ТЕХНИЧЕСКИЕ НАУКИ

TECHNICAL SCIENCES

УДК 004.056.3
ББК 32.973.53
И 88

Очередько Андрей Романович
Аспирант Кубанского государственного технологического университета, Краснодар, e-mail: andrewlisten@mail.ru
Герасименко Виталий Сергеевич
Аспирант Кубанского государственного технологического университета, Краснодар, e-mail: vs.gerasimenko@mail.ru
Путято Михаил Михайлович
Кандидат технических наук, доцент кафедры компьютерных технологий и информационной безопасности Кубанского государственного технологического университета, Краснодар, e-mail: putyato.m@gmail.com
Макарян Александр Самвелович
Кандидат технических наук, доцент кафедры компьютерных технологий и информационной безопасности Кубанского государственного технологического университета, Краснодар, e-mail: msanya@yandex.ru

Исследование SIEM-систем на основе анализа механизмов выявления кибератак

(Рецензирована)

Аннотация. *Представлен анализ и классификация SIEM-систем на основе анализа используемых механизмов корреляции и прогнозирования. Произведен обзор современного состояния рынка SIEM-систем и введены группы критериев для оценки механизмов идентификации с учетом специфики применяемых корреляционных методов, а также модулей прогнозирования инцидентов. Производится изучение SIEM-системы Splunk: методы и алгоритмы корреляции и прогнозирования. Приведен пример создания трендовой линии множества событий инцидентов информационной безопасности. Выводом может служить тот факт, что существующие алгоритмы на данный момент не позволяют полностью использовать методы и методики выявления киберугроз, так как разработка и реализация этого исследовательского направления требует дополнительных расходов и ресурсов. Необходимо продолжать постоянное расширение и модернизацию функций SIEM-систем для поддержания высокого уровня безопасности в условиях роста количества киберугроз.*

Ключевые слова: *кибербезопасность, SIEM-системы, инцидент информационной безопасности, кибератака.*

Ocheredko Andrey Romanovich
Post-gradient student, Kuban State University of Technology, Krasnodar, e-mail: andrewlisten@mail.ru
Gerasimenko Vitaliy Sergeevich
Post-gradient student, Kuban State University of Technology, Krasnodar, e-mail: vs.gerasimenko@mail.ru
Putyato Mikhail Mikhaylovich
Candidate of Technical Sciences, Associate Professor of the Department of Computer Technologies and Information Security, Kuban State University of Technology, Krasnodar, e-mail: putyato.m@gmail.com
Makaryan Aleksandr Samvelovich
Candidate of Technical Sciences, Associate Professor of the Department of Computer Technologies and Information Security, Kuban State University of Technology, Krasnodar, e-mail: msanya@yandex.ru

Research of SIEM systems based on the analysis of mechanisms for detecting cyber attacks

Abstract. *The article presents the analysis and classification of SIEM systems based on the analysis of the correlation and forecasting mechanisms used. A review of the current state of the SIEM systems market was made and groups of criteria were introduced for evaluating identification mechanisms, taking into account the specifics of the applied correlation methods, as well as incident forecasting modules. The study of the Splunk SIEM system is carried out: methods and algorithms for correlation and forecasting. An example of creating a trend line for a set of information security incident events is given. The conclusion can be the fact that the existing algorithms currently do not allow the full use of methods and techniques for detecting cyber threats, since the development and implementa-*

tion of this research direction requires additional costs and resources. It is necessary to continue to constantly expand and modernize the functions of SIEM systems to maintain a high level of security in the face of an increasing number of cyber threats.

Keywords: *cybersecurity, SIEM systems, information security incident, cyber attack.*

Введение. В настоящее время любая система защиты информации обязана развиваться и адаптироваться к постоянно появляющимся новым видам угроз. Например, одна из таких угроз недалекого будущего – квантовые вычисления, которые одновременно с угрозой предоставляют и новые способы защиты информации в виде квантовой криптографии. Определением параметров таких угроз занимается довольно большое количество ученых [1]. Если рассматривать контекст более «реальных» киберугроз, то это одно из актуальных направлений информационной безопасности.

