Методы анализа логов Sysmon для обнаружения киберугроз

Е. В. Костиков

***Аннотация*—В современном мире, где киберугрозы становятся все более изощренными, анализ системных логов играет ключевую роль в обеспечении безопасности информационных систем. Логи предоставляют ценную информацию о событиях, происходящих в сети, позволяя обнаруживать аномальные действия, которые могут указывать на атаки или нарушения безопасности. Регулярный мониторинг и анализ этих данных помогают в быстром выявлении инцидентов, что способствует оперативной реакции и минимизации ущерба. Кроме того, системные логи являются важным инструментом для расследования инцидентов, позволяя установить причины и масштабы атак. В условиях постоянного роста числа киберугроз, грамотный анализ логов становится жизненно необходимым для защиты организаций и их данных.**

Sysmon является мощным решением для анализа логов, которое может значительно улучшить безопасность информационной инфраструктуры. С помощью Sysmon можно отслеживать изменения в файловой системе, сетевые подключения и запущенные процессы, что позволяет выявлять подозрительные действия и аномалии. Этот инструмент интегрируется с SIEM-системами, что упрощает анализ и корреляцию данных.

Настоящая работа посвящена обзору существующих методов и решений для анализа логов Sysmon для обнаружения вредоносного программного обеспечения.

***Ключевые слова*—кибербезопасность, sysmon, анализ логов, вредоносное программное обеспечение, выявление вредоносной активности.**

I. Введение

В условиях стремительного развития информационных технологий и увеличения объемов цифровых данных, киберугрозы становятся все более изощренными и сложными. Вредоносное программное обеспечение представляет одну из наиболее серьезных угроз для информационных систем, от кражи конфиденциальной информации до нарушения нормального функционирования корпоративных сетей. Кибератаки продолжают набирать обороты, представляя серьезную угрозу для организаций по всему миру. По данным исследования, глобальная экономика в 2020 году потеряла около 1 триллиона долларов из-за киберпреступлений, что примерно на 50 процентов больше по сравнению с 2018 годом, когда эти потери оценивались в 600 миллиардов долларов. В этих значениях учитываются как прямые финансовые убытки от киберинцидентов, так и затраты на меры по обеспечению безопасности [1]. В ответ на это возросло использование продвинутых методов анализа для предотвращения и расследования инцидентов безопасности. Для эффективной защиты от угроз критически важно не только внедрение передовых средств защиты, но и использование методов анализа журналов событий (логов), которые позволяют своевременно выявлять и реагировать на подозрительные активности.

Sysmon предназначен для детального мониторинга и записи событий на уровне операционной системы. Он фиксирует широкий спектр данных, включая создание процессов, изменения в файловой системе, сетевые соединения и другие важные события, которые могут указывать на наличие вредоносной активности. Анализ логов, генерируемых Sysmon, предоставляет уникальные возможности для глубокой диагностики и расследования инцидентов безопасности. Логи содержат исчерпывающую информацию о поведении системы, что позволяет обнаруживать аномалии, которые могут указывать на компрометацию системы. Регулярный анализ логов помогает выявлять скрытые угрозы, которые могут оставаться незамеченными традиционными антивирусными программами и системами обнаружения вторжений. Также на основе логов Sysmon можно разрабатывать поведенческие модели, способные предсказывать и предотвращать будущие атаки. Данная утилита хорошо зарекомендовала себя как среди исследователей кибербезопасности, так и в среде аналитиков SOC-центров.

В разделе II данной работы приводится описание службы Sysmon и основных особенностей использования данного средства для сбора и анализа системных логов. Раздел III посвящен рассмотрению существующих подходов и методов к проведению анализа логов Sysmon для обнаружения вредоносной активности, активности вредоносного программного обеспечения в частности.

II. Описание Sysmon

Sysmon (System Monitor) – это системная служба Windows, используемая для мониторинга поведения системы [2]. Sysmon представляет собой инструмент для отслеживания происходящей в системе активности, сбора подробных структурированных данных о

Статья получена 28 октября 2024.

Егор Вячеславович Костиков, Московский государственный университет имени М.В. Ломоносова, магистратура Кибербезопасность (МГУ-Сбер) (email: kostikovev@mail.ru).

произошедших событиях, которые могут быть использованы для дальнейшего анализа на наличие аномальной активности.

Sysmon отслеживает системную активность и обеспечивает ее журналирование, не предоставляя при этом средств для анализа или визуализации собираемых данных.

Таблица 1: События, регистрируемые Sysmon версии 15.11

ID	Описание регистрируемого события
1	Создание процесса
2	Изменение в дате создания файлов
3	Сетевая активность процессов
4	Запуск и остановка службы Sysmon
5	Завершение процесса
6	Загрузка драйверов в ОС
7	Загрузка модулей (DLL-библиотек) в память процессов
8	Создание потока одного процесса в адресном пространстве другого процесса
9	Прямой доступ к дискам и томам
10	Доступ одного процесса к памяти другого процесса
11	Создание файлов
12	Создание и удаление ключей и значений реестра
13	Изменение значений в реестре
14	Внесение изменений в названия ключей и значений в реестре
15	Создание альтернативного потока данных
16	Изменение конфигурации Sysmon
17	Создание именованного канала
18	Соединение по именованному каналу между клиентом и сервером
19	Создание фильтра событий WMI
20	Создание потребителя событий WMI
21	Подключение потребителя событий к фильтру событий WMI
22	DNS-запрос от какого-либо процесса
23	Удаление файла с сохранением его копии с каталоге-архиве Sysmon
24	Изменение содержимого буфера обмена
25	Попытки сокрытия процессов
26	Удаление файла
27	Обнаружение и блокирование создания исполняемых файлов
28	Обнаружение и блокирование удаления файлов
29	Обнаружение создания исполняемых файлов
255	Ошибка службы Sysmon

Sysmon может собирать данные о создании новых процессов и файлов, обнаружении сетевых подключений, изменении содержимого реестра и многих прочих событий. Полный перечень регистрируемых службой Sysmon версии 15.11 событий приведен в Таблице 1.

Стоит отметить, что регистрируемые службой Sysmon события, как правило, более информативны, чем аналогичные события, регистрируемые стандартными средствами Windows. Например, для события создания процесса, имеющего идентификатор 1 для Sysmon (Рис. 1) и идентификатор 4688 журнала безопасности Windows (Рис. 2), средствами Sysmon регистрируется более полная информация о создаваемом процессе и о родительском процессе.

