обработки данных: она временно хранит события, предотвращая их потерю при высоких нагрузках или сбоях отдельных компонентов SIEM, и позволяет обрабатывать события асинхронно, но почти в режиме реального времени.

Не все источники событий могут напрямую отправлять данные в Kafka - особенно это касается устаревших систем или решений, не поддерживающих современные протоколы интеграции. Для таких случаев используется промежуточный компонент - коллектор. Коллектор принимает логи от источников по различным протоколам и передаёт далее в Kafka, никак не видоизменяя исходное сообщение. Такой подход обеспечивает обратную совместимость с существующей инфраструктурой SIEM.

В последнее время из-за быстрого роста компании каждый месяц появляется один или два новых источника, уже сейчас количество источников исчисляется десятками. В связи с этим замечены две проблемы.

1. Создание коллектора в текущий момент занимает до двух суток времени, и на протяжении этого времени инженер, ответственный за создание коллектора, должен держать контекст в голове.
2. Инвентаризация источников логов ведется вручную параллельно процессу настройки приема событий из источника.

Указанные проблемы обусловлены недостаточным уровнем автоматизации существующих процессов. Цель данной работы – разработать и внедрить автоматизированный процесс настройки приема логов из различных источников, включающий в себя развёртывание и конфигурирование коллекторов логов и других вспомогательных элементов.

Для достижения этой цели в работе необходимо решить следующие задачи.

1. Проанализировать существующий процесс развертывания коллекторов, выявить операции, подлежащие автоматизации.
2. Спроектировать и разработать инструменты автоматизации, которые позволят исключить или минимизировать ручные операции
3. Внедрить разработанное решение в инфраструктуру SOC и провести оценку его эффективности.

# Введение в предметную область

## Инфраструктура SIEM

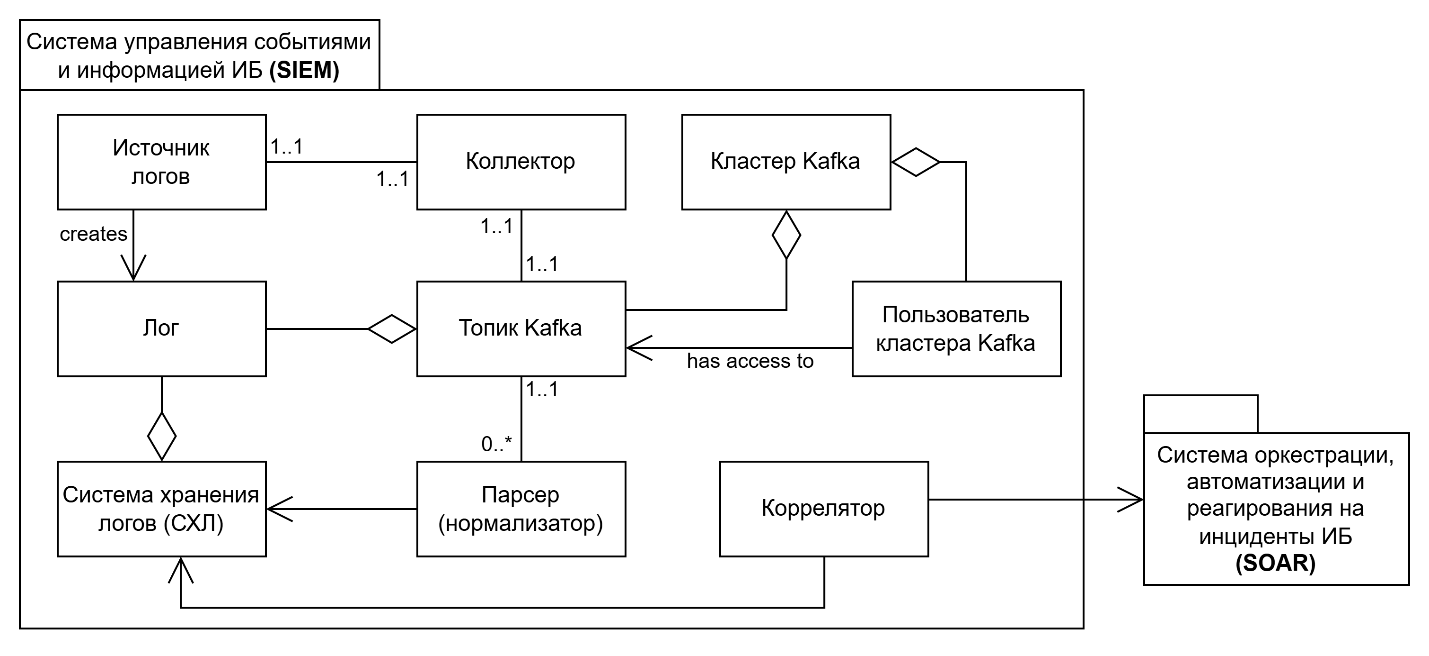


Рисунок 1 – Предметная область SIEM

На рисунке 1 отражены основные сущности системы управления информацией и событиями информационной безопасности. Рассмотрим их подробнее.