Оценка состояния защищенности основывается на данных, поступающих из источников информации, количество которых с каждым днем растет. Для решения задач оперативного мониторинга и реагирования на инциденты безопасности существует определенный класс программного обеспечения – SIEM-системы (Security Information and Event Management). Программные продукты реализуют процесс, объединяющий сетевую активность в единый адресный набор данных. Функциональный набор таких систем включает сбор и анализ, а также предоставление информации из сетевых, мобильных и стационарных устройств с учетом информационной безопасности. Управление идентификацией и доступом осуществляется с использованием инструментов менеджмента уязвимостей баз данных, программного обеспечения, хранимых данных и приложений мобильных устройств [2]. Основные функции таких систем:

- Анализ событий;
- обнаружение и уведомление:
  - выявленных аномалий сетевого трафика;
  - неожиданных действий пользователя;
  - неопознанных устройств;
- Сбор и хранение журналов событий;
- Предоставление инструментов для анализа результатов кибератак;
- Использование корреляционных правил;
- Автоматическое оповещение и инцидент-менеджмент;
- Возможность прогнозирования атак;
- Создание отчетов:
  - ежедневных отчетов об инцидентах;
  - еженедельных отчетов о рейтинге нарушителей;
  - отчет по работоспособности устройств;
- Мониторинг событий:
  - Устройств;
  - Серверов;
  - критически важных систем;
- Оповещение заинтересованных лиц;
- Организация базы данных инцидентов информационной безопасности.

Современный рынок SIEM-систем сформировался и базируется на следующих программных продуктах: Micro Focus (HP) ArcSight, IBM Qradar, McAfee ESM, RSA NetWitness, Splunk, RuSIEM и MaxPatrol SIEM [3, 4]. Представленные системы имеют как общие структурные элементы, так и собственные индивидуальные решения. Главной целью любой сферы применения таких программных продуктов является выбор наиболее функционального и в то же время соответствующего по цене решения. Ниже сформируем набор критериев и произведем сравнение SIEM-систем для определения оценки каждой из них.

Проведем анализ использования механизмов обнаружения и идентификации кибератак в SIEM-системах. Сформируем группы критериев для оценки механизмов идентификации с

учетом специфики применяемых корреляционных методов, а также модулей прогнозирования инцидентов. Исходя из принципа комплексности, критерии объединены в группы влияния на эффективность работы SIEM-системы:

1. Корреляция данных – 40%.
    1.1. Наличие предустановленных правил корреляции.
    1.2. Количество предустановленных правил корреляции.
    1.3. Количество используемых методов корреляции.
    1.4. Наличие конструктора управления правилами корреляции.
2. Прогнозирование инцидентов – 40%.
    2.1. Наличие модулей прогнозирования.
    2.2. Относительная размерность основного информационного ресурса в прогнозировании.
    2.3. Использование технологий искусственного интеллекта.
    2.4. Использование алгоритмов машинного обучения.
3. Наличие необходимых функций – 20%.
    3.1. Карточка инцидента.
    3.2. Количество настраиваемых полей.
    3.3. Возможность выделения ложных срабатываний.
    3.4. Журналирование изменений объектов – инициированных пользователями.
    3.5. Журналирование изменений объектов – инициированных системными компонентами.
    3.6. Агрегация событий по типу.

Произведем оценку характеристик SIEM-систем, используя описанные выше формальные критерии. Результаты анализа представлены в таблице 1.