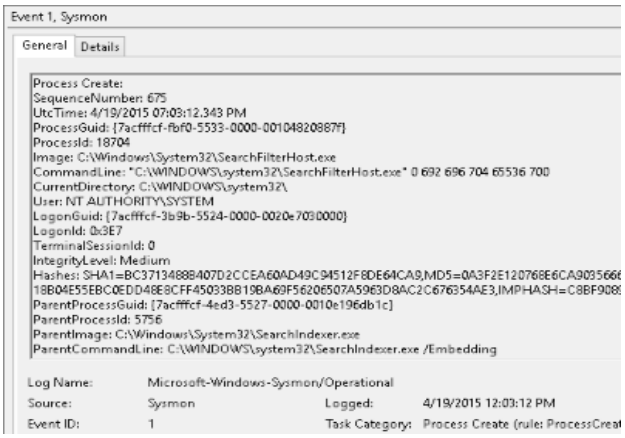

Рис. 1: Пример события с идентификатором 1, регистрируемого Sysmon

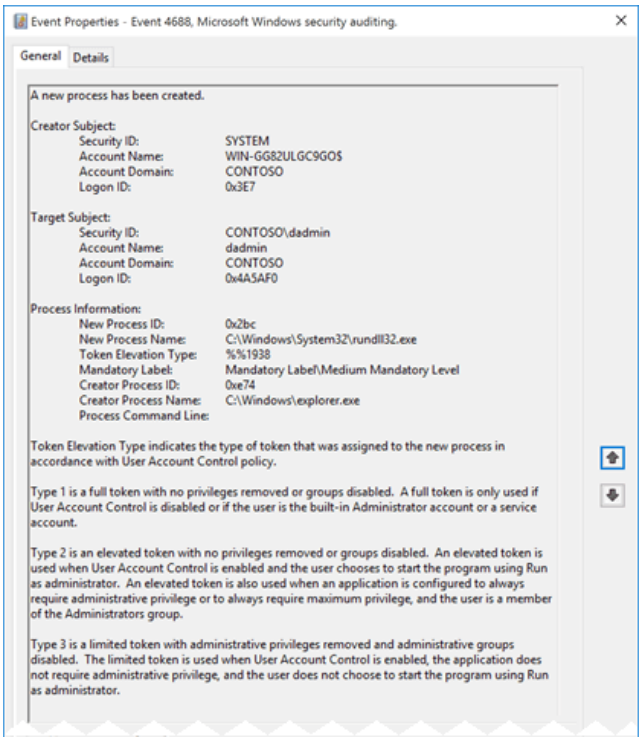

Рис. 2: Пример события с идентификатором 4688, регистрируемого стандартными средствами Windows

Кроме того, Sysmon предоставляет достаточно широкие возможности для своего конфигурирования и фильтрации собираемых событий.

Процесс сбора и анализа логов Sysmon включает в себя несколько ключевых этапов:

1. Установка и развертывание – установка и

развертывание Sysmon в системах, чтобы начать сбор информации о происходящих в интересующих системах событиях.

2. Конфигурация – процесс настройки Sysmon, при котором указываются необходимые для мониторинга события, а также место назначения для ведения журнала. Для настройки Sysmon можно использовать файл конфигурации, в котором указаны события для мониторинга и ведения журнала. Настройка конфигурационного файла службы может быть проведена в соответствии с уникальными особенностями и потребностями системы, определяя, например, необходимые для регистрации типы событий.
3. Сбор данных. Журналы Sysmon обычно публикуются в формате XML в журнале событий Windows (Windows Event Log). Для сбора журналов Sysmon могут использоваться различные методы, такие как переадресация событий Windows (WEF), централизованное решение для ведения журнала или решение SIEM. Используя эти методы, можно объединить журналы из нескольких систем в одном месте для их дальнейшего анализа.
4. Анализ собранных логов – проведение анализа собранных на предыдущем этапе данных, используя ручные методы или инструменты автоматизированной обработки. Логи Sysmon содержат различные типы событий, включая создание процессов, сетевые подключения, создание или модификацию файлов, изменения реестра и многое другое, для выявления необычной и подозрительной активности, индикаторов компрометации и понимания поведения системы.
5. Реагирование на инциденты и форензика. Проанализированные во время реагирования на инциденты безопасности Sysmon логи могут использоваться для восстановления временных рамок инцидента, отслеживания действий злоумышленников и определения последствий инцидентов безопасности.

Также стоит отметить, что существует аналогичное решение для операционных систем на базе ядра Linux – SysmonForLinux [3]. Данное решение представляет собой порт Sysmon для операционных системы на базе Linux, который был также разработан компанией Microsoft. SysmonForLinux поддерживает некоторое ограниченное подмножество событий, регистрируемых Sysmon, при этом собирая все те же данные о событии, что и оригинальное приложение. Перечень регистрируемых приложением SysmonForLinux версии 1.3.2 приведен в Таблице 2.

Таблица 2: События, регистрируемые SysmonForLinux версии 1.3.2

ID	Описание регистрируемого события
1	Создание процесса
3	Сетевая активность процессов
4	Запуск и остановка службы Sysmon
5	Завершение процесса
9	Прямой доступ к дискам и томам
10	Доступ одного процесса к памяти другого процесса
11	Создание файлов
16	Изменение конфигурации Sysmon
23	Удаление файла с сохранением его копии с каталоге-архиве Sysmon
255	Ошибка службы Sysmon

III. Подходы к анализу логов Sysmon

A. Использование графов

Как уже было отмечено, сама по себе служба Sysmon не предоставляет возможности для анализа собранных ею сведений. Один из подходов для анализа и выявления вредоносного программного обеспечения заключается в построении по имеющимся логам ориентированного графа, отражающего логическую последовательность произошедших в системе действий. Построенный и визуализированный граф представляет собой графический способ выявления аномалий в системе, заключающийся в поиске и анализе наиболее обособленных вершин графа.

