

ИССЛЕДОВАНИЕ МЕТОДОВ КОРРЕЛЯЦИИ СОБЫТИЙ МОНИТОРИНГА И РАЗРАБОТКА СИСТЕМЫ ОБНАРУЖЕНИЯ ИНЦИДЕНТОВ В ГЕТЕРОГЕННОЙ ИТ-ИНФРАСТРУКТУРЕ

Т.В. Скворцова¹, А.К. Брайко¹, О.А. Авилова¹

ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Аннотация. В статье рассматриваются вопросы автоматизации процессов мониторинга в микросервисных архитектурах, где традиционные подходы часто оказываются недостаточно эффективными. Проанализированы ограничения пороговых методов обнаружения сбоев, которые приводят к избыточному количеству уведомлений. Описан новый подход к агрегации данных телеметрии, включающий метрики, логи и распределенные трассировки, для выявления составных инцидентов. Представлена математическая модель нормализации событий и алгоритм их корреляции, направленный на снижение уровня информационного шума и сокращение времени идентификации инцидента. Приведены результаты экспериментальной оценки на синтетическом наборе данных, подтверждающие работоспособность предложенного метода с показателем F1-меры на уровне 0,86.

Ключевые слова: мониторинг, микросервисная архитектура, событие, ИТ-инфраструктура.

STUDYING MONITORING EVENT CORRELATION METHODS AND DEVELOPING AN INCIDENT DETECTION SYSTEM IN A HETEROGENEOUS IT INFRASTRUCTURE

T.V. Skvortsova¹, A.K. Braiko¹, O.A. Avilova¹

¹Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. This article examines the automation of monitoring processes in microservice architectures, where traditional approaches often prove ineffective. The limitations of threshold-based failure detection methods, which lead to excessive notifications, are analyzed. A new approach to telemetry data aggregation, including metrics, logs, and distributed traces, is described for identifying complex incidents. A mathematical model for event normalization and an algorithm for their correlation

are presented, aimed at reducing information noise and reducing incident identification time. The results of an experimental evaluation on a synthetic dataset are presented, confirming the effectiveness of the proposed method with an F1 score of 0.86.

Keywords: monitoring, microservice architecture, event, IT infrastructure.

Цифровая трансформация и вызовы наблюдаемости

Цифровая трансформация и повсеместное внедрение облачных технологий (того, что называют Cloud Native) серьёзно поменяли ИТ-инфраструктуру компаний. Вместо монолитов всё чаще ставят распределённые микросервисные архитектуры, которые крутятся под управлением оркестраторов. И этот процесс неизбежно ведёт к одному – система становится сложнее, а данных, которые она генерирует, – экспоненциально больше. Системы мониторинга крупных предприятий каждый день перерабатывают терабайты телеметрии. Но тут есть проблема: чем больше данных собираешь, тем прозрачнее система не становится. Наоборот. Инженеры тонут в информационном шуме. В потоке событий – от загрузки CPU до строчек в логах приложений – трудно быстро понять, где следствие, а где коренная причина сбоя.

Почему это важно? Потому что нужно снижать операционные риски и держать ИТ-сервисы стабильными, даже когда вокруг много неопределённости. А инструменты мониторинга часто пилят тонны избыточных уведомлений. От этого персонал устаёт – есть даже термин «усталость от алертов» – и в какой-то момент можно пропустить действительно критический инцидент. Наша цель – разработать методику и программное средство для автоматической корреляции событий. Такое, чтобы точность диагностики стала выше.

Проблематика мониторинга гетерогенных сред

Подходы к мониторингу, основанные на статичных пороговых значениях, например, когда загрузка процессора превышает определенный процент, имеют существенные ограничения при работе в динамической среде. Во-первых, сама динамика среды предполагает автоматическое создание и удаление контейнеров, поэтому кратковременные изменения нагрузки могут являться штатным режимом работы при запуске сервиса, а не аномалией. Во-вторых, в распределенных системах часто возникают каскадные сбои, когда отказ одного компонента инфраструктуры вызывает цепную реакцию ошибок в зависимых сервисах. В таких

случаях оператор системы мониторинга получает множество уведомлений, сигнализирующих о локальных проблемах, но не указывающих на первоисточник неисправности.

Для решения указанной задачи требуется переход к комплексному анализу трех типов данных наблюдаемости. К ним относятся метрики, представляющие собой количественные показатели вроде задержек или пропускной способности, логи, содержащие текстовые записи событий об ошибках и исключениях, и трассировки, показывающие цепочки вызовов между сервисами. Объединение данных потоков в единую модель позволяет контекстуализировать события и более точно определять состояние системы.

Математическая модель и методы корреляции

Для алгоритмической обработки разнородных данных в работе используется унифицированная векторная модель события. Пусть E обозначает множество всех зарегистрированных событий. Каждое отдельное событие e_i описывается вектором, компоненты которого включают временную метку регистрации события t_i , идентификатор источника $source_i$, тип события $type_i$, значение события $value_i$, метаданные контекста $context_i$ и метку аномальности y_i . Значение события может быть числовым или текстовым, а метаданные связывают событие с конкретной транзакцией. Метка аномальности представляет собой бинарный признак, принимающий значение 0 или 1.

Инцидент I_j в данной модели определяется не как одиночное событие, а как кортеж, агрегирующий множество связанных событий. Он включает временной интервал T_j , подмножество событий E_j , список затронутых компонентов C_j , а также уровень критичности $severity_j$.

