Аналитическая статья
УДК 004.056; DOI: 10.61260/2218-13X-2025-1-120-134

МЕТОДИКА АНАЛИЗА ЖУРНАЛОВ СОБЫТИЙ ИНФОРМАЦИОННЫХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ ELASTICSEARCH ДЛЯ ОБНАРУЖЕНИЯ СИГНАЛОВ О ВРЕДОНОСНЫХ ДЕЙСТВИЯХ В ИНФОРМАЦИОННОЙ СИСТЕМЕ

✉**Дудников Иван Алексеевич;**
Шариков Павел Иванович;
Майоров Александр Владимирович;
Санкт-Петербургский государственный университет телекоммуникаций имени профессора М.А. Бонч-Бруевича, Санкт-Петербург, Россия
✉***van.dy@mail.ru***

Аннотация. Рассматривается использование стека ELK (Elasticsearch, Logstash, Kibana) для автоматизированного анализа журналов событий информационных систем с целью повышения эффективности обнаружения аномалий, указывающих на вредоносные действия. В качестве основного инструмента используется Elasticsearch, который позволяет эффективно хранить и анализировать большие объемы данных, а также интегрировать различные системы для мониторинга безопасности. Работа направлена на развитие методов корреляции событий и использования машинного обучения для автоматического выявления угроз в реальном времени. Особое внимание уделено оптимизации процессов мониторинга информационной безопасности, сокращению времени реакции на инциденты и улучшению точности диагностики угроз. Предложенный подход интегрируется в существующие инфраструктуры и адаптируется к меняющимся условиям, обеспечивая гибкость и эффективность работы с логами. В дальнейших исследованиях планируется провести экспериментальное применение метода и сравнение с другими решениями для оценки его эффективности.

Ключевые слова: анализ журналов событий, информационная безопасность, стек ELK, Elasticsearch, аномалии, машинное обучение, автоматизация, мониторинг

Для цитирования: Дудников И.А., Шариков П.И., Майоров А.В. Методика анализа журналов событий информационных систем с использованием Elasticsearch для обнаружения сигналов о вредоносных действиях в информационной системе // Науч.-аналит. журн. «Вестник С.-Петерб. ун-та ГПС МЧС России». 2025. № 1. С. 120–134. DOI: 10.61260/2218-13X-2025-1-120-134.

Analytical article

METHODOLOGY FOR ANALYZING EVENT LOGS OF INFORMATION SYSTEMS USING ELASTICSEARCH TO DETECT SIGNALS OF MALICIOUS ACTIVITIES IN INFORMATION SYSTEMS

✉**Dudnikov Ivan A.;**
Sharikov Pavel I.;
Maiorov Alexander V.
Saint-Petersburg State university of telecommunications named after professor M.A. Bonch-Bruevich, Saint-Petersburg, Russia
✉***van.dy@mail.ru***

Abstract. The work explores the use of the ELK stack (Elasticsearch, Logstash, Kibana) for automated analysis of event logs in information systems to improve the efficiency of anomaly detection, indicating malicious activities. Elasticsearch is used as the main tool, enabling efficient storage and analysis of large data volumes, as well as the integration of various systems for security monitoring. The paper focuses on the development of event correlation methods and the use

of machine learning for real-time threat detection. Special attention is given to optimizing information security monitoring processes, reducing response times to incidents, and improving threat diagnosis accuracy. The proposed approach integrates into existing infrastructures and adapts to changing conditions, ensuring flexibility and efficiency in working with logs. Future research will include experimental implementation of the method and comparison with other solutions to evaluate its effectiveness.

Keywords: event log analysis, information security, ELK stack, Elasticsearch, anomalies, machine learning, automation, monitoring

For citation: Dudnikov I.A., Sharikov P.I., Maiorov A.V. Methodology for analyzing event logs of information systems using Elasticsearch to detect signals of malicious activities in information systems // Scientific and analytical journal «Vestnik Saint-Petersburg university of State fire service of EMERCOM of Russia». 2025. № 1. P. 120–134. DOI: 10.61260/2218-13X-2025-1-120-134.

Введение

Анализ журналов событий информационных систем (или логов) является важным процессом в обеспечении безопасности, производительности и стабильности IT-инфраструктуры. Журналы событий содержат ценные данные о действиях пользователей, операциях приложений и сетевых взаимодействиях. Однако ручной анализ журналов событий становится непрактичным по нескольким причинам: во-первых, это экономически нецелесообразно, поскольку большие объемы данных требуют значительных затрат на привлечение специалистов и инфраструктуры для их обработки. Во-вторых, высок риск пропуска важной информации из-за человеческого фактора, особенно при обработке огромных массивов однотипных данных. Кроме того, ручной анализ занимает много времени, что может привести к задержкам в выявлении угроз и своевременном реагировании на инциденты. Автоматизация с использованием инструментов, таких как Elasticsearch, позволяет значительно повысить эффективность и точность обработки логов.

Elasticsearch – одна из популярных платформ для хранения и анализа логов, благодаря своей масштабируемости, скорости поиска и интеграции с другими инструментами, такими как Logstash для сбора данных и Kibana для визуализации. Этот набор инструментов, известный как стек ELK (Elasticsearch, Logstash, Kibana), стал стандартом для анализа журналов событий. В данной статье будет рассмотрено использование Elasticsearch и стека ELK для анализа журналов событий информационных систем с особым вниманием на выявление аномалий, которые могут свидетельствовать о вредоносных действиях.

Релевантные работы

Для анализа событий информационных систем, разработанных на платформе Java, стек ELK предоставляет мощные возможности, позволяющие автоматизировать обработку и анализ больших объемов данных. Одной из задач, для которых стек ELK может быть применен, является анализ байт-кода class-файлов и выявление аномалий, связанных с изменениями модулей или обфускацией. Это особенно актуально для Java-приложений, где большое количество файлов и сложные интеграции делают ручной анализ крайне трудоемким. Например, использование ELK позволяет автоматизировать процесс обнаружения подмены Java-модулей, что важно для предотвращения атак, направленных на замену модулей с вредоносными версиями. Применение скрытого цифрового водяного знака в байт-коде, устойчивого к обфускации, в сочетании с анализом логов помогает обнаруживать и предотвращать подобные атаки, что подтверждено в ряде исследований [1].