Например, приложение Grafiki [4] с открытым исходным кодом предоставляет возможность построения и визуализации по имеющемуся набору логов Sysmon ориентированного графа, вершинами которого являются присутствующие в полях логов Sysmon сущности, такие как, например, имя процесса, сетевой адрес, имя файла, ключ реестра, а ориентированными дугами – производимые над данными сущностями действия, соответствующие типам событий Sysmon, такие как, например, создание, изменение, удаление, загрузка, подключение. Пример графа, построенного средствами Grafiki, приведен на Рис. 3.

Для выявления аномалий в Sysmon логах также может быть применен подход с построением взвешенного графа угроз (см. [5] и [6]). Идея такого подхода заключается в построении ориентированного графа процессов, дугам которого присвоены некоторые веса, соответствующие частоте запуска дочернего процесса в присутствующем наборе логов. Подграфы с более низким средним весом в сравнении со средним весом всего графа являются подозрительным и подлежат более внимательному анализу, так как редко посещаемая вершина графа является аномальной зоной.

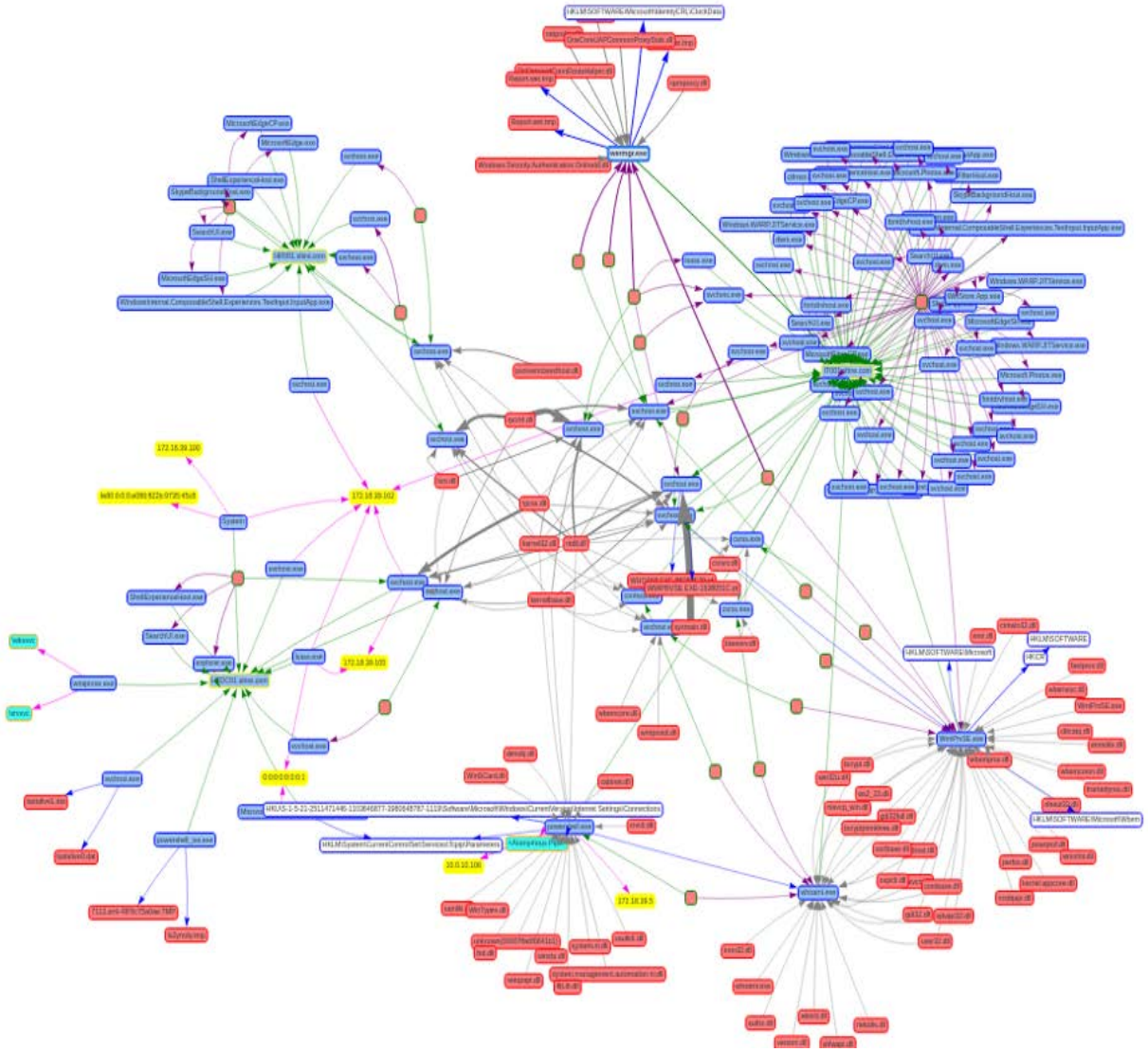

Рис. 3: Граф, построенный с помощью средства Grafiki

B. Использование специально настроенных конфигураций Sysmon

В работе [7] авторами проводится исследование, сосредоточенное на определении оптимальных конфигураций систем для методов обнаружения следов перемещения внутри периметра (LM, Lateral Movement). Перемещение внутри периметра – это тактика, используемая злоумышленниками для перемещения по сети после начальной компрометации, с целью доступа к ценным данным и повышения своих привилегий. На основе базы знаний MITRE ATT&CK [8] и Sysmon в ходе исследования был разработан инструмент PeX для автоматизации разбора и анализа журналов Sysmon. Инструмент и набор данных находятся в открытом доступе, предоставляя ценный ресурс для дальнейших исследований и практического применения [9].

Авторы сосредоточились на исследовании атак с использованием удаленных служб и атак с использованием учетных данных. PeX использует настраиваемые правила, основанные на знаниях о методах атаки из базы данных MITRE ATT&CK. Это позволяет адаптировать алгоритм под специфические требования различных организаций и улучшать его эффективность в зависимости от текущих угроз. PeX показал высокую точность в обнаружении перемещения внутри периметра.