### Источники логов

К источникам логов могут относиться разные системы, но чаще это сетевое оборудование, антивирусное ПО, приложения для аудита безопасности операционных систем, системы управления учетными данными (англ. Identity and Access Management System, IdM/IAM) и другие сервисы, которыми пользуется большое количество сотрудников (хоть и не всегда явно).

Как было сказано во введении, не всегда источники логов умеют передавать данные через современные протоколы взаимодействия. Например, большое количество серверного оборудования предлагает экспорт логов только через UDP (англ. User Datagram Protocol, сетевой протокол транспортного уровня). А некоторые приложения (например, система управления проектами Jira) сохраняют данные аудита только в базу данных (БД), в таком случае, единственный путь получения данных – читать их прямо из БД. Для таких случаев необходим какой-то сервис-адаптер, преобразующий один формат данных в другой. Таким сервисом выступает коллектор.

### Коллекторы

Коллектор — это сервис, который принимает события от различных источников по разным протоколам и интерфейсам, и передаёт их в стандартизированном виде в систему Apache Kafka. Основная задача коллектора заключается в обеспечении совместимости устаревших или ограниченных в плане интеграции систем с современной инфраструктурой SIEM. При этом коллектор не изменяет смысл или содержимое события, а только адаптирует транспортный и форматный уровень передачи данных, что позволяет централизованно и унифицированно обрабатывать логи в SIEM-системе.

Коллектор должен уметь пропускать через себя большой поток данных в единицу времени, оставаясь при этом легковесным и нересурсозатратным. Среди инструментов с открытым исходным кодом можно выделить следующие инструменты, подходящие на роль коллектора:

* Filebeat;
* FluentD;
* Vector.dev.

Сюда что-то стоит дописать?

### Apache Kafka

### Компоненты SIEM

Блаблаблаблабла

### Процессы SIEM

Дорогой дневник. Сегодня был неплохой день. Жить можно, но осторожно. Я камыыыыыыш. Главным признаком того, что этот диплом золото,является его автор. Он выдающийся челопиздрик. Хорошо, что не выпадающий. Люблю его, а ты ??

# Анализ существующих процессов

## Процесс настройки сбора событий

### Создание коллектора

### Прямая интеграция с Kafka

Блаблаблабла

## Процесс настройки нормализации событий

### Создание парсера

Блаблаблабла

### Написание правил нормализации

#### Обогащение данных

Блаблабла

#### Исключение данных

Блаблабла

# Автоматизация процесса

## Автоматизация поиска свободного порта

блаблабла

### Варианты решения проблемы

Блаблабла

### Обоснование выбранного алгоритма

Блаблабла

## Автоматизация создания топика

### Создание топика в собственном кластере

Блаблаб

### Создание топика в кластере KaaS решения

Блаблабла

## Автоматизация создания коллектора

### Требуемые параметры для запуска

### Генерация конфигурационного файла

### Развертывание сервиса

### Предоставление информации о развернутом экземпляре пользователю

## Автоматизация создания парсера

### Генерация конфигурационных файлов из шаблона

### Механизм развертывания парсера

Заключение

В рамках данной работы …..

Список использованных источников и литературы

1. Site Reliability Engineering: How Google Runs Production Systems // Бетси Бейер, Нейл Ричард Мёрфи, Дэвид Рензин, Кент Кавахара и Стивен Торн, 2016.
2. UML. Основы. Третье издание. Краткое руководство по стандартному языку объектного моделирования // Мартин Фаулер, 2013.
3. Docker Documentation // Docker – [Б.м.]., 2024 – URL: [https://docs.docker.com](https://docs.docker.com/) (дата обращения: 20.06.2024).
4. What is SIEM? // Microsoft – [Б.м.]., 2025 – URL: https://www.microsoft.com/en-us/security/business/security-101/what-is-siem