ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
САНКТ-ПЕТЕРБУРГСКИЙ ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(СПИИРАН)
лаборатория проблем компьютерной безопасности
профильная лаборатория

ДОПУСТИТЬ К ПРЕДСТАВЛЕНИЮ ГЭК
Заведующий лабораторией

_____ _____И.В. Котенко_____
подпись *И.О.Фамилия*

«_____» _____ 2018 г.

Федорченко Андрей Владимирович

КОРРЕЛЯЦИЯ БОЛЬШИХ МАССИВОВ ГЕТЕРОГЕННЫХ ДАННЫХ ДЛЯ МОНИТОРИНГА И УПРАВЛЕНИЯ БЕЗОПАСНОСТЬЮ В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ

НАУЧНЫЙ ДОКЛАД

об основных результатах подготовленной научно – квалификационной работы (диссертации)
по основной образовательной программе подготовки научно-педагогических кадров в аспирантуре

направление подготовки 09.06.01 «Информатика и вычислительная техника»
направленность 05.13.01 «Системный анализ, управление и обработка информации»

Научный руководитель
д-р техн. наук, профессор

_____ _____И.В. Котенко_____
подпись *И.О.Фамилия*

«_____»_____ 2018 г.

Автор работы
аспирант

_____ _____А.В. Федорченко____
подпись *И.О.Фамилия*

Санкт-Петербург – 2018

Работа выполнена в федеральном бюджетном учреждении науки «Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)».

Научный руководитель: доктор технических наук,
профессор
Котенко Игорь Витальевич,
СПИИРАН,
заведующий лабораторией
проблем компьютерной безопасности.

Рецензенты: доктор технических наук,
профессор
Саенко Игорь Борисович,
Федеральное бюджетное учреждение науки «Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)»,
Ведущий научный сотрудник лаборатории проблем компьютерной безопасности.

доктор технических наук,
профессор
Молдовян Александр Андреевич,
Федеральное бюджетное учреждение науки «Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)»,
Заместитель директора по информационной безопасности, научный руководитель отдела проблем информационной безопасности.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы научно-квалификационной работы. Корреляция данных для обеспечения безопасности в киберфизических системах (КФС) является важной и актуальной задачей, поскольку их инфраструктуры отличаются высокой гетерогенностью и объемами информации, а существующие методы не в состоянии эффективно выполнять задачи корреляции в них. К системам данного класса относятся сети «Интернета вещей» (Internet of Things, IoT), автоматизированные системы управления технологическим процессом (АСУ ТП), бортовые компьютеры транспорта с функцией корректировки курса и многие другие.

Основная проблема выполнения корреляции данных в КФС обусловлена структурной, конфигурационной и функциональной неопределенностью их инфраструктуры. Эффективность выполнения задачи корреляции значительно снижается при высокой гетерогенности источников информации, несогласованности их форматов представления данных, условно-неограниченном количестве, а также возрастающей сложности и скрытости потенциальных атак. Одним из классов средств защиты информации в подобных условиях являются системы управления информацией и событиями безопасности (Security Information and Event Management, SIEM) следующего поколения. Предполагается, что для выполнения задач корреляции в неопределенных инфраструктурах необходим этап предварительного анализа данных, в результате которого производится автоматизированная адаптация модуля корреляции к целевой инфраструктуре, а именно, определение ее структуры, типов активов и их иерархии. Дальнейший анализ направлен на обучение моделей поведения различных объектов целевой инфраструктуры (ЦИ) с целью проактивного мониторинга, позволяющего заранее вычислить вероятность наступления конкретного инцидента.

Различные методы корреляции данных исследуются научным сообществом на протяжении более 20 лет. Однако существующие методики, как правило, требуют тонкой ручной либо автоматизированной настройки под конкретную задачу, а на практике преимущественно применяются методы на основе правил. Таким образом, разработка адаптивного подхода корреляции данных в условиях неопределенной инфраструктуры является актуальной темой исследований.

Степень разработанности темы научно-квалификационной работы. Вопросам корреляции гетерогенных данных для различных прикладных задач посвящено большое количество работ как отечественных исследователей: О.Ю. Воробьев, С.В. Клочков, И.В. Котенко, И.Б. Саенко, так и зарубежных: C. Kruegel, F. Valeur, G. Vigna, R. Sadoddin, A. Ghorbani, A. Muller, T. Limmer, F. Dressler. Анализ предметной области показал, что предлагаемые методики

преимущественно не обладают возможностью комплексной адаптации к ЦИ, а существующие подходы адаптации направлены на частные случаи неопределенностей (неопределенность поведения, неопределенность состояния и др.). В связи с данным фактом в текущем исследовании была поставлена задача разработки комплексного подхода к корреляции данных для выполнения задач безопасности в SIEM-системах на основе адаптации к неопределенной инфраструктуре КФС.