В работе [10] авторами разрабатывается инструмент для автоматического извлечения техник ATT&CK из журналов событий Sysmon, а также предлагается эффективная система прогнозирования перемещения внутри периметра с использованием метода Quantification Theory Type 3 и техник ATT&CK. Авторами было разработано веб-приложение, которое извлекает данные Sysmon логов с устройств Windows и сопоставляет их с матрицей ATT&CK. Это также позволяет визуализировать техники, используемые злоумышленниками, и понять, какие именно этапы атаки были выполнены. Сначала логи собираются с различных устройств внутри сети, затем собранные данные анализируются и сопоставляются с соответствующими техниками в матрице ATT&CK, после чего результаты отображаются в удобной для понимания визуальной форме, что помогает быстро идентифицировать угрозы. Предлагаемая система обнаружения перемещения внутри периметра предназначена для обнаружения перемещений злоумышленников внутри сети после первоначального проникновения. Она использует Quantification Theory

Type 3 для анализа данных и выявления закономерностей, указывающих на перемещение внутри периметра. Quantification Theory Type 3 – это статистический метод, который используется для анализа категориальных данных [11]. В контексте данного исследования он помогает количественно оценить взаимосвязи между различными техниками атак и их потенциальным влиянием на безопасность сети.

Еще одно открытое решение – Sysmon modular [12]. Это решение представляет собой настраиваемую структуру, разработанную для улучшения возможностей Sysmon. Sysmon Modular выделяется благодаря своему модульному подходу к конфигурации, что делает его весьма эффективным для анализа логов и обнаружения вредоносного ПО. Sysmon Modular разбивает конфигурацию Sysmon на несколько небольших, управляемых частей. Каждый модуль фокусируется на определенных типах событий или данных журналов, таких как создание процессов, сетевые подключения, изменения файлов и т.д. Такой модульный подход облегчает настройку и обновление конфигураций по мере необходимости. Пользователи могут включать или отключать определенные модули в зависимости от своих нужд. Эта гибкость позволяет создавать индивидуальные конфигурации, которые могут фокусироваться на конкретных аспектах мониторинга системы, повышая эффективность анализа логов. Фокусируясь на релевантных событиях и уменьшая шум, Sysmon Modular помогает специалистам по безопасности эффективно выявлять и реагировать на угрозы.

C. Использование интеллектуального анализа процессов

Под интеллектуальным анализом процессов (Process mining) понимается технология извлечения знаний из журналов событий информационных систем и восстановление моделей данных систем на основе извлеченных данных (см. [13]).

Журналы событий (в том числе логи Sysmon) могут быть использованы для осуществления трех форм process mining: извлечение процесса, проверка соответствия и усовершенствование процесса. Методы извлечения процессов используют журнал событий для получения модели процесса без использования какой-либо априорной информации. Извлечение процессов является самой значимой составляющей process mining. При проверке соответствия производится сопоставление существующей модели процесса с журналом событий этого же процесса. Проверка соответствия может быть использована для оценки того, насколько реальные данные журнала соответствуют модели, и наоборот. При усовершенствовании процесса существующую модель процесса улучшают с использованием информации о реально осуществляемом процессе, зафиксированном в каком-либо журнале событий. В данном случае, в отличие от проверки соответствия, речь идет об изменении и/или расширении априорной модели процесса.

В работах [14] и [15] описывается построение модели с использованием алгоритмов Process mining, реализованных средствами библиотеки *Pm4py* языка программирования Python. В результате применения и оценки различных алгоритмов строится модель со значением уровня соответствия модели журналам событий в 96%. Для построенной таким образом модели, соответствующей штатному поведению системы, восстановленному по записям Sysmon логов, был сделан вывод о том, что она может быть использована для выявления событий, выходящих за рамки обыденного поведения системы.

D. Использование онтологий

В работе [16] проектируется высокоуровневая система оценки угроз программного обеспечения, использующая логи Sysmon, а также онтологию Cyber Threat Intelligence Ontology (CTIO) для классификации исполняющегося программного обеспечения на основе четырех категорий: высокий уровень угрозы (вредоносное программное обеспечение; легитимное или неизвестное ранее программное обеспечение, связанное с вредоносными индикаторами), средний уровень угрозы (легитимное программное обеспечение, имеющее уязвимость; легитимное программное обеспечение, использующееся злоумышленником для проведения атаки), небольшой уровень угрозы (потенциально не вредоносное программное обеспечение), неизвестное ранее программное обеспечение (неизвестное программное обеспечение без известных связей с вредоносными индикаторами).

Онтология CTIO состоит из нескольких субонтологий, основанных на существующих универсально используемых таксономиях, таких как CVE, CWE, ATT&CK, разработанных онтологиях вредоносного программного обеспечения и ExtendedCPE и других. Онтология вредоносного программного обеспечения и онтология ExtendedCPE являются основными компонентами CTIO и предназначены для предоставления точных сведений о вредоносном и безопасном программном обеспечении.

Архитектура предлагаемой высокоуровневой системы автоматизированной оценки угроз приведена на Рис. 4. Авторами определяется следующий порядок работы данной системы:

1. Производится сбор журналов Sysmon с различных устройств;
2. С учетом идентификатора события происходит извлечение полей события для дальнейшего проведения оценки угрозы. Например, для события с идентификатором 1 (создание процесса) извлекаются в частности имя пользователя, хэш процесса, содержимое командной строки данного и родительского процессов.
3. Механизм поиска (Lookup Engine) проверяет, включен ли процесс во внутренний белый список хэшей компонента онтологии ExtendedCPE, и извлекает соответствующий уровень угрозы.

Уровень угрозы неопасного процесса может изменяться на основе новой информации, например, в случае обновления информации в базе данных общеизвестных уязвимостей. При этом экземпляры неопасного программного обеспечения, связанные с определенным уровнем угрозы, могут быть дополнительно проверены в соответствии с их поведением. Сведения о событиях, которые подлежат дальнейшему изучению, включаются в специальную базу знаний.

4. Механизм поиска проверяет, были ли ранее запрошены извлеченные значения элементов, такие как хэши и командные строки, в течение указанного периода времени, и извлекает соответствующую информацию (быстрая проверка безопасного или вредоносного программного обеспечения). При наличии уже классифицированного процесса система передает информацию непосредственно в механизм процесса принятия решений.
5. Значения элементов неопознанных процессов становятся частью запросов SPARQL, выполняемых механизмом SPARQL (SPARQL Engine), которые выполняют семантический поиск в базе знаний CTIO. Основываясь на полученной информации механизм принятия решений классифицирует процесс как процесс высокого уровня угрозы, среднего уровня угрозы, небольшого уровня угрозы или как неизвестный процесс.
6. Процессы, которые были классифицированы как неизвестные, либо считаются безопасными после ручной проверки, либо подвергаются дальнейшему исследованию через определенные промежутки времени путем сопоставления с новыми данными.

