МЕТОДОЛОГИЯ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ УРОВНЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ГЕТЕРОГЕННЫХ СИСТЕМНЫХ ЛОГОВ

Автор: Усмонов Фазлиддин Шарофиддин ўғли
(ООО Unicon.UZ центр научно-технических и маркетинговых исследований, gentlemanfu@gmail.com, +99833 0075700)

Аннотация: *В данной работе мы представляем развернутое исследование методов трансформации «сырых» системных журналов в объективные метрики защищенности. В отличие от стандартных SIEM-подходов, здесь предлагается многоуровневая теоретическая конструкция, объединяющая формальные модели аудита, энтропийный анализ и байесовскую логику. Мы детально разбираем, как корреляция гетерогенных данных позволяет преодолеть проблему «информационного взрыва» и неопределенности. Итогом работы является архитектура аналитического конвейера, способного не просто фиксировать инциденты, но и количественно оценивать динамику вектора состояния безопасности системы.*

Ключевые слова*: информационная безопасность, системные логи, SIEM, обнаружение аномалий, энтропийный анализ, байесовский вывод, машинное обучение, Zero Trust; скрытые марковские модели (HMM); нечеткая логика (fuzzy logic), метрики защищенности.*

Современная ситуация в сфере кибербезопасности заставляет нас признать: реактивные методы защиты окончательно исчерпали свой потенциал. Когда злоумышленник может находиться в сети незамеченным в среднем 16 суток, классические антивирусы и брандмауэры превращаются в «посмертный» инструмент анализа уже случившегося краха. Мы сталкиваемся с необходимостью перехода к парадигме Zero Trust («никому не доверяй, всегда проверяй»), где ключевым активом становятся системные логи — наиболее детальный и объективный след жизнедеятельности ИТ-инфраструктуры.

Однако здесь возникает фундаментальное противоречие. С одной стороны, логи фиксируют всё: от сетевых сессий до атомарных вызовов функций. С другой — колоссальный объем и гетерогенность этих данных создают ситуацию, когда найти реальную угрозу в «белом шуме» практически невозможно без строгой математической базы. Наша цель — предложить такую базу, которая позволит превратить хаотичный поток событий в измеримую шкалу уровня безопасности.

Теоретическая база: переход от простого логирования к глубокому аудиту

Прежде всего, необходимо разделить понятия, которые часто путают в технической литературе. В нашей модели регистрация (logging) — это лишь механизм

записи фактов, тогда как анализ (auditing) — это механизм интерпретации этих фактов относительно политики безопасности. Мы формализуем лог как последовательность кортежей $L = {(t_i, s_i, e_i, m_i)}$, где учитываются временные метки, источники и типы событий. С точки зрения кибернетики, мы рассматриваем поток событий как стохастический процесс.

Для полноты охвата мы выделяем четыре ключевых класса данных, каждый из которых требует своего подхода к анализу:

1. Операционные логи: Жизненный цикл процессов и попытки эскалации привилегий.
2. Сетевые журналы: Паттерны «бокового перемещения» (lateral movement), которые критически важны для обнаружения продвинутых атак.
3. Журналы аутентификации: База для выявления атак типа brute-force или pass-the-hash.
4. Прикладные логи: Отражение внутренней логики ПО, где скрываются инъекционные атаки. Для их унификации мы предлагаем использовать онтологию OCSF, что делает модель инвариантной к конкретным платформам.

Математическое ядро системы оценки защищенности Центральным элементом нашего подхода является использование информационной энтропии Шеннона для поиска аномалий без использования сигнатур. Аномальное событие всегда оставляет «след» в распределении вероятностей типов событий:

- Снижение энтропии часто сигнализирует о концентрации однотипных действий, что характерно для DDoS-атак.
- Скачкообразный рост энтропии, напротив, указывает на хаотизацию — типичный признак работы программ-шифровальщиков.