Другим примером использования стека ELK является анализ уязвимостей, таких как эксплойты в библиотеке Log4j, которые широко применяются в информационных системах. Подобные уязвимости требуют оперативного выявления подозрительных действий, например, аномальных вызовов функций или нехарактерного поведения приложения.

Настройка фильтров и визуализаций в Kibana позволяет автоматизировать этот процесс и оперативно генерировать оповещения о возможных угрозах [2].

Кроме того, ELK может быть использован для предотвращения атак, связанных с обфускацией байт-кода. Такие атаки направлены на сокрытие изменений в коде или нарушение работы встроенных защитных механизмов, таких как цифровые водяные знаки. Использование стека ELK позволяет обнаруживать аномалии, вызванные обфускацией, и предотвращать разрушение структуры данных и функциональности системы [3].

Протодьяконова М.С., Н.Н. Богдашина, С.Г. Ноговицын, Н.И. Иванов исследуют использование стека ELK для анализа сетевого трафика и обнаружения аномалий на веб-сайтах [4]. Авторы детально рассматривают особенности анализа веб-журналов, делают сравнительный анализ различных систем мониторинга и логирования. Они приводят практический пример настройки Logstash для сбора логов, использования Elasticsearch для их индексирования и Kibana для визуализации, подчеркивая важность быстрой и точной обработки данных. Особое внимание уделено плагину ElastAlert, который настраивается для отправки уведомлений при обнаружении подозрительных изменений в логах, что помогает оперативно реагировать на потенциальные угрозы. Авторы приходят к выводу, что применение этих технологий в информационных системах значительно повышает их устойчивость к кибератакам за счет интеграции эффективных инструментов анализа данных.

Котенко И.В., А.А. Кулешов, И.А. Ушаков предлагают системный подход к созданию платформы мониторинга и управления инцидентами безопасности на базе Elastic Stack [5]. В их работе проводится глубокий анализ современных SIEM-систем и выявляются проблемы, связанные с обработкой большого объема событий. В данном контексте Elasticsearch выступает в качестве инструмента для обработки и анализа больших данных. В работе говорится о необходимости горизонтального масштабирования в рамках крупных компаний с высокой нагрузкой на системы безопасности. Разработанный прототип включает инструменты для сбора данных с множества источников, их индексирования и хранения с возможностью последующего анализа событий безопасности. В работе выделяется важность открытых решений, таких как Elastic Stack, которые позволяют адаптировать систему под уникальные нужды организации и эффективно справляться с угрозами безопасности за счет гибкой настройки логирования и аналитики.

Балашов Н.А., М.В. Балашова, С.Р. Книгин, Н.А. Кутовский описали практический опыт внедрения стека ELK для сбора и анализа системных журналов в облачной инфраструктуре Объединенного института ядерных исследований [6]. Авторы отмечают, что рост инфраструктур приводит к значительному увеличению объемов обрабатываемых данных, что делает традиционные методы анализа логов недостаточно эффективными. Elasticsearch обеспечивает централизованное хранение и обработку событий с распределенной архитектурой, что позволяет динамически масштабировать систему в зависимости от потребностей. Особое внимание уделяется вопросам отказоустойчивости и безопасности, где ELK Stack демонстрирует высокую гибкость и адаптивность к специфике научных вычислительных инфраструктур. Кроме того, в статье обсуждается важность использования дополнительных инструментов, таких как Nginx и Open Distro для Elasticsearch, для улучшения функциональности и безопасности при работе с журналами событий.

Dzik C.S. и I.I. Piletski в своем исследовании [7] рассматривают подход к централизованному мониторингу и аналитике облачных приложений, работающих на базе Amazon Web Services (AWS). Авторы акцентируют внимание на проблемах, возникающих при масштабировании облачных сервисов, когда данные журналов и метрик поступают из множества разнородных источников. Для решения этой задачи предлагается интеграция Amazon CloudWatch и Kinesis с передачей данных в Elasticsearch для более глубокой аналитики. Использование Elasticsearch обусловлено его возможностями поиска и анализа в реальном времени, что критично для своевременного обнаружения проблем и аномалий,

которые могут сигнализировать о вредоносных действиях. Данный подход подчеркивает актуальность применения Elasticsearch для анализа событий в реальном времени в облачных инфраструктурах, особенно с точки зрения обеспечения безопасности и мониторинга.

Применение методов интеллектуального анализа и моделей работы с большими данными в системах информационной безопасности тесно связано с задачами автоматизации обнаружения и предотвращения атак. Например, исследования, посвященные архитектуре систем обнаружения атак, демонстрируют, как подходы на основе интеллектуального анализа данных могут быть интегрированы в корпоративные и государственные информационные системы для повышения их устойчивости к угрозам [8].

Кроме того, современные модели представления больших данных, включая использование NoSQL-баз данных, играют ключевую роль в обработке и анализе событий безопасности. Такие подходы позволяют организовать хранение и быстрый доступ к большим объемам данных о кибератаках, что критично для своевременного обнаружения угроз [9].

Особого внимания заслуживают работы, посвященные детектированию стегоинсайдеров в корпоративных сетях, которые объединяют гибридные модели NoSQL и методы обнаружения скрытых угроз. Подходы, описанные в исследованиях, подчеркивают важность использования гибридных баз данных для выявления внутренних угроз, таких как скрытые данные или внедренные инсайдеры [10, 11].

Авторы исследования на тему «Big Data Processing for Full-Text Search and Visualization with Elasticsearch» [12] рассматривают использование Elasticsearch для анализа больших данных с целью идентификации пользователей на основе косвенных признаков их активности. Они акцентируют внимание на том, что обработка данных такого масштаба требует высокопроизводительных систем для полнотекстового поиска и кластеризации. Elasticsearch был выбран благодаря своей способности работать с неструктурированными данными и поддерживать высокую скорость поиска, что позволяет эффективно решать задачи анализа активности пользователей. Важным аспектом работы является способность Elasticsearch интегрироваться с MapReduce для обработки больших объемов данных, что делает его подходящим инструментом для анализа событийных журналов. Это подчеркивает его значимость в контексте задачи анализа логов и обнаружения аномалий в системах безопасности.