E. Использование алгоритмов машинного и глубокого обучения

Для более эффективного анализа Sysmon логов могут быть применены технологии машинного и глубокого обучения. Обучая модели машинного и глубокого обучения на известных исторических данных системных журналов, а также применяя известные сведения о тактиках, техниках и процедурах, можно разработать прогностическую модель, которая будет с высокой точностью выявлять аномалии и потенциальные атаки в системе.

В статье [17] предлагается построить нейронную сеть на основе управляемых рекуррентных блоков (GRU) для автоматической идентификации подозрительной активности. Предлагаемая модель бинарной классификации состоит из четырех слоев: входного слоя, embedding слоя, слоя GRU и выходного слоя. В качестве функции активации используется сигмоидная функция активации.

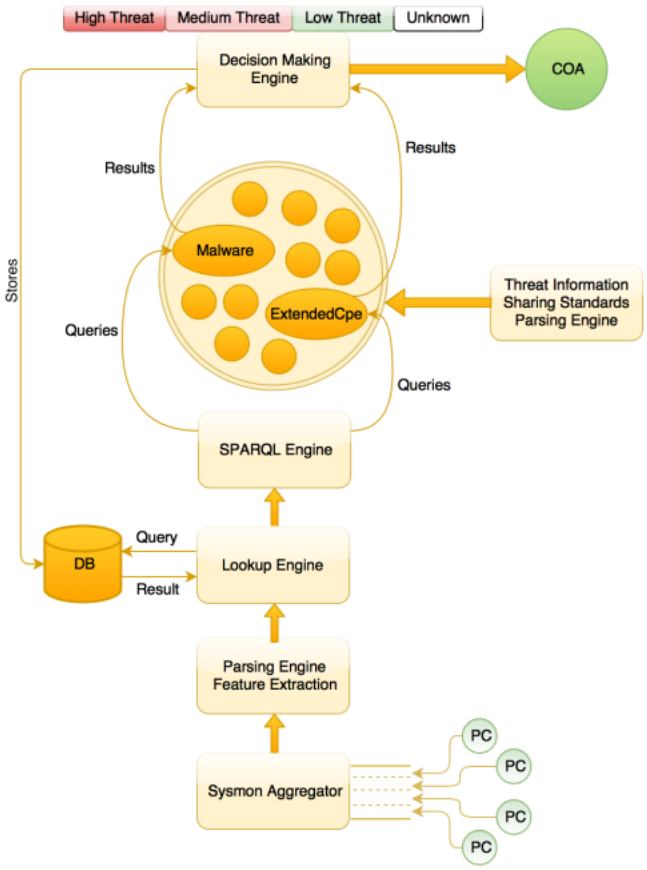

Рис. 4: Предлагаемая в работе [16] архитектура системы оценки угроз

Схема модели приведена на Рис. 5. В результате исследований авторов работы было установлено, что большинство атак содержат следующие четыре категории поведения: процесс, доступ к файлу, реестр, доступ к сети. Обучение модели производится на основе относящихся к данным категориям типов событий Sysmon, перечень которых (вместе с их атрибутами, непосредственно участвующими в обучении) приведен в Таблице 3.

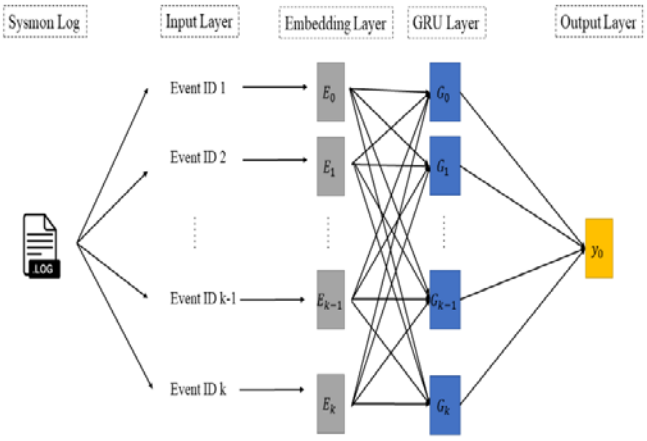

Рис. 5: Предлагаемая в работе [17] схема модели

Таблица 3: Выбранные для анализа в работе [17] типы событий и их атрибуты

ID	Атрибуты события
1	Event ID, Image, User, ParentImage
2	Event ID, TargetFilename, CreationUtcTime, PreviousCreationTime
3	Event ID, Protocol, Initiated, SourcePort, DestinationPort
5	Event ID
7	Event ID, ImageLoaded, Signed
8	Event ID, TargetImage
9	Event ID
10	Event ID, TargetImage, GrantedAccess
11	Event ID, TargetFileName
12	Event ID, EventType
13	Event ID
14	Event ID, EventType
15	Event ID, TargetFileName

При построении модели было использовано 47175 записей логов, соответствующих активности легитимного программного обеспечения, и 10048 записей логов, соответствующих активности вредоносного программного обеспечения.

На кросс-валидации была достигнута точность в 95.52%. На экземплярах ранее неизвестного вредоносного программного обеспечения была получена точность в 93.23%.

На основе результатов проведенных экспериментов был сделан вывод о том, что предлагаемая модель эффективно классифицирует поведение как легитимного, так и вредоносного программного обеспечения в системе, а также хорошо идентифицирует неизвестное вредоносное программное обеспечение.

В работе [18] спроектирован и представлен подход LogGD, представляющий собой метод обнаружения аномалий с использованием графовых нейронных сетей (GNN). Последовательности логов преобразуются в графы, где узлы представляют события журнала, а ребра – взаимосвязи между ними. Используя пространственную структуру и семантическую информацию о событиях журнала, LogGD эффективно обнаруживает аномалии в системных журналах, включая те, которые генерируются Sysmon.

LogGD состоит из трех основных компонентов: построение графов на основе предобработанных логов, обучение представлений графов и классификация графов. Схема работы изображена на Рис. 6.