Таблица 1

Сравнение SIEM-систем по группе критериев

№ п/п	Наименование критерия	Micro Focus (HP) ArcSight	IBM Qradar	McAfee ESM	RSA NetWitness	Splunk
1.1.	Наличие предустановленных правил корреляции	0.4	0.4	0.4	0.4	0.4
1.2.	Количество предустановленных правил корреляции	0.4	0.2	0.1	0.2	0.2
1.3.	Количество используемых методов корреляции	0.4	0.4	0.4	0.4	0.4
1.4.	Наличие конструктора управления правилами корреляции	0.4	0.4	0.4	0.4	0.4
2.1.	Наличие модулей прогнозирования	0.4	0.4	0.4	0	0.4
2.2.	Относительная размерность основного информационного ресурса	0.4	0.4	0.4	0.4	0.4
2.3.	Использование технологий искусственного интеллекта	0	0.2	0	0	0.4
2.4.	Использование алгоритмов машинного обучения	0.4	0.4	0.4	0	0.4
3.1.	Карточка инцидента	0.2	0.2	0.2	0.2	0.2
3.2.	Количество настраиваемых полей	0.3	0.2	0.1	0.1	0.3
3.3.	Возможность выделения ложных срабатываний	0.2	0.2	0.2	0.2	0.2
3.4.	Журналирование изменений объектов – инициированных пользователями	0.2	0.2	0.2	0.2	0.2
3.5.	Журналирование изменений объектов – инициированных системными компонентами	0.2	0.2	0.2	0.2	0.2
3.6.	Агрегация событий по типу	0.2	0.2	0.2	0.1	0.1
	Итог	4.1	4	3.6	2.8	4.2

Анализ использования механизмов корреляции, прогнозирования и других параметров в SIEM-системах показывает, что из рассмотренных вариантов лучшие результаты с точки зрения эффективности обнаружения инцидентов показывают приложения Micro Focus (HP), ArcSight и Splunk.

Реализация механизмов корреляции и прогнозирования в SIEM-системах. Следует выделить два вида методов корреляции [5]:

– *сигнатурные* (rule-based);
– *бессигнатурные.*

Для надежности и предсказуемости поведения системы производители в большинстве случаев прибегают к сигнатурным методам, в которых администратор системы должен определить правила идентификации инцидентов. Бессигнатурные методы менее гибкие в настройках, все параметры в них заложены на этапе создания систем, поэтому говорить о полном управлении инцидентами в данных системах невозможно. Приведем основные практические методы (рис. 1).