Hårek Haugerud, Mohamad Sobhie и Anis Yazidi предложили метод оптимизации конфигурации Elasticsearch через стохастическое приближение с возмущением [13]. В работе исследуется проблема автоматической оптимизации параметров Elasticsearch для повышения его производительности в условиях больших нагрузок. Одной из ключевых проблем Elasticsearch является снижение эффективности при увеличении объема хранимых данных и нагрузки на систему. Авторы предлагают использовать алгоритм стохастического приближения с возмущением (SPSA) для автоматической настройки параметров Elasticsearch без необходимости перезагрузки системы. В результате такой оптимизации достигается улучшение времени отклика и увеличение производительности системы более чем на 40 %. Этот подход демонстрирует необходимость и важность правильной настройки Elasticsearch для работы с большими массивами данных, что критично для анализа журналов событий.

Несмотря на преимущества в области анализа журналов событий с использованием Elasticsearch и стека ELK, которые были приведены в исследованиях выше, многие методики имеют ряд недостатков, ограничивающих их эффективность в выявлении вредоносных действий. Многие исследования в этой области сосредоточены на решении конкретных задач, например, на анализе сетевого трафика или наблюдении за облачными системами, и не предлагают универсальных инструментов для выявления сложных атак или многоэтапных сценариев действий злоумышленников. Проблемой также остаются высокие затраты на внедрение и настройку существующих решений, их неполная совместимость с существующей инфраструктурой и ограниченная способность адаптироваться к новым угрозам. В связи с этим разработка методики анализа журналов событий на базе Elasticsearch, предназначенной для автоматического выявления признаков вредоносных

активностей, позволяет решить обозначенные проблемы. Предложенный подход, благодаря своей универсальности, способности адаптироваться к меняющимся условиям и экономической эффективности, является актуальным и востребованным инструментом для обеспечения безопасности информационных систем.

Основные компоненты и принципы работы стека ELK

Стек ELK состоит из трех основных компонентов:

1. Elasticsearch – это поисковая система, предназначенная для эффективного хранения и быстрого поиска данных. Она использует распределенную архитектуру, что делает её оптимальной для работы с большими массивами информации.

2. Logstash – инструмент для сбора, обработки и передачи данных в Elasticsearch. Он поддерживает различные источники и форматы данных, позволяя извлекать логи из разнообразных систем, таких как файлы, базы данных, приложения и устройства.

3. Kibana – интерфейс для визуализации и анализа данных, хранящихся в Elasticsearch. С помощью Kibana можно создавать дашборды, отслеживать ключевые показатели и исследовать тренды, что упрощает анализ логов.

На рис. 1 показано, как эти компоненты взаимодействуют в рамках информационной системы, где каждый из них выполняет свою роль, обеспечивая сбор, обработку, хранение и визуализацию данных.

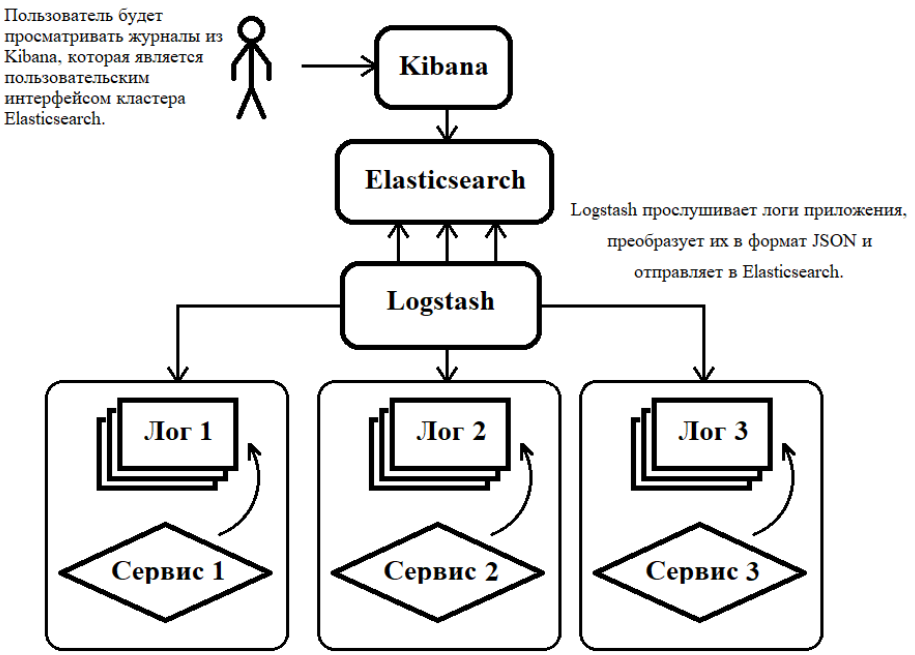

Рис. 1. **Взаимодействие компонентов стека ELK**

Logstash выполняет сбор данных из различных источников, включая серверные журналы, журналы приложений, сетевые события и информацию из системы управления базами данных. На этапе обработки система осуществляет предварительную фильтрацию записей, структурный анализ информации и трансформацию данных в стандартизированные форматы для последующей интеграции в систему хранения.

Обработанные наборы данных передаются в Elasticsearch, где происходит их индексация в распределённом хранилище с возможностью оперативного поиска и аналитической обработки. Платформа обеспечивает высокопроизводительную работу с крупными массивами информации, поддерживая расширенные функции текстового поиска и аналитические агрегации данных. Для оптимизации хронологических данных реализованы механизмы управления жизненным циклом индексов (ILM). Эти политики автоматизируют

ротацию хранилищ, архивацию устаревших данных и удаление неактуальной информации, обеспечивая рациональное использование дискового пространства и поддержку исторических данных в соответствии с заданными регламентами.

Kibana подключается к Elasticsearch для создания визуализаций и проведения анализа данных. Дашборды Kibana позволяют отслеживать ключевые метрики, такие как количество запросов, частота ошибок и распределение активности по IP-адресам. Также создаются графики и оповещения, чтобы оперативно реагировать на подозрительную активность.

Схема отражает потоки данных в системе на всех этапах обработки. Logstash служит начальным звеном, получая данные из источников и отправляя их в Elasticsearch, который отвечает за хранение и управление большими объемами данных. Kibana завершает цикл, визуализируя информацию, полученную из Elasticsearch, для создания аналитических отчетов и мониторинга системы. Все три компонента работают вместе для создания единой платформы управления и анализа данных.

Инструменты для журналирования Java-приложений

Java-приложения нередко выступают в качестве ключевых компонентов IT-инфраструктуры, предназначенной для государственных, муниципальных или корпоративных организаций. Поэтому логирование их работы и анализ этих логов имеют решающее значение для обеспечения стабильности и безопасности системы.