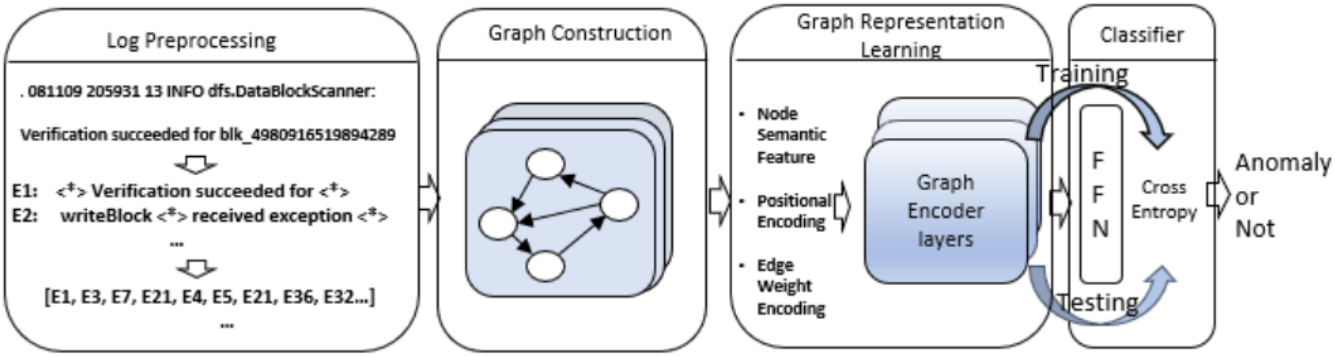

Рис. 6: Предлагаемая в работе [18] схема работы LogGD

На этапе построения графов последовательности предобработанных логов преобразуются в направленные графы, где узлы представляют собой события журналов, а рёбра представляют частоту возникновения и совместную встречаемость этих событий. Возникающая графовая структура включает в себя локальную структуру, ориентированную на узлы, представленную матрицей степеней графа, глобальную структуру местоположения узлов, кодируемую матрицей расстояний, и количественные связи между узлами, представленные матрицей весов построенного графа. Это преобразование отражает как семантическую, так и структурную информацию событий журналов, улучшая представление данных логов для обнаружения аномалий.

Суть следующего этапа – обучение представлений графов – закодировать структурную информацию о графе в пространство меньшей размерности, например, представить вершины графа или целиком граф (подграфы) как точки в новом графе. При этом цель алгоритма/модели – чтобы в получившемся пространстве геометрические соотношения отражали структуру исходного графа, например, близкие вершины в пространстве были также близки (связаны ребром, имели небольшой кратчайший путь) в графе. Для построения сети GNN был выбран метод Graph Transformer Network (GTN) [19], так как он демонстрирует лучшую производительность при классификации графов с использованием структурного кодирования.

На последнем этапе после изучения представлений графов модель классифицирует их, чтобы определить, является ли последовательность журналов нормальной или аномальной. Этот процесс включает обучение GNN на размеченных данных и использование изученных представлений для определения аномалий по новым последовательностям логов. Представления графа подаются на слой FFN с функцией активации GELU (Gaussian Error Linear Unit), после чего для классификации выход подается функции *softmax*.

Результаты тестирования построенной сети на различных наборах данных и подсчета основных метрик приведены на Рис. 7. По двум основным причинам

LogGD работает лучше, чем аналогичные подходы. Во-первых, LogGD может фиксировать более выразительную структурную информацию из графиков, чем просто последовательные связи между событиями журнала. Эти расширенные возможности помогают LogGD лучше идентифицировать аномальные последовательности логов. Во-вторых, настраиваемая модель GTN отражает взаимодействие между узловыми объектами и структурой графа, представленной кратчайшим относительным расстоянием пути, что также может повысить эффективность обнаружения аномалий.

Dataset	Metrics	LogGD	LR	SVM	LogRobust	CNN	NeuralLog
HDFS	F1	**0.9877**	0.9616	0.8330	0.9819	0.9872	0.9827
	Precision	0.9774	0.9603	0.9519	0.9688	**0.9852**	0.9627
	Recall	**0.9982**	0.9629	0.7405	0.9954	0.9891	0.9956
BGL	F1	**0.9719**	0.2799	0.4558	0.9402	0.9140	0.9535
	Precision	**0.9708**	0.1684	0.8190	0.9229	0.8669	0.9586
	Recall	**0.9731**	0.8286	0.3158	0.9596	0.9702	0.9484
Spirit	F1	**0.9789**	0.9652	0.9736	0.9757	0.9652	0.9510
	Precision	0.9889	0.9580	0.9773	**0.9957**	0.9740	0.9694
	Recall	0.9691	**0.9724**	0.9699	0.9566	0.9566	0.9349
TDB	F1	**0.9284**	0.4651	0.7797	0.4043	0.5533	0.7704
	Precision	**0.9772**	0.3390	0.7188	0.4329	0.5405	0.9683
	Recall	**0.8889**	0.7407	0.8519	0.4198	0.5802	0.6437

Рис. 7: Результаты тестирования LogGD

В работе [20] также с использованием графовых нейронных сетей предлагается метод, который представляет логи в виде атрибутивных, направленных и взвешенных графов, что позволяет эффективно обнаруживать аномалии. Узлы в графе соответствуют лог-событиям, а рёбра указывают на последовательность событий. Вес ребра показывает количество последовательных пар событий. Авторами предлагается метод для выявления таких аномалий, как количественные аномалии (возникают, если частота появления некоторых событий выше или ниже ожидаемой) и последовательные аномалии (возникают, если порядок событий отличается от нормального). Для обучения представлений узлов в графах используются сверточные сети DiGCN [21]. Эта модель обучается на графах для выявления аномалий, учитывая как атрибуты узлов, так и веса рёбер. DiGCN позволяет эффективно выявлять аномалии в структуре графа, что повышает точность обнаружения. Предложенный метод обладает несколькими ключевыми преимуществами, как, например, высокая выразительность графовых представлений, что позволяет более точно моделировать логи, а также возможность обнаружения сложных структурных аномалий.

В работе [22] исследуется зависимость типов событий безопасности и используемых моделей машинного обучения для их обнаружения, которые при использовании на определенных событиях лучше бы моделировали данные и приводили бы к наименьшему количеству ложноположительных событий. В частности, в данной работе предлагается использовать модель One Class SVM (OCSVM) [23] для обнаружения аномалий на событиях Sysmon, связанных с созданием процессов (события с ID 1).