Для работы в режиме реального времени мы интегрируем байесовский подход. Он позволяет обновлять апостериорную вероятность угрозы $P(\Theta|E)$ каждый раз, когда система видит новое событие. Чтобы избежать вычислительного коллапса, мы применяем наивный байесовский классификатор, что дает возможность инкрементного обновления оценок.

Кроме того, учитывая временные зависимости между событиями, мы задействуем аппарат скрытых марковских моделей (HMM). Здесь события в логах рассматриваются как наблюдаемые выходы скрытых состояний системы (например, «норма» или «вторжение»). Если вероятность наблюдаемой цепочки падает ниже порога θ, система сигнализирует о деградации уровня безопасности.

Корреляция данных и управление качеством в условиях неопределенности: Просто собрать данные недостаточно — их нужно коррелировать. В современных сетях использование только одного источника ведет к «информационной слепоте». Мы выделяем ключевые атрибуты качества корреляции, которые напрямую влияют на достоверность итоговой оценки: обработка пропущенных значений, борьба с ложными связями и устойчивость к выбросам.

В условиях высокой неопределенности, когда логи зашумлены или намеренно искажены, мы предлагаем методы нечеткой логики (fuzzy logic). Это позволяет оперировать не только жесткими бинарными категориями («атака/норма»), но и лингвистическими переменными («уровень угрозы — средний»), что делает систему более адаптивной к сложным целевым атакам.

Архитектура интеллектуального аналитического конвейера Предложенная нами многоуровневая структура объединяет все методы в единый поток обработки:

1. Слой предобработки: Нормализация форматов и NLP-парсинг неструктурированных записей.
2. Слой первичной детекции: Энтропийный мониторинг в динамическом скользящем окне.
3. Слой классификации: Использование байесовского вывода для атрибуции аномалий к техникам базы MITRE ATT&CK.
4. Слой свёртки: Вычисление интегрального показателя $\Sigma(t)$, который учитывает критичность узлов, интенсивность событий и достоверность источников.

В качестве технологического стека мы опираемся на ELK (Elasticsearch, Logstash, Kibana), что позволяет эффективно применять алгоритмы машинного обучения, такие как Isolation Forest и DBSCAN, для формирования «цифрового профиля» нормального поведения.

Критический анализ и перспективы развития: Мы признаем, что модель сталкивается с практическими вызовами. Главный из них — концептуальный дрейф (concept drift): профиль «нормального» поведения системы неизбежно меняется из-за обновлений ПО или смены штата, что вызывает ложные срабатывания. Аналитики тратят до 40% времени на расследование таких тревог.

Решение мы видим в добавлении семантического контекста и использовании графовых нейронных сетей для анализа причинно-следственных связей. Перспективным также является применение трансформеров (BERT/GPT) для глубокого разбора логов в гетерогенных средах.

Заключение: Разработанная методология превращает системный аудит из пассивного архива в активный инструмент измерения безопасности. Комбинация статистических и интеллектуальных методов позволяет достичь главного — сопоставимости оценок защищенности в динамике. Это создает фундамент для построения систем класса SIEM нового поколения, способных функционировать в условиях неопределенности и постоянной эволюции киберугроз.

СПИСОК ЛИТЕРАТУРЫ:

1. Mandiant. M-Trends 2024: Special Report.
2. IBM Security. Cost of a Data Breach Report 2023.
3. Gerhards R. The Syslog Protocol: RFC 5424.

4. Lunt T.F., Jagannathan R. A model of security monitoring.
5. Chuah E., et al. A systematic literature review of log-correlation tools.
6. Veeramachaneni K., et al. AI2: Training a Big Data Machine to Defend.
7. Sommer R., Paxson V. Outside the Closed World.
8. Козлов А.Д., Нога Н.Л. Применение методов нечеткой логики.
9. Kotenko I.V., et al. Система на основе средств Elastic Stack.
10. MITRE ATT&CK Framework v14.
11. Gollmann D. Computer Security.
12. Кузнецов А. А. и др. Теоретические основы ИБ.