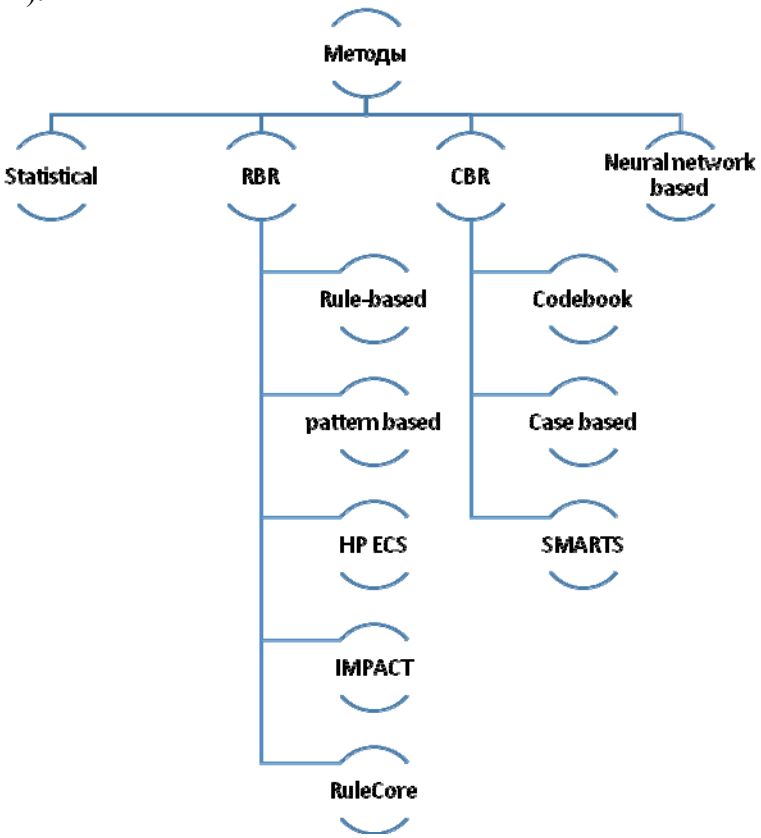

Рис. 1. Основные практические методы

Как правило, в одной реализации SIEM-системы используются один или два метода корреляции. Важно не путать корреляцию с лжекорреляцией – с выводом на экран или в отчет однотипных событий или событий заданного класса. Существует ошибочное мнение, будто технологии продвинулись настолько, что алгоритм может без участия человека «отличать хорошее от плохого», а сигнатурные методы имеют очень много ложных срабатываний.

Разработчики SIEM-систем в основном применяют сигнатурные методы, поскольку они гораздо более гибкие, имеют большую эффективность при обнаружении угроз.

Что касается выбранных двух SIEM-систем, то и ArcSight и Splunk используют в своем решении методы, основанные на rule-based правилах. Естественно, на этом все не заканчивается, разработчики комбинируют методы и постоянно обновляют методику корреляции в их собственном решении.

В Splunk Machine Learning представлено несколько видов алгоритмов (рис. 2.) [6].

Алгоритм линейной регрессии используется для прогнозирования числовых полей. Формирует прогноз значения числового поля, используя взвешенную комбинацию значений

других полей в этом событии. Обычно эти прогнозы используются для выявления аномалий: прогнозы, которые значительно отличаются от фактического значения, могут считаться аномальными. Примерами таких прогнозов могут служить прогнозирование энергопотребления сервера, прогнозирование использования VPN, прогнозирование средней стоимости строения, прогнозирование мощности электростанции и т.д.

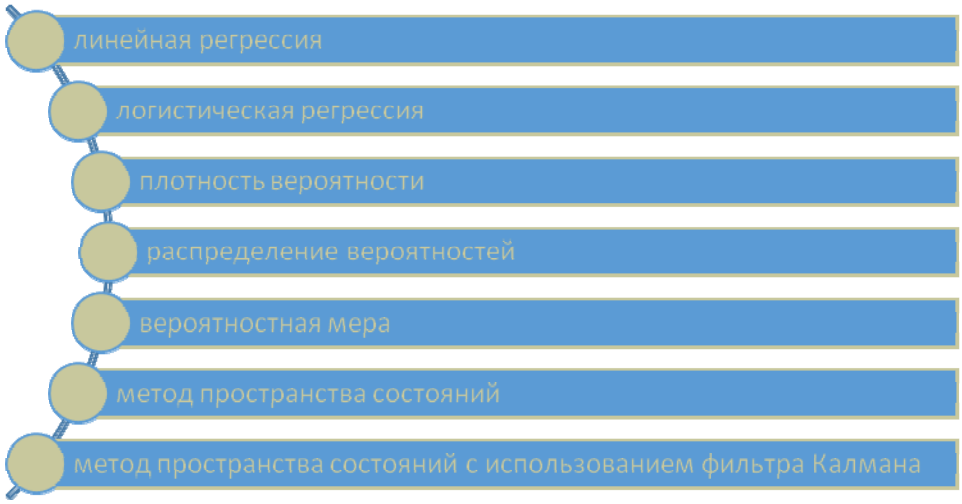

Рис. 2. Виды алгоритмов

Алгоритм логистической регрессии используется для прогнозирования категориальных полей, прогнозирования значения категориального поля, при использовании значений других полей в этом событии. Обычно эти прогнозы используются для выявления аномалий: прогнозы, которые значительно отличаются от фактического значения, могут считаться аномальными. Примерами таких прогнозов могут служить прогнозирование вероятности сбоя жесткого диска, прогнозирование наличия вредоносных программ, прогнозирование внешних аномалий и т.д.