Для Java-приложений широко используются такие инструменты, как Log4j, Logback и SLF4J для ведения журнала событий. Эти библиотеки позволяют разработчикам настраивать уровни логирования (DEBUG, INFO, WARN, ERROR) и формат сообщений. Тем не менее для эффективного сбора, передачи и анализа журналов событий в реальном времени необходимо более функциональное решение, способное обрабатывать большие объемы данных. В данной области оптимальным выбором является использование стека ELK.

Logstash и Filebeat часто используются для сбора логов Java-приложений. Logstash позволяет захватывать и фильтровать логи, передавая их в Elasticsearch для дальнейшего анализа [14]. Этот процесс может включать парсинг логов, нормализацию данных и добавление метаданных (например, временные метки или данные о сервере), что облегчает их дальнейший анализ.

Однако для более легковесных решений, таких как высоконагруженные системы, можно использовать Filebeat. Filebeat – это агент для передачи логов, который потребляет меньше ресурсов, чем Logstash, что делает его подходящим для масштабных инфраструктур [15].

Fluent Bit как альтернатива для облегченного сбора логов

Еще одной альтернативой Logstash является Fluent Bit – малотребовательный к системным ресурсам и высокопроизводительный агент для сбора логов, который также может интегрироваться с Elasticsearch. Этот инструмент отличается низким потреблением ресурсов и высокой скоростью работы, что делает его отличным выбором для крупных распределенных систем, где объем логов может достигать миллионов записей в сутки [16].

Использование Fluent Bit особенно полезно в сценариях, где критически важно минимизировать задержки при передаче данных в Elasticsearch и снизить нагрузку на систему логирования.

Логирование Spring Boot приложений

Для приложений на базе Spring Boot, которые широко используются в микросервисных архитектурах, логирование и мониторинг становятся особенно важными. Приложения могут генерировать огромное количество логов, что делает их анализ в реальном времени необходимым для быстрого выявления проблем и аномалий.

Spring Boot предлагает встроенные механизмы для ведения логов, такие как Logback и SLF4J, которые могут быть настроены для передачи логов в стек ELK. Однако стандартные настройки логирования не всегда достаточны для крупных распределенных систем, поэтому требуется централизованное решение для сбора и анализа логов.

Для этого можно использовать Filebeat или Logstash для агрегации логов из всех сервисов Spring Boot [17]. Эти инструменты захватывают логи и передают их в Elasticsearch, где они индексируются и становятся доступными для поиска и анализа.

Преимущества централизованного логирования

Централизованное логирование позволяет не только улучшить процесс поиска и анализа логов, но и создавать централизованные дашборды, где можно отслеживать общие метрики производительности и выявлять аномалии в поведении приложений. В частности, такие метрики, как частота ошибок, время выполнения запросов, объем трафика и количество активных пользователей, могут быть легко визуализированы в Kibana.

Используя возможности Elasticsearch и Kibana, команды разработчиков могут быстро выявлять проблемы в производительности приложений, а также идентифицировать аномальные действия, которые могут свидетельствовать о вредоносной активности.

Анализ журналов безопасности с использованием стека ELK

Журналы безопасности содержат ценную информацию о событиях, которые происходят в системе, и могут служить источником данных для обнаружения угроз безопасности. Правильная организация анализа журналов безопасности позволяет оперативно выявлять подозрительную активность и реагировать на потенциальные инциденты.

Журналы безопасности могут включать в себя данные о попытках входа в систему, неудачных аутентификациях, изменениях конфигурации, обращениях к чувствительным данным и других критически важных событиях. Эти данные, собранные с различных систем (сетевые устройства, операционные системы, базы данных, приложения), могут быть централизованы в Elasticsearch для дальнейшего анализа.

Ключевыми этапами анализа журнала безопасности являются:

1. Сбор логов: логи собираются с различных источников с помощью Logstash или Filebeat. Эти инструменты позволяют настроить парсинг сложных форматов логов, таких как журналы аудита или данные из систем обнаружения вторжений.

2. Обработка данных: важным шагом является обработка данных – фильтрация, нормализация, обогащение метаданными и преобразование форматов логов. Это позволяет стандартизировать данные и упростить их анализ.

3. Индексация в Elasticsearch: после обработки данные передаются в Elasticsearch, где они индексируются и становятся доступными для поиска и анализа.

4. Анализ и визуализация: используя Kibana, можно настроить дашборды для мониторинга ключевых метрик безопасности, таких как количество неудачных попыток входа в систему, подозрительные IP-адреса, активность пользователей и другие индикаторы компрометации.

Обнаружение аномалий и автоматизация реагирования в Elasticsearch

Одной из ключевых задач анализа логов является своевременное выявление аномалий, которые могут свидетельствовать о вредоносных действиях. Elasticsearch включает в себя встроенные возможности машинного обучения (Machine Learning – ML), которые помогают выявлять аномалии в данных. С помощью ML можно создавать модели, которые автоматически обучаются на данных и распознают отклонения от нормального

поведения. Например, модель может отслеживать частоту определённых событий, таких как попытки входа в систему, и выявлять пики активности, которые могут указывать на атаки методом перебора или внутренние злоупотребления.

Ключевые функции Elasticsearch включают:

– анализ временных рядов: ML в Elasticsearch анализирует временные ряды событий и находит отклонения, которые могут сигнализировать о потенциальной угрозе;

– корреляционный анализ: возможность выявлять связи между различными событиями помогает обнаружить скрытые паттерны, которые могут указывать на сложные атаки;

– автоматическое обнаружение аномалий: модели машинного обучения в Elasticsearch автоматически находят аномалии в данных без необходимости вручную настраивать пороговые значения;

– прогнозирование поведения: ML прогнозирует дальнейшие события на основе предыдущих данных, что помогает подготовиться к возможным угрозам;

– кластеризация событий: Elasticsearch способен группировать схожие события, что упрощает анализ больших объёмов данных и позволяет выявить основные причины инцидентов;

– анализ текстовых данных: Elasticsearch анализирует неструктурированные данные, такие как текстовые логи, для извлечения полезных паттернов и ключевых событий, связанных с безопасностью.

