Проект РНФ № 21-71-20078

Аналитическая обработка больших массивов гетерогенных данных о событиях кибербезопасности в интересах оценки состояния, поддержки принятия решений и расследования компьютерных инцидентов в критически важных инфраструктурах

Описание выполненных в 2021 году работ и полученных научных результатов

1. Разработана концепция аналитической обработки больших массивов гетерогенных данных о событиях кибербезопасности в интересах оценки состояния, поддержки принятия решений и расследования компьютерных инцидентов в критически важных инфраструктурах (КВИ). Концепция описывает: принципы организации аналитической обработки больших массивов гетерогенных данных; цели и задачи перспективных систем мониторинга и управления безопасностью КВИ, а также предъявляемые к ним требования; особенности применения технологии суперкомпьютерных вычислений для анализа данных по кибербезопасности; механизмы обеспечения оценки состояния, поддержки принятия решений и расследования инцидентов кибербезопасности; описание КВИ категорий «Интернет вещей» и «киберфизическая система»; обобщенную архитектуру разрабатываемой системы. Сформулированы основные задачи, решаемые системой: оперативное обнаружение атак и нарушений политики безопасности; выявление инцидентов безопасности и их приоритезация; автоматическое реагирование на инциденты безопасности; ведение базы знаний по инцидентам безопасности; аудит и расследование инцидентов безопасности; оценка угроз для отдельных ресурсов КВИ. Их решение в среде суперкомпьютерных вычислений предполагает использование графических CUDA-ускорителей, библиотек TensorFlow, TyTorch и Theano, технологии MPI, акторной модели. Определены основные аналитические компоненты системы: обнаружения атак; обнаружения аномалий; оценки защищенности; анализа рисков; принятия решений; визуализации; расследования компьютерных инцидентов. Функционирование системы осуществляется в режимах обучения и эксплуатации.

2. Разработан общий подход и требования, предъявляемые к компонентам обнаружения в реальном времени атак на основе имитационного и графо-ориентированного моделирования. Проведенный анализ современного состояния исследований позволил выделить методы обнаружения, основанные на графах, байесовских сетях, Марковских моделях, сетях Петри, имитационном моделировании и технологиях больших данных для обнаружения многошаговых атак. Требования, предъявляемые к компонентам обнаружения, рассмотрены с точки зрения возможных атакующих воздействий и процесса их обнаружения, защищаемой КВИ, структуры и объема обрабатываемых данных, вариативности и масштабируемости подхода, а также тестирования. Обнаружение проводится в режиме, близком к режиму реального времени, с возможностью оперативного выявления известных видов инцидентов безопасности в отведенные временные рамки. Структура компонентов обнаружения атак предполагает возможность встраивания в процессе настройки частных модулей выявления атак. Все встроенные и включенные в работу модули обнаружения функционируют параллельно, в результате чего их функции могут быть распределены в рамках вычислительного кластера суперкомпьютера.

3. Разработан общий подход и требования, предъявляемые к компонентам обнаружения в реальном времени аномальной активности и нарушений критериев и политик безопасности на основе аналитической обработки больших массивов гетерогенных данных о событиях кибербезопасности. Проблема обнаружения в реальном времени аномальной активности и нарушений критериев и политик безопасности связана с решением задач, обусловленных большими объемами анализируемых данных и их размерностью, и высокой скоростью генерации новых потоков данных. Эти задачи порождают научно-практические вызовы, связанные с асинхронной генерацией данных, выявлением динамических связей между событиями, разнородностью используемых форматов и схем описания событий, а также смещением концепта в данных. Ключевые требования к компонентам обнаружения аномальной активности и нарушений критериев и политик безопасности связаны с обеспечением масштабируемости, оперативности, полноты и достоверности принимаемого решения, а также адаптивностью процедур обработки больших массивов данных о событиях безопасности. В структурной схеме компонентов можно выделить два уровня: планирования задач, связанных с обработкой событий, и обработки событий безопасности. Второй уровень представлен множеством программных модулей, обеспечивающих несколько режимов работы: модули обнаружения аномалий, модули верификации модели, а также модули, реализующие дообучение моделей, которые будут позволять не переобучать все обученные ранее модели машинного обучения, а расширять и уточнять состав типов нарушений критериев и политик безопасности и аномальной активности.

4. Разработан общий подход и требования, предъявляемые к компонентам оперативной оценки защищенности информационных, телекоммуникационных и других критически важных ресурсов на основе аналитической обработки больших массивов гетерогенных данных. Подход включает этапы разработки требований к оценке защищенности, спецификации оценки защищенности, разработки проекта оценки защищенности, проведения оценки защищенности и формирование заключения об оценке защищенности. В его основу положена разработка модельно-методического аппарата для оценки информационных, телекоммуникационных и других критически важных ресурсов на основе совместного применения графов атак и графов зависимостей сервисов. При этом в среде суперкомпьютерных вычислений могут использоваться технология MPI и распределенная память суперкомпьютера. Выделенные функциональные требования определяют необходимость комплексной оценки защищенности, поддержки принятия мер защиты, фиксации критериев, процедур, инструментов оперативной оценки защищенности и форм представления ее результатов, указания условий и границ применения компонента оценки защищенности. К квалиметрическим требованиям относятся: несмещенность, состоятельность, эффективность и достаточность оценки защищенности ресурсов КВИ.