Алгоритм плотности вероятности используется для интеллектуального обнаружения резко отклоняющихся значений. Используя пошаговый управляемый рабочий процесс, происходит поиск числовых выбросов, чтобы использовать алгоритм плотности и сегментировать данные перед поиском аномалий. Примером может служить прогнозирование аномалий в метриках жесткого диска.

Алгоритм распределения вероятностей используется для определения числовых резко отклоняющихся показателей и поиска данных, которые значительно отличаются от предыдущих значений. Примером может служить прогнозирование резко отклоняющихся значений во время ответа сервера, прогнозирование резко отклоняющихся значений в количестве входов в учетную запись (по сравнению с прогнозируемым значением), прогнозирование резко отклоняющихся значений в условиях повышенной влажности на территории электростанции, обнаружение резко отклоняющихся значений в данных колл-центра и т.д.

Алгоритм мер вероятности определяет категорически отклоняющиеся значения и осуществляет поиск событий, которые содержат необычные комбинации значений. Примером может служить обнаружение резко отклоняющихся значений при сбоях диска, обнаружение резко отклоняющихся значений в транзакциях сетей блокчейн, обнаружение резко отклоняющихся значений в ипотечных договорах, обнаружение резко отклоняющихся значений в активности мобильного телефона и т.д.

Алгоритм метод пространства состояний производит интеллектуальное прогнозирование будущих числовых данных временных рядов с помощью пошагового рабочего процесса с возможностью ввода данных из разных источников и учета особых календарных «особых дней», таких, как праздники, дни, памятные для компании. Примером может служить прогнозирование количества звонков в колл-центр, прогнозирование входов в приложение в праздничные дни, прогнозирование расходов на различные приложения и т.д.

Алгоритм метод пространства состояний с использованием фильтра Калмана производит прогноз будущих значений с учетом прошлых значений метрики (числовой временной ряд). Примером может служить прогноз Интернет-трафика, прогнозирование количества входов в учетную запись сотрудников, прогноз ежемесячных продаж, прогноз количества устройств Bluetooth, прогноз обменного курса TWI с использованием ARIMA и т.д.

Для создания трендовой линии множества событий инцидентов информационной безопасности в системе Splunk, используя встроенную команду поиска для создания трендовых линий на основе скользящих средних и линейной регрессии [7]. Опишем алгоритм анализа с помощью встроенного объектно-ориентированного языка в виде макросов:

```
eventstats count as numevents sum($x$) as sumX sum($y$) as sumY sum(eval
($x$*$y$)) as sumXY sum(eval($x$*$x$)) as sumX2 sum(eval($y$*$y$)) as sumY2
    | eval slope=((numevents*sumXY)-(sumX*sumY))/((numevents*sumX2)-
(sumX*sumX))
    | eval yintercept= (sumY-(slope*sumX))/numevents
    | eval newY=(yintercept + (slope*$x$))
    | eval R=((numevents*sumXY) - (sumX*sumY))/sqrt(((numevents*sumX2)-
(sumX*sumX))* ((numevents*sumY2)-(sumY*sumY)))
    | eval R2=R*R
```

Макрос получает два аргумента, значения x и y от каждого события и создает следующие поля для каждого события:

- slope – наклон линии тренда;
- yintercept – y-пересечение линии тренда;
- R – коэффициент корреляции;
- R2 – коэффициент детерминации;
- NewY – значения линии тренда.

Первая временная диаграмма генерирует исходные данные выборки (график количества событий во времени), который затем передается в макрос, который генерирует новые значения оси y тренда. Во втором разделе временной диаграммы отображаются как исходные значения, так и линия тренда на диаграмме, которую можно визуализировать.