Кроме обнаружения аномалий важным аспектом безопасности является автоматизация реагирования на инциденты. Современные системы, построенные на базе стека ELK, могут автоматизировать ответ на угрозы. Это достигается через настройку автоматических уведомлений, которые срабатывают при выявлении подозрительной активности, например, неудачных попыток входа в систему или аномальных сетевых запросов. Такие системы могут не только уведомлять администраторов о возможных угрозах, но и предпринимать шаги для нейтрализации инцидентов, например, блокировать подозрительные IP-адреса, завершать сессии с аномальным поведением или запускать восстановление системы. Автоматизация этого процесса ускоряет отклик на угрозы и минимизирует риски компрометации системы.

В Kibana можно настроить систему уведомлений, которая будет срабатывать при обнаружении подозрительного поведения. Например, оповещение может быть настроено на случай нескольких неудачных попыток входа в систему с одного IP-адреса, что может указывать на попытку несанкционированного доступа.

Интеграция с SIEM-системами (системы управления информацией и событиями безопасности) позволяет автоматически принимать защитные меры, такие как блокировка IP-адресов или завершение подозрительных сессий.

Методика

Методика включает настройку стека ELK, установку всех его компонентов и их конфигурирование для хранения и обработки логов. На первом этапе выполняется установка Elasticsearch, Logstash и Kibana с проверкой совместимости версий. Далее производится настройка Elasticsearch для распределенного хранения и индексации, что обеспечивает обработку больших объемов данных. Также создаются индексы для различных типов логов (системные, сетевые, логи безопасности и приложений) и настраиваются параметры долгосрочного хранения, включая политику управления жизненным циклом индексов (ILM) для оптимизации использования хранилища.

После настройки Elasticsearch подключается Kibana, которая используется для анализа данных и визуализации, включая создание дашбордов для мониторинга показателей, таких как количество запросов к системе, распределение по IP-адресам и частота ошибок.

На следующем этапе выполняется настройка сбора данных с использованием Logstash и Filebeat. Filebeat устанавливается на серверах, с которых собираются логи, и передает данные в Logstash или напрямую в Elasticsearch. Filebeat также настраивается для добавления метаданных (например, имени сервера, пути к файлу лога), что облегчает поиск и фильтрацию. В Logstash создаются фильтры для структурирования логов и фильтрации данных по уровню критичности (DEBUG, INFO, WARN, ERROR). Это позволяет выделить важную информацию и установить фильтры для аномальных событий, например, высокого процента ошибок или попыток входа в систему.

Для расширения данных добавляются геолокация и информация об источнике событий, а также присваиваются метки событиям, связанным с подозрительной активностью, например, «неудачная авторизация» или «атака полного перебора».

Настройка индексов и правил выявления аномалий в Elasticsearch включает создание отдельных индексов для каждого типа данных, что упрощает их анализ. Устанавливаются шаблоны индексов для упорядочивания структуры данных, задаются пороговые значения и лимиты для выявления подозрительных действий, таких как частые попытки входа или активность в нерабочее время. Также производится настройка фильтрации событий по географическим IP-адресам для отслеживания подозрительных подключений и анализа логов приложений на наличие уязвимостей.

Для визуализации и анализа данных в Kibana создаются дашборды, которые отображают метрики, такие как активность пользователей, неудачные попытки входа и распределение по IP-адресам. Настраиваются предупреждения (оповещения) для своевременного реагирования на угрозы, например, при превышении лимита подключений за минуту или множественных неудачных попытках входа с одного IP. Также реализуется автоматическое реагирование на события, такие как временная блокировка IP-адресов при выявлении атак полного перебора.

Для анализа временных рядов создаются корреляции между событиями, позволяющие обнаруживать сложные паттерны активности, например, одновременные подключения с разных IP-адресов или доступ к чувствительным данным в нестандартное время. В целях автоматизации выявления угроз и инцидентов используется машинное обучение для обнаружения аномалий. Встроенные модели Elastic анализируют частоту неудачных попыток входа и всплески активности. Настраивается автоматический ответ на инциденты, включая временную блокировку IP при подозрительной активности и завершение активных сессий при аномальных всплесках подключений.

На заключительном этапе осуществляется интеграция стека ELK с SIEM-системами для расширения возможностей мониторинга и автоматического реагирования. Настраивается передача данных о событиях в SIEM, что позволяет централизовать информацию о безопасности и оперативно реагировать на инциденты.

На рис. 2 изображен порядок выполнения разработанной методики.

Шаг 1. Установка Elasticsearch, Logstash и Kibana

Перед установкой проверяется совместимость версий Elasticsearch, Logstash и Kibana для обеспечения стабильной работы.

Выполняется установка Elasticsearch, после чего настраиваются параметры конфигурации, такие как cluster.name, node.name, path.data, path.logs. Запуск Elasticsearch позволяет убедиться в его доступности, обычно на порту 9200.

Устанавливается Kibana, и в конфигурационном файле указывается адрес Elasticsearch (обычно elasticsearch.hosts: ["http://localhost:9200"]). После запуска Kibana проверяется её доступность на порту 5601.

Устанавливается Logstash, и создается файл конфигурации для его подключения к Elasticsearch. Запуск Logstash проверяет его корректное подключение к Elasticsearch.