IV Заключение

В работе были рассмотрены различные современные методы и подходы к анализу логов Sysmon для поиска и обнаружения вредоносной активности и активности вредоносного программного обеспечения. Рассмотренные исследования и открытые решения демонстрируют существующее многообразие методологий и инструментов для анализа логов Sysmon, начиная от систем на основе графов, онтологий и настроенных пользовательских наборов конфигураций и заканчивая передовыми методами машинного и глубокого обучения, такими как графические нейронные сети. Использование этих подходов может значительно повысить эффективность обнаружения вредоносных действий и повысить общий уровень защиты и кибербезопасности.

Благодарности

Работа написана в рамках развития программы магистратуры факультета ВМК МГУ имени М.В. Ломоносова “Кибербезопасность” (МГУ-Сбер) [24, 25].

Редакция журнала традиционно отмечает, что все публикации в журнале INJOIT, связанные с цифровой повесткой, начинались с работ В.П. Куприяновского и его многочисленных соавторов [26-28]

Библиография

[1] The Hidden Costs of Cybercrime // McAfee Report, 2020
[2] Sysmon https://learn.microsoft.com/en-us/sysinternals/downloads/sysmon (дата обращения 10.05.2024)
[3] Sysmon for Linux https://github.com/Sysinternals/SysmonForLinux (дата обращения 10.05.2024)
[4] Grafiki https://github.com/lucky-luk3/Grafiki (дата обращения 10.05.2024)
[5] Sysmon Threat Analysis Guide https://www.varonis.com/blog/sysmon-threat-detection-guide (дата обращения 10.05.2024)
[6] Sysmon Visualizaton and Tools https://github.com/agreenjay/sysmon/tree/master (дата обращения 10.05.2024)
[7] Smiliotopoulos, C.; Barmpatsalou , K.; Kambourakis, G. Revisiting the Detection of Lateral Movement through Sysmon. Appl. Sci. 2022, 12, 7746
[8] Mitre https://attack.mitre.org/ (дата обращения 24.05.2024)
[9] Python_Evtx_Analyzer https://github.com/ChristosSmiliotopoulos/Python_Evtx_Analyzer (дата обращения 24.05.2024)
[10] Okada, Satoshi, et al. "Predicting and Visualizing Lateral Movements Based on ATT&CK and Quantification Theory Type 3." JCIT vol.26, no.1 2024: pp.1-14.
[11] Landini, Gregory. “Quantification Theory in *9 of Principia Mathematica.” History and Philosophy of Logic 21, no. 1 (2000): 57–77.
[12] A Sysmon configuration repository for everybody to customise https://github.com/olafhartong/sysmon-modular (дата обращения 25.05.2024)

[13] Van Der Aalst W. et al. Process Mining manifesto //International Conference on Business Process Management. 2011. Т. 99, С. 169-194.
[14] Хасанова А.М., Интеллектуальный анализ процессов по данным журналов событий информационных систем // International Journal of Open Information Technologies. 2022. Т. 10, № 10.
[15] Хасанова А.М., Дунаев М.Е. Применение технологии Process mining для выявления аномальных ситуаций в работе наукоемкого оборудования // International Journal of Open Information Technologies. 2022. Т. 9, № 8.
[16] Mavroeidis V, Jøsang A Data-Driven Threat Hunting Using Sysmon // Proceedings of the 2nd international conference on cryptography, security and privacy. 2018. С. 82-88
[17] Chen C, Syu G, Cai Z Analyzing System Log Based on Machine Learning Model // International Journal of Network Security. 2020. Т. 22, № 6, С. 925-933.
[18] Y. Xie, H. Zhang and M. A. Babar, "LogGD: Detecting Anomalies from System Logs with Graph Neural Networks," 2022 IEEE 22nd International Conference on Software Quality, Reliability and Security (QRS), Guangzhou, China, 2022, pp. 299-310
[19] Dwivedi, Vijay Prakash and Xavier Bresson. “A Generalization of Transformer Networks to Graphs.” ArXiv abs/2012.09699 (2020)
[20] Li, Zhong and Shi, Jiayang and van Leeuwen, Matthijs, Graph Neural Networks Based Log Anomaly Detection and Explanation. Available at SSRN: https://ssrn.com/abstract=4627217 or http://dx.doi.org/10.2139/ssrn.4627217
[21] Z. Tong, Y. Liang, C. Sun, X. Li, D. Rosenblum, and A. Lim. Digraph inception convolutional networks. Advances in neural information processing systems, 33:17907–17918, 2020.
[22] H. M. Farooq and N. M. Otaibi, "Optimal Machine Learning Algorithms for Cyber Threat Detection," 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, 2018, pp. 32-37, doi: 10.1109/UKSim.2018.00018.
[23] Schölkopf, Bernhard, and Alexander J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002.
[24] Магистерская программа “Кибербезопасность” МГУ-Сбер https://cyber.cs.msu.ru/ (дата обращения 30.10.2024)
[25] Сухомлин В. А. Концепция и основные характеристики магистерской программы" Кибербезопасность" факультета ВМК МГУ //International Journal of Open Information Technologies. – 2023. – Т. 11. – №. 7. – С. 143-148.
[26] Розничная торговля в цифровой экономике / В. П. Куприяновский, С. А. Синягов, Д. Е. Намиот [и др.] // International Journal of Open Information Technologies. – 2016. – Т. 4, № 7. – С. 1-12. – EDN WCMIWN.
[27] Развитие транспортно-логистических отраслей Европейского Союза: открытый BIM, Интернет Вещей и кибер-физические системы / В. П. Куприяновский, В. В. Аленьков, А. В. Степаненко [и др.] // International Journal of Open Information Technologies. – 2018. – Т. 6, № 2. – С. 54-100. – EDN YNIRFG.
[28] Умная инфраструктура, физические и информационные активы, Smart Cities, BIM, GIS и IoT / В. П. Куприяновский, В. В. Аленьков, И. А. Соколов [и др.] // International Journal of Open Information Technologies. – 2017. – Т. 5, № 10. – С. 55-86. – EDN ZISODV.