Получим коэффициент корреляции (R) или коэффициент детерминации (R2). Вышеприведенный макрос также автоматически вычисляет R и R2 для графика. Чтобы отобразить только R2 (например, чтобы поместить его под диаграммой на приборной панели), необходимо выполнить следующий поиск:

```
sourcetype=my_data | timechart count as yvalue | `lineartrend(_time,yvalue)` | stats first(R2)
```

Для более объективной оценки построим график с использованием визуальных возможностей системы Splunk (рис. 3.).

Рис. 3. Трендовая линия возникновения инцидентов безопасности

Выводы. Современный рынок достаточно насыщен реализациями SIEM-систем. Эти системы масштабируются в зависимости от размера бизнеса. Однако существующие алгоритмы на данный момент не позволяют полностью использовать методы и методики выявления киберугроз, так как разработка и реализация этого исследовательского направления требует дополнительных расходов и ресурсов. Благодаря политике свободного доступа к алгоритмам компании Splunk, мы будем продолжать исследования в области модернизации существующих алгоритмов, методов и технологий.

Анализ использования механизмов корреляции и прогнозирования в результате сравнения характеристик SIEM-систем с использованием предложенных критериев показывает, что наиболее эффективно обрабатывают данные и формируют лучшие результаты обнаружения инцидентов информационной безопасности Micro Focus (HP), ArcSight и Splunk.

Необходимо продолжать постоянное расширение и модернизацию функций SIEM-систем для поддержания высокого уровня безопасности в условиях роста количества киберугроз.

Примечания:

1. Исследование реализации механизмов инкапсуляции ключей постквантовых криптографических методов / А.В. Власенко, М.В. Евсюков, М.М. Путято, А.С. Макарян // Прикаспийский журнал: управление и высокие технологии. Астрахань, 2019. С. 121–127. URL: https://elibrary.ru/item.asp?id=41869980
2. Путято М.М., Макарян М.М. Классификация мессенджеров на основе анализа уровня безопасности хранимых данных // Прикаспийский журнал: управление и высокие технологии. Астрахань, 2019. С. 135–143. URL: https://elibrary.ru/item.asp?id=41869982
3. Сравнение SIEM-систем. Ч. 1. URL: https://www.anti-malware.ru/compare/SIEM-systems#part315 (свободный).
4. Сравнение SIEM-систем. Ч. 2. URL: https://www.anti-malware.ru/compare/SIEM-systems-part2 (свободный).
5. Корреляция SIEM – это просто. Сигнатурные методы. URL: https://www.securitylab.ru/analytics/431459.php (свободный).
6. Splunk® MachineLearningToolkit. URL: http://easycalculation.com/statistics/learn-regression.php (свободный).
7. Splunk // Википедия. URL: https://wiki.splunk.com/Community:Plotting_a_linear_trendline (свободный).

References:

1. Research of key encapsulation mechanisms based on postquantum cryptographic algorithms / A.V. Vlasenko, M.V. Evsyukov, M.M. Putyato, A.C. Makaryan // Caspian Journal: Management and High Technologies. Astrakhan, 2019. P. 121–127. URL: https://elibrary.ru/item.asp?id=41869980
2. Putyato M.M., Makaryan A.C. Classification of messengers based on analysis of the security level of stored data // Caspian Journal: Management and High Technologies. Astrakhan, 2019. P. 135–143. URL: https://elibrary.ru/item.asp?id=41869982
3. Comparison of SIEM systems. Part 1. URL: https://www.anti-malware.ru/compare/SIEM-systems#part315 (free).
4. Comparison of SIEM systems. Part 2. URL: https://www.anti-malware.ru/compare/SIEM-systems-part2 (free).
5. SIEM correlation is easy. Signature Methods. URL: https://www.securitylab.ru/analytics/431459.php (free).
6. Splunk® Machine Learning Toolkit. URL: http://easycalculation.com/statistics/learn-regression.php (free).
7. Splunk // Wikipedia. URL: https://wiki.splunk.com/Community:Plotting_a_linear_trendline (free).