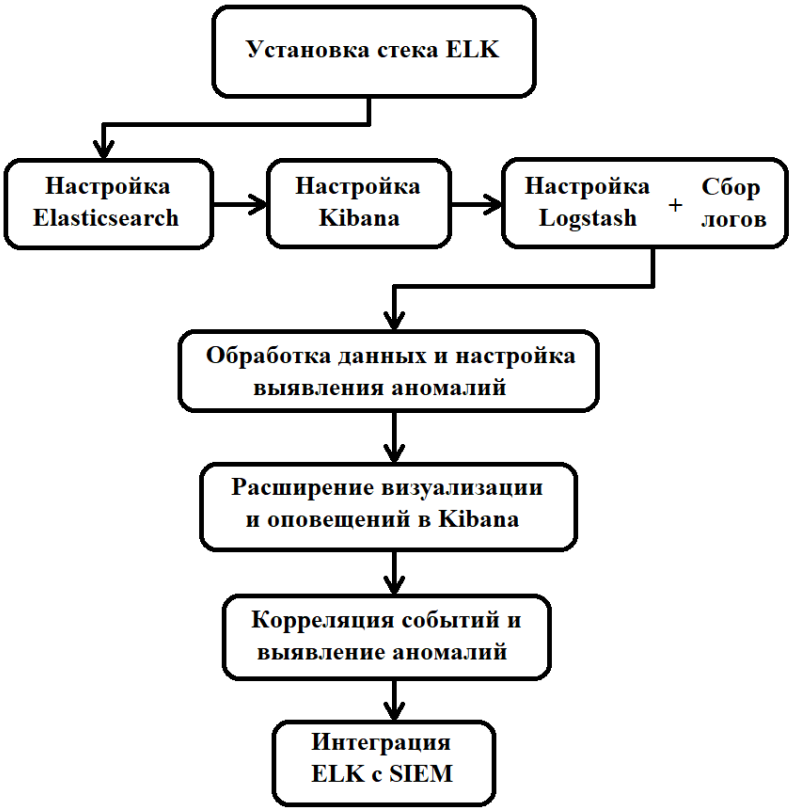

Рис. 2. **Методика**

Шаги для настройки и интеграции стека ELK

Шаг 2. Настройка Elasticsearch для распределенного хранения и индексации

На этапе настройки распределенного хранения в Elasticsearch задаются параметры, такие как number_of_shards и number_of_replicas, для обработки больших объемов данных. Дополнительно оптимизируются параметры производительности (refresh_interval, index.translog.durability).

Создаются отдельные индексы для хранения логов различных типов (системные, сетевые, логи безопасности и приложений).

Определяется политика управления жизненным циклом индексов (ILM), указывающая периоды хранения данных и условия перехода между фазами, что позволяет автоматизировать архивирование логов и оптимизировать использование ресурсов.

Шаг 3. Настройка Kibana для визуализации данных

Kibana подключается к Elasticsearch, что позволяет получить доступ к индексам для последующей визуализации данных.

Создаются визуализации и дашборды для анализа логов: графики запросов, распределение по IP-адресам, частота ошибок и другие метрики.

Настраиваются оповещения, позволяющие оперативно реагировать на потенциальные угрозы, такие как превышение лимита подключений или множественные неудачные попытки входа.

Шаг 4. Сбор логов с помощью Logstash и Filebeat

Filebeat устанавливается на серверах, с которых требуется собирать логи, и настраивается для передачи логов в Logstash или напрямую в Elasticsearch.

В конфигурации Filebeat добавляются метаданные (например, имя сервера и путь к файлу лога), что упрощает фильтрацию и поиск данных.

В Logstash создаются фильтры, структурирующие данные по уровню критичности (DEBUG, INFO, WARN, ERROR), что позволяет выделить важные события и обнаруживать аномалии, такие как высокий процент ошибок или частые попытки неудачной авторизации.

Шаг 5. Обработка данных и настройка выявления аномалий

В систему добавляются модули геолокации и меток событий, что позволяет обозначить потенциально подозрительную активность (например, неудачные попытки авторизации).

Настраиваются шаблоны индексов для структурирования данных и устанавливаются пороговые значения для выявления подозрительных действий.

Выполняется фильтрация событий по географическим IP-адресам, что позволяет отслеживать подозрительные подключения и анализировать логи на наличие уязвимостей.

Шаг 6. Расширение визуализации и оповещений в Kibana

Разрабатываются дополнительные дашборды для отображения ключевых метрик, таких как активность пользователей, неудачные попытки входа и распределение по IP-адресам.

Настраиваются автоматические оповещения и реакции на угрозы, включая временную блокировку IP-адресов при обнаружении подозрительной активности.

Шаг 7. Корреляция событий и выявление аномалий

Корреляция событий и выявление аномалий является ключевым этапом методики, направленным на своевременное обнаружение вредоносных действий в информационной системе. Этот шаг основывается на автоматизированной обработке логов, поступающих из различных источников, и анализе взаимосвязей между событиями. Суть предлагаемого подхода заключается в идентификации аномального поведения, которое может указывать на целевые атаки (APT), попытки эксплуатации уязвимостей, несанкционированный доступ или другие подозрительные действия.

Выделяются следующие аномалии:

1. Аномалии доступа:

– множественные неудачные попытки входа (атака полного перебора);

– авторизация с географически удаленных или редко используемых IP-адресов, особенно в сочетании с успешным входом в короткий промежуток времени;

– подключение с IP-адресов, попадающих в известные списки подозрительной активности.

2. Аномалии поведения пользователей:

– попытки доступа к нехарактерным для пользователя системам или данным;

– резкое увеличение объема операций за короткий промежуток времени (например, массовая выгрузка данных);

– изменение ролей или прав пользователя без явного запроса.

3. Сетевые аномалии:

– подозрительные подключения к редко используемым портам или услугам;

– всплески активности, такие как многократные запросы с одного IP-адреса (DDoS-подобная активность);

– необычное количество исходящих соединений, которые могут указывать на утечку данных.

4. Аномалии в системных процессах:

– неожиданные изменения конфигурации системы или приложений;

– частые ошибки или перезапуски служб, которые могут быть следствием атаки.

Обнаружение этих аномалий позволяет оперативно выявлять угрозы, которые часто остаются незамеченными при традиционном анализе логов. Например, множественные неудачные попытки входа могут быть признаком попытки подбора пароля, а подключение из необычного региона – индикатором компрометации учетной записи. Анализ таких событий в связке позволяет быстрее идентифицировать сложные атаки, такие как атаки на привилегии или многошаговые атаки, которые используют комбинацию уязвимостей и обходят защиту.

На базе Elasticsearch настраиваются автоматизированные правила для поиска подозрительных шаблонов в логах. Например, если в логах входа фиксируются пять неудачных попыток авторизации с разных IP-адресов в течение минуты, система генерирует оповещение. Используются алгоритмы машинного обучения, встроенные в Elastic Stack (например, модуль машинного обучения X-Pack), для определения нетипичного поведения на основе анализа исторических данных.

Logstash собирает данные из журналов приложений, сетевых устройств, систем безопасности (IDS/IPS) и серверов. Для каждого события добавляются метаданные, такие как геолокация IP-адресов или хост, с которого оно поступило. Это помогает связывать действия, происходящие в разных подсистемах.

В логах анализируются взаимосвязи. Например, подозрительное поведение можно выявить, если успешный вход из одной страны сопровождается множественными попытками доступа к конфиденциальным данным с другого IP-адреса. Создаются цепочки событий, например: авторизация, запуск команды с повышением прав, копирование большого объема данных. Такие последовательности сигнализируют о возможных атаках.

Настраиваются пороговые значения для критических метрик (например, число запросов в секунду или число ошибок). При превышении порогов Elasticsearch генерирует оповещение через Kibana, а встроенные плагины, такие как ElastAlert, могут автоматически выполнять действия: временно блокировать IP-адрес, уведомлять администратора, завершать подозрительные сессии.