Метод корреляции базируется на предположении о пространственно-временной связи событий, относящихся к одному инциденту. Процесс включает несколько этапов. На этапе пространственной привязки алгоритм использует карту зависимостей сервисов для определения топологической близости источников. Учитывается связь между зависимыми сервисами при одновременном возникновении ошибок. На этапе временной группировки события фильтруются через скользящее временное окно, размер которого в базовых сценариях составляет от одной до пяти минут. На этапе контекстной связи используется идентификатор транзакции для объединения метрик и логов в единый кластер. Разметка трассировки $Y(\tau_k)$ определяется как максимальное значение аномальности среди всех

входящих в неё событий, что выражается формулой $Y(\tau_k) = \max\{y_i: e_i \in \tau_k\}$. Данный подход позволяет выявлять скрытые проблемы, при которых функциональность системы сохраняется, но происходит деградация качественных характеристик обслуживания.

Архитектура и реализация системы

Для проверки теоретической модели был разработан прототип системы мониторинга, архитектура которого построена на открытых технологиях. Сбор данных осуществляется компонентами Prometheus для метрик, Fluentd для логов и OpenTelemetry для трассировок. Ядро обработки реализовано на языке Python и выполняет нормализацию поступающих данных, их фильтрацию и корреляцию. Результаты сохраняются в базу данных Elasticsearch.

Для тестирования была развернута микросервисная архитектура в среде Docker Compose. С помощью генератора синтетической нагрузки эмулировались различные сценарии сбоев, включая утечку памяти, сетевые задержки, отказ базы данных и высокую нагрузку на шлюз. Был сформирован набор данных, содержащий размеченные инциденты для обучения и валидации алгоритма.

Результаты экспериментальной оценки

Мы оценивали эффективность по двум группам метрик – качество классификации и скорость работы. По точности обнаружения получили Precision = 0,88. Это значит, что большая часть уведомлений, которые выдала система, действительно указывала на реальную проблему. Показатель Recall = 0,85 – столько инцидентов из всех смоделированных мы успешно зафиксировали. Итоговая F1-мера (среднее гармоническое между точностью и полнотой) вышла 0,865.

Если смотреть на время, то средняя продолжительность между началом сбоя и формированием оповещения (MTTI) составила от двух до четырёх минут. Ложных срабатываний стало меньше – это хорошо, потому что операторы тратят меньше времени на разбор пустышек и начинают больше доверять системе мониторинга.

Заключение

В работе была рассмотрена задача автоматической корреляции событий в гетерогенных ИТ-системах. Предложенная векторная модель события позволяет эффективно объединять числовые и текстовые данные для последующего анализа. Использование идентификаторов трассировки обеспечивает надежную

связь между метриками и логами при анализе работы микросервисов. Разработанный алгоритм был протестирован на синтетических данных и продемонстрировал высокую эффективность с F1-мерой на уровне 0,86. Полученные результаты могут быть использованы при организации процессов эксплуатации, а перспективным направлением дальнейших исследований является внедрение методов машинного обучения для прогнозирования сбоев до момента их возникновения.

Список литературы

1. Москвичев А. Д., Долгачев М. В. Алгоритмы корреляции событий информационной безопасности // Автоматизация процессов управления. 2020. № 3. С. 50–59.
2. Гайфулина Д. А., Котенко И. В. Анализ моделей глубокого обучения для задач обнаружения сетевых аномалий интернета вещей // Информационно-управляющие системы. 2021. №. 1 (110). С. 28–37.
3. Общий подход к построению перспективных систем мониторинга распределенных информационно-телекоммуникационных сетей / Аллакин В.В., Будко Н.П., Васильев Н.В. // Системы управления, связи и безопасности. 2021. № 4. С. 125-227.
4. Системная инженерия в проблемах национальной безопасности / Анищенко А.В., Артемьев В.Б., Бондарева М.К., Будзко В.И., Бурмака А.А., Бычков И.В., Васильев В.И., Вульфин А.М., Гаденин М.М., Гарбук С.В., Гедзюн В.С., Головин С.А., Емельянов С.Г., Жуков А.О., Зацаринный А.А., Зорин А.В., Игнатьев А.Г., Капустин А.Я., Климов С.М., Козлов С.В. и др. Тематический блок «Национальная безопасность». Москва, 2025.

References

1. Moskvichev A. D., Dolgachev M. V. Information Security Event Correlation Algorithms // Automation of Control Processes. 2020. No. 3. pp. 50-59.
2. Gaifulina D. A., Kotenko I. V. Analysis of Deep Learning Models for Internet of Things Network Anomaly Detection Problems // Information Control Systems. 2021. No. 1 (110). pp. 28-37.
3. General Approach to Building Promising Systems for Monitoring Distributed Information and Telecommunication Networks / Allakin V. V., Budko N. P., Vasiliev N. V. // Control, Communication, and Security Systems. 2021. No. 4. pp. 125-227.

4. Systems Engineering in National Security Issues. Anishchenko A.V., Artemyev V.B., Bondareva M.K., Budzko V.I., Burmaka A.A., Bychkov I.V., Vasiliev V.I., Vulfin A.M., Gadenin M.M., Garbuk S.V., Gedzyn V.S., Golovin S.A., Emel'yanov S.G., Zhukov A.O., Zatsarinny A.A., Zorin A.V., Ignatiev A.G., Kapustin A.Ya., Klimov S.M., Kozlov S.V. and others. Thematic block "National Security". Moscow, 2025.