Sysmon Log Analysis Methods for Cyber Threat Detection

Egor V. Kostikov

***Abstract*— In the modern world, where cyber threats are becoming more sophisticated, the analysis of system logs plays a key role in ensuring the security of information systems. Logs provide valuable information about events occurring on the network, allowing you to detect abnormal actions that may indicate attacks or security breaches. Regular monitoring and analysis of this data helps in the rapid identification of incidents, which contributes to a prompt response and minimization of damage. In addition, system logs are an important tool for investigating incidents, allowing you to determine the causes and extent of attacks. With the constant increase in the number of cyber threats, competent log analysis is becoming vital to protect organizations and their data.**

Sysmon is a powerful log analysis solution that can significantly improve the security of the information infrastructure. With Sysmon, you can monitor changes in the file system, network connections, and running processes, which allows you to identify suspicious activities and anomalies. This tool integrates with SIEM systems, which simplifies data analysis and correlation.

This paper is devoted to an overview of existing methods and solutions for analyzing Sysmon logs to detect malicious software.

***Keywords*—cybersecurity, sysmon, log analysis, malware, detection of malicious activity.**

REFERENCES

[1] The Hidden Costs of Cybercrime // McAfee Report, 2020
[2] Sysmon https://learn.microsoft.com/en-us/sysinternals/downloads/sysmon (data obrashhenija 10.05.2024)
[3] Sysmon for Linux https://github.com/Sysinternals/SysmonForLinux (data obrashhenija 10.05.2024)
[4] Grafiki https://github.com/lucky-luk3/Grafiki (data obrashhenija 10.05.2024)
[5] Sysmon Threat Analysis Guide https://www.varonis.com/blog/sysmon-threat-detection-guide (data obrashhenija 10.05.2024)
[6] Sysmon Visualizaton and Tools https://github.com/agreenjay/sysmon/tree/master (data obrashhenija 10.05.2024)
[7] Smiliotopoulos, C.; Barmpatsalou , K.; Kambourakis, G. Revisiting the Detection of Lateral Movement through Sysmon. Appl. Sci. 2022, 12, 7746
[8] Mitre https://attack.mitre.org/ (data obrashhenija 24.05.2024)
[9] Python_Evtx_Analyzer https://github.com/ChristosSmiliotopoulos/Python_Evtx_Analyzer (data obrashhenija 24.05.2024)
[10] Okada, Satoshi, et al. "Predicting and Visualizing Lateral Movements Based on ATT&CK and Quantification Theory Type 3." JCIT vol.26, no.1 2024: pp.1-14.
[11] Landini, Gregory. “Quantification Theory in *9 of Principia Mathematica.” History and Philosophy of Logic 21, no. 1 (2000): 57–77.
[12] A Sysmon configuration repository for everybody to customise https://github.com/olafhartong/sysmon-modular (data obrashhenija 25.05.2024)
[13] Van Der Aalst W. et al. Process Mining manifesto //International Conference on Business Process Management. 2011. T. 99, S. 169-194.
[14] Hasanova A.M., Intellektual'nyj analiz processov po dannym zhurnalov sobytij informacionnyh sistem // International Journal of Open Information Technologies. 2022. T. 10, # 10.
[15] Hasanova A.M., Dunaev M.E. Primenenie tehnologii Process mining dlja vyjavlenija anomal'nyh situacij v rabote naukoemkogo oborudovanija // International Journal of Open Information Technologies. 2022. T. 9, # 8..
[16] Mavroeidis V, Jøsang A Data-Driven Threat Hunting Using Sysmon // Proceedings of the 2nd international conference on cryptography, security and privacy. 2018. С. 82-88
[17] Chen C, Syu G, Cai Z Analyzing System Log Based on Machine Learning Model // International Journal of Network Security. 2020. Т. 22, № 6, С. 925-933.
[18] Y. Xie, H. Zhang and M. A. Babar, "LogGD: Detecting Anomalies from System Logs with Graph Neural Networks," 2022 IEEE 22nd International Conference on Software Quality, Reliability and Security (QRS), Guangzhou, China, 2022, pp. 299-310
[19] Dwivedi, Vijay Prakash and Xavier Bresson. “A Generalization of Transformer Networks to Graphs.” ArXiv abs/2012.09699 (2020)
[20] Li, Zhong and Shi, Jiayang and van Leeuwen, Matthijs, Graph Neural Networks Based Log Anomaly Detection and Explanation. Available at SSRN: https://ssrn.com/abstract=4627217 or http://dx.doi.org/10.2139/ssrn.4627217
[21] Z. Tong, Y. Liang, C. Sun, X. Li, D. Rosenblum, and A. Lim. Digraph inception convolutional networks. Advances in neural information processing systems, 33:17907–17918, 2020.
[22] H. M. Farooq and N. M. Otaibi, "Optimal Machine Learning Algorithms for Cyber Threat Detection," 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, 2018, pp. 32-37, doi: 10.1109/UKSim.2018.00018.
[23] Schölkopf, Bernhard, and Alexander J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002.
[24] Magisterskaja programma “Kiberbezopasnost’” MGU-Sber https://cyber.cs.msu.ru/ (data obrashhenija 30.10.2024)
[25] Suhomlin V. A. Koncepcija i osnovnye harakteristiki magisterskoj programmy" Kiberbezopasnost'" fakul'teta VMK MGU //International Journal of Open Information Technologies. – 2023. – T. 11. – #. 7. – S. 143-148
[26] Roznichnaja torgovlja v cifrovoj jekonomike / V. P. Kuprijanovskij, S. A. Sinjagov, D. E. Namiot [i dr.] // International Journal of Open Information Technologies. – 2016. – T. 4, # 7. – S. 1-12. – EDN WCMIWN.
[27] Razvitie transportno-logisticheskih otraslej Evropejskogo Sojuza: otkrytyj BIM, Internet Veshhej i kiber-fizicheskie sistemy / V. P. Kuprijanovskij, V. V. Alen'kov, A. V. Stepanenko [i dr.] // International Journal of Open Information Technologies. – 2018. – T. 6, # 2. – S. 54-100. – EDN YNIRFG.
[28] Umnaja infrastruktura, fizicheskie i informacionnye aktivy, Smart Cities, BIM, GIS i IoT / V. P. Kuprijanovskij, V. V. Alen'kov, I. A. Sokolov [i dr.] // International Journal of Open Information Technologies. – 2017. – T. 5, # 10. – S. 55-86. – EDN ZISODV.