В Kibana создаются дашборды для анализа ключевых метрик, таких как частота попыток авторизации, количество ошибок и активности IP-адресов. Графики и диаграммы помогают быстро выявить всплески активности или отклонения от нормы.

Шаг 8. Интеграция ELK с SIEM для централизованного мониторинга

Настраивается передача данных из стека ELK в SIEM-систему для расширенного мониторинга и автоматического реагирования на инциденты.

Определяются политики автоматического реагирования, которые позволяют централизованно управлять безопасностью, выявлять и анализировать потенциальные угрозы в режиме реального времени.

Заключение

В данной работе представлена новая методика анализа журналов событий информационных систем, основанная на использовании стека ELK для выявления сигналов, указывающих на вредоносные действия. Основное преимущество предложенного подхода заключается в интеграции современных методов корреляции событий и автоматизированного выявления аномалий, что позволяет значительно повысить эффективность мониторинга информационной безопасности. Благодаря гибкости настройки, автоматизации процессов и использованию алгоритмов машинного обучения предложенная методика позволяет сократить время реагирования на действия злоумышленников. Это достигается за счет быстрой идентификации подозрительных действий, таких как неудачные попытки авторизации, всплески сетевой активности или аномалии в поведении пользователей, что минимизирует риски компрометации систем.

Методика предоставляет универсальный инструмент для анализа событий в реальном времени, который легко интегрируется в существующую инфраструктуру и адаптируется к меняющимся условиям. Ее применение способствует не только своевременному обнаружению угроз, но и автоматизации реагирования на инциденты, включая блокировку IP-адресов, завершение подозрительных сессий и отправку уведомлений.

В рамках дальнейшего развития исследования планируется реализация предложенной методики на практике с целью проведения экспериментов в реальной среде. Особое внимание будет уделено сравнению эффективности предлагаемого подхода с существующими решениями. Также планируется расширение возможностей автоматизации, в том числе более глубокая интеграция с SIEM-системами и использование продвинутых алгоритмов машинного обучения для прогнозирования потенциальных угроз и защиты информационных систем.

Список источников

1. A Technique for Detecting the Substitution of a Java-Module of an Information System Prone to Pharming with Using a Hidden Embedding of a Digital Watermark Resistant to Decompilation / Sh. Pavel [et al.] // International Congress on Ultra Modern Telecommunications and Control Systems and Workshops: Virtual, Online, 2021. P. 219–223. DOI: 10.1109/ICUMT54235.2021.9631736. EDN YVVEUX.

2. Исследование и алгоритм предотвращения эксплуатации уязвимостей библиотеки журналирования Log4j в информационных системах Java-приложений / П.И. Шариков [и др.] // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Сер. 1: Естественные и технические науки. 2023. № 4. С. 100–106. DOI: 10.46418/2079-8199_2023_4_19. EDN BULSOH.

3. Шариков П.И. Исследование атаки обфускацией на байт-код java-приложения с целью разрушения или повреждения цифрового водяного знака // I-methods. 2022. Т. 14. № 1. EDN GQGKIV.

4. Применение стека ELK для анализа сетевого трафика / М.С. Протодьяконова [и др.] // World science: problems and innovations. 2018. Т. 1. С. 105–106.

5. Котенко И.В., Кулешов А.А., Ушаков И.А. Система сбора, хранения и обработки информации и событий безопасности на основе средств Elastic Stack // Труды СПИИРАН. 2017. № 5 (54). С. 5–34.

6. Применение стека технологий ELK для сбора и анализа системных журналов событий / Н.А. Балашов [и др.] // Современные информационные технологии и IT-образование. 2021. Т. 17. № 1. С. 61–68.

7. Dzik C.S., Piletski I.I. Real-Time AWS resources monitoring and analytics // Big data and advanced analytics. 2021. № 7-1. С. 25–30.

8. Майоров А.В. Архитектура и программная реализация системы обнаружения компьютерных атак в корпоративных и государственных информационных системах на основе методов интеллектуального анализа // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Сер. 1: Естественные и технические науки. 2023. № 2. С. 40–46. DOI: 10.46418/2079-8199_2023_2_8. EDN HEPDFF.

9. Майоров А.В., Красов А.В., Ушаков И.А. Модель представления больших данных о компьютерных атаках в формате nosql // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Сер. 1: Естественные и технические науки. 2023. № 2. С. 47–54. DOI: 10.46418/2079-8199_2023_2_9. EDN GDZKWM.

10. An approach for stego-insider detection based on a hybrid nosql database / I. Kotenko [et al.] // Journal of Sensor and Actuator Networks. 2021. Vol. 10. № 2. DOI: 10.3390/jsan10020025. EDN IKOMVS.

11. Detection of stego-insiders in corporate networks based on a hybrid NoSQL database model / I. Kotenko [et al.] // ACM International Conference Proceeding Ser.: 4. SPb., 2020. P. 3442612. DOI: 10.1145/3440749.3442612. EDN EYKYHJ.

12. Big Data Processing for Full-Text Search and Visualization with Elasticsearch / A. Voit [et al.] // (IJACSA) International Journal of Advanced Computer Science and Applications. 2017. Т. 8. № 12. P. 76–83.

13. Hårek Haugerud, Mohamad Sobhie, Anis Yazidi Tuning of Elasticsearch Configuration: Parameter Optimization Through Simultaneous Perturbation Stochastic Approximation // Frontiers in big data. 2022. Т. 8.

14. Logging Java Apps with ELK. URL: https://logz.io/blog/logging-java-elk-stack (дата обращения: 20.11.2024).

15. Store Java application's logs in Elasticsearch. URL: https://mostafa-asg.github.io/post/ship-app-logs-to-elasticsearch-elk-filebeat (дата обращения: 20.11.2024).

16. Java logging with Fluent Bit and Elasticsearch. URL: https://chronosphere.io/learn/java-logging-with-fluent-bit-and-elasticsearch (дата обращения: 20.11.2024).

17. Spring Boot Logs Aggregation and Monitoring Using ELK Stack. URL: https://auth0.com/blog/spring-boot-logs-aggregation-and-monitoring-using-elk-stack (дата обращения: 20.11.2024).