5. Разработан общий подход и требования, предъявляемые к компонентам оперативного анализа и управления рисками информационной безопасности на основе аналитической обработки больших массивов гетерогенных данных о событиях кибербезопасности в интересах оценки состояния, поддержки принятия решений и расследования инцидентов. Цель этих компонентов - определить уровень рисков кибербезопасности, обеспечить баланс между стоимостью возможных негативных последствий и стоимостью мер защиты, и выработать рекомендации по обработке выявленных рисков. В условиях больших объемов данных компоненты анализа и управления рисками должны оперативно реагировать на изменяющуюся ситуацию, пересчитывать и сравнивать с критерием интегральные оценки. Достижение сформулированных требований к данным компонентам возможно при использовании

комплекса методов, в том числе формирования модели атак в виде графа в условиях обработки больших данных и с учетом неизвестных уязвимостей, обработки больших графов атак с целью вычисления метрик безопасности, обработки графов атак с циклами, формировании объективных интегральных метрик и объяснении их смысла на основе онтологического подхода, параллельной обработки больших данных, в частности при формировании и обработке графовых моделей и онтологий, реализованных для исполнения на высокопроизводительном кластере с поддержкой горизонтального масштабирования вычислительных ресурсов в рамках данного компонента.

6. Разработан общий подход и требования, предъявляемые к компонентам оперативной визуализации больших массивов гетерогенных данных о событиях кибербезопасности в интересах оценки состояния, поддержки принятия решений и расследования инцидентов. Были выявлены основные проблемы оперативной визуализации больших массивов гетерогенных данных, которые включают в себя выбор модели визуализации, обеспечение оперативности, восприятие данных большого объема, обработку гетерогенных данных и многомерность данных. Сформированный список требований к оперативной визуализации включает возможности агрегации данных путем агрегации объектов и измерений, поддержки преобразования гетерогенных данных к количественному или категориальному виду, а также определения структуры данных, обеспечения оперативности путем возможности горизонтального масштабирования вычислительных мощностей, включения моделей визуализации, которые способны отображать агрегированные объекты и человеко-машинного взаимодействия. Общий подход к визуализации представлен в виде стандартного конвейера визуализации, адаптированного для визуализации больших массивов гетерогенных данных, в виде схемы: анализ данных - агрегация - разметка - отрисовка. На основе данного общего подхода предложена архитектура компонента визуализации с поддержкой масштабирования модулей за счет методов параллельной обработки данных.

7. Разработан общий подход и требования, предъявляемые к компонентам принятия решений по защите информационных, телекоммуникационных и других критически важных ресурсов на основе аналитической обработки больших массивов гетерогенных данных о событиях кибербезопасности. Основное ограничение при принятии решений состоит в большом количестве вариантов решений и их характеристик, которые необходимо проанализировать. Поэтому для эффективного решения поставленной задачи требуется адаптация применяемого модельного аппарата под условия и ограничения современных систем параллельной обработки, в том числе с использованием суперкомпьютерных технологий и кластерных вычислений. В структурной схеме компонентов выделяются уровни планирования задач и принятия решений. Второй уровень представлен множеством программных модулей, обеспечивающих статический и динамический режим работы компонента. В том числе выделяются модули: обработки данных о событиях и данных безопасности из внешних источников; инвентаризации активов; формирования и обновления модели атак; формирования модели атакующего; формирования модели контрмер, интеграции с моделью атак; взаимодействия с компонентами оценивания и анализа защищенности, обнаружения атак и аномалий; принятия решений и вычисления показателей выбора мер реагирования; определения времени, доступного на принятие решения; взаимодействия с суперкомпьютерным центром; взаимодействия с компонентами визуализации.

8. Разработан общий подход и требования, предъявляемые к компонентам проведения расследований компьютерных инцидентов на основе аналитической обработки больших массивов гетерогенных данных о кибербезопасности. Общий подход к расследованию инцидентов включает следующие процедуры: индексация,

хеширование и подписание цифровой подписью входных данных для удобства поиска, гарантии целостности и подлинности; данные о сценариях атак и их контексте поступают от компонентов обнаружения в реальном времени атак; данные об аномалиях поступают от компонентов обнаружения в реальном времени аномальной активности и нарушений критериев и политик безопасности; приоритет обнаруженных атак и аномалий для расследования инцидентов определяется на основе данных о критичности элементов КВИ, предоставляемых компонентами оперативного анализа и управления рисками информационной безопасности; в качестве выходных данных предоставляются отчеты о хранимых инцидентах безопасности, включая их контекст, а также связанных с ними инцидентах. Кроме того, формируются данные для компонентов оперативной визуализации с целью их отображения для аналитики.

Результаты исследований были опубликованы в 12 статьях, индексируемых WoS и Scopus, и 14 статьях и тезисах докладов, индексируемых РИНЦ. Опубликована одна монография. При выполнении проекта было подготовлено 9 свидетельств о государственной регистрации программ для ЭВМ. Члены коллектива участвовали в апробации результатов на 12 российских и международных конференциях и семинарах.

Используя результаты проведенных исследований, были подготовлены лекции в рамках курса по повышению квалификации в сфере киберкриминалистики для сотрудников Следственного комитета из разных регионов России. Занятия направлены на повышение компетенций следователей по раскрытию преступлений в цифровой среде [https://tass.ru/obschestvo/13166641/amp]. Для имитации действий нарушителей и сбора данных о событиях безопасности был разработан прототип информационной системы и организован Хакатон (соревнование по этичному хакингу) для студентов вузов Санкт-Петербурга. В мероприятии приняло участие 29 команд, более 70 студентов [https://spark.ru/user/139694/blog/82190/uchyonie-spb-fits-ran-testiruyut-algoritmi-obnaruzheniya-atak].