References

1. A Technique for Detecting the Substitution of a Java-Module of an Information System Prone to Pharming with Using a Hidden Embedding of a Digital Watermark Resistant to Decompilation / Sh. Pavel [et al.] // International Congress on Ultra Modern Telecommunications and Control Systems and Workshops: Virtual, Online, 2021. P. 219–223. DOI: 10.1109/ICUMT54235.2021.9631736. EDN YVVEUX.

2. Issledovanie i algoritm predotvrashcheniya ekspluatacii uyazvimostej biblioteki zhurnalirovaniya Log4j v informacionnyh sistemah Java-prilozhenij / P.I. Sharikov [i dr.] // Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizajna. Ser. 1: Estestvennye i tekhnicheskie nauki. 2023. № 4. S. 100–106. DOI: 10.46418/2079-8199_2023_4_19. EDN BULSOH.

3. Sharikov P.I. Issledovanie ataki obfuskaciej na bajt-kod java-prilozheniya s cel'yu razrusheniya ili povrezhdeniya cifrovogo vodyanogo znaka // I-methods. 2022. T. 14. № 1. EDN GQGKIV.

4. Primenenie steka ELK dlya analiza setevogo trafika / M.S. Protod'yakonova [i dr.] // World science: problems and innovations. 2018. T. 1. S. 105–106.

5. Kotenko I.V., Kuleshov A.A., Ushakov I.A. Sistema sbora, hraneniya i obrabotki informacii i sobytij bezopasnosti na osnove sredstv Elastic Stack // Trudy SPIIRAN. 2017. № 5 (54). S. 5–34.

6. Primenenie steka tekhnologij ELK dlya sbora i analiza sistemnyh zhurnalov sobytij / N.A. Balashov [i dr.] // Sovremennye informacionnye tekhnologii i IT-obrazovanie. 2021. T. 17. № 1. S. 61–68.

7. Dzik C.S., Piletski I.I. Real-Time AWS resources monitoring and analytics // Big data and advanced analytics. 2021. № 7-1. S. 25–30.

8. Majorov A.V. Arhitektura i programmnaya realizaciya sistemy obnaruzheniya komp'yuternyh atak v korporativnyh i gosudarstvennyh informacionnyh sistemah na osnove metodov intellektual'nogo analiza // Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizajna. Ser. 1: Estestvennye i tekhnicheskie nauki. 2023. № 2. S. 40–46. DOI: 10.46418/2079-8199_2023_2_8. EDN HEPDFF.

9. Majorov A.V., Krasov A.V., Ushakov I.A. Model' predstavleniya bol'shih dannyh o komp'yuternyh atakah v formate nosql // Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizajna. Ser. 1: Estestvennye i tekhnicheskie nauki. 2023. № 2. S. 47–54. DOI: 10.46418/2079-8199_2023_2_9. EDN GDZKWM.

10. An approach for stego-insider detection based on a hybrid nosql database / I. Kotenko [et al.] // Journal of Sensor and Actuator Networks. 2021. Vol. 10. № 2. DOI: 10.3390/jsan10020025. EDN IKOMVS.

11. Detection of stego-insiders in corporate networks based on a hybrid NoSQL database model / I. Kotenko [et al.] // ACM International Conference Proceeding Ser.: 4. SPb., 2020. P. 3442612. DOI: 10.1145/3440749.3442612. EDN EYKYHJ.

12. Big Data Processing for Full-Text Search and Visualization with Elasticsearch / A. Voit [et al.] // (IJACSA) International Journal of Advanced Computer Science and Applications. 2017. T. 8. № 12. P. 76–83.

13. Hårek Haugerud, Mohamad Sobhie, Anis Yazidi Tuning of Elasticsearch Configuration: Parameter Optimization Through Simultaneous Perturbation Stochastic Approximation // Frontiers in big data. 2022. T. 8.

14. Logging Java Apps with ELK. URL: https://logz.io/blog/logging-java-elk-stack (data obrashcheniya: 20.11.2024).

15. Store Java application's logs in Elasticsearch. URL: https://mostafa-asg.github.io/post/ship-app-logs-to-elasticsearch-elk-filebeat (data obrashcheniya: 20.11.2024).

16. Java logging with Fluent Bit and Elasticsearch. URL: https://chronosphere.io/learn/java-logging-with-fluent-bit-and-elasticsearch (data obrashcheniya: 20.11.2024).

17. Spring Boot Logs Aggregation and Monitoring Using ELK Stack. URL: https://auth0.com/blog/spring-boot-logs-aggregation-and-monitoring-using-elk-stack (data obrashcheniya: 20.11.2024).

Информация о статье:
Статья поступила в редакцию: 10.02.2025; одобрена после рецензирования: 15.03.2025;
принята к публикации: 17.03.2025
Information about the article:
The article was submitted to the editorial office: 10.02.2025; approved after review: 15.03.2025;
accepted for publication: 17.03.2025

Информация об авторах:
Дудников Иван Алексеевич, студент магистратуры кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций имени профессора М.А. Бонч-Бруевича (193232, Санкт-Петербург, пр. Большевиков, д. 22, к. 1), e-mail: van.dy@mail.ru
Шариков Павел Иванович, старший преподаватель кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций имени профессора М.А. Бонч-Бруевича (193232, Санкт-Петербург, пр. Большевиков, д. 22, к. 1), кандидат технических наук, e-mail: sharikov.pavel@ro.ru, https://orcid.org/0000-0003-3996-9217, SPIN-код: 5844-6493
Майоров Александр Владимирович, аспирант кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций имени профессора М.А. Бонч-Бруевича (193232, Санкт-Петербург, пр. Большевиков, д. 22, к. 1)

Information about authors:
Dudnikov Ivan A., master's degree student of the department of security communication systems of Saint-Petersburg State university of telecommunications named after professor M.A. Bonch-Bruevich (193232, Saint-Petersburg, Bolshevikov ave, 22, k. 1), e-mail: van.dy@mail.ru
Sharikov Pavel I., senior lecturer of the department of security communication systems of Saint-Petersburg State university of telecommunications named after professor M.A. Bonch-Bruevich (193232, Saint-Petersburg, Bolshevikov ave, 22, k. 1), candidate of technical sciences, e-mail: sharikov.pavel@ro.ru, https://orcid.org/0000-0003-3996-9217, SPIN: 5844-6493
Maiorov Alexander V., graduate student of the department of security communication systems of Saint-Petersburg State university of telecommunications named after professor M.A. Bonch-Bruevich (193232, Saint-Petersburg, Bolshevikov ave, 22, k. 1)