Научная задача. Разработка модельно-методического аппарата для выполнения процесса корреляции данных безопасности в КФС на основе адаптации к их инфраструктурам в условиях неопределенности.

Объект исследования. КФС, а также их информация и события безопасности.

Предмет исследования. Модели, методики и алгоритмы корреляции условно-неограниченного количества разнородных событий и информации безопасности в КФС.

Основной целью научно-квалификационной работы является повышение эффективности выполнения процесса корреляции событий и информации безопасности в КФС за счет применения комбинированной методики корреляции, ориентированной на обработку исходных данных от условно-неограниченного количества гетерогенных источников и адаптацию к неопределенной ЦИ. Для достижения указанной цели в работе поставлены и решены следующие **задачи**:

1. Анализ информации и событий безопасности, а также методов и подходов к их корреляции для формального описания исходных данных и формирования модели неопределенной ЦИ КФС.
2. Разработка модели корреляции данных безопасности с условно-статичным содержимым на основе онтологического подхода.
3. Разработка модели и комплексной методики корреляции данных с динамичным содержимым (событий) с возможностью автоматизированной адаптации к неопределенной ЦИ КФС.
4. Экспериментальная оценка предложенных алгоритмов и методик по показателям оперативности, масштабируемости и ресурсопотребления.

Положения, выносимые на защиту:

1. Онтологическая модель корреляции данных безопасности с условно-статичным содержимым.
2. Модель корреляции данных безопасности с динамичным содержимым.
3. Комплексная методика корреляции информации и событий безопасности с адаптацией к неопределенной ЦИ КФС.

Научная новизна диссертационной работы состоит в следующем:

1. Предложенная онтологическая модель корреляции данных безопасности с условно-статичным содержимым ориентирована на использование взаимосвязей как между разнотипной информацией: уязвимостями, эксплойтами, слабостями, шаблонами атак и др., так и между различными источниками однотипной информации, например, между записями об уязвимостях в различных базах. Основным аспектом установления факта наличия связи между разнотипными данными безопасности является реализация той или иной сущности (эксплойт реализует уязвимость, уязвимость реализует слабость и т.д.). В свою очередь связь между экземплярами однотипной информации безопасности из различных источников строится за счет пересекающихся ссылок на ресурсы их описания.

2. Модель корреляции данных с динамичным содержимым (событий безопасности) опирается на извлечение характеристик объектов инфраструктуры из свойств событий. Множество различных комбинаций характеристик и их типов, а также связей между свойствами по различным показателям эквивалентности формируют конкретную модель неопределенной ЦИ КФС.

3. Разработанная комплексная методика корреляции информации и событий безопасности отличается адаптацией к неопределенной ЦИ КФС за счет выявления различных типов объектов и их иерархии на основе структурного анализа журналов событий. Дальнейшая обработка данных опирается на исследование поведения экземпляров объектов конкретных типов, а также связи информации с условно-статичным и динамичным содержимым.

Обоснованность и достоверность представленных научных положений обеспечивается тщательным анализом состояния исследований в предметной области, подтверждается согласованностью результатов, полученных при экспериментах, успешной апробацией на ряде научных конференций всероссийского и международного уровня, и публикацией в ведущих рецензируемых научных изданиях.

Теоретическая и практическая значимость результатов исследования. Разработанные модели и комплексная методика корреляции данных безопасности КФС развивают теоретические положения в данной области и позволят повысить эффективность предварительной оценки вероятности наступления инцидента за счет адаптивного определения ЦИ и идентификации типов активов, а также анализа поведения отдельных объектов. Предложенная методика должна стать основой модуля корреляции в SIEM-системах следующего поколения. Применимость результатов исследований определяется необходимостью проактивного мониторинга подобными системами состояния безопасности КФС для своевременного предотвращения потенциальных инцидентов и (или) минимизации ущерба в случае

их неизбежности. С учетом критичности различных КФС, а также их разнообразия, предлагаемый подход корреляции данных имеет обширную область применения.

Реализация результатов работы. Отраженные в научно-квалификационной работе исследования проведены в рамках следующих научно-исследовательских работ: гранта РФФИ № 16-37-00338-мол_а, гранта РНФ № 15-11-30029, гранта Президента Российской Федерации № МК-314.2017.9, проектов Минобрнауки России № 14.604.21.0137, № 14.604.21.0033, № 14.616.21.0028 и № 14.604.21.0147, проекта 2017-2018 гг. НИР-ФУНД Университета ИТМО № 717075 «Методы, модели, методики, алгоритмы, протоколы и приложения для обеспечения информационной безопасности киберфизических систем» и др.

Апробация результатов работы. Основные положения и результаты работы докладывались на научных конференциях: 26-я международная конференция по параллельной, распределенной и сетевой обработке PDP-2018 (Кембридж, Великобритания, 2018); 9-я международная конференция по интеллектуальному сбору данных и передовым вычислительным системам IDAACS-2017 (Бухарест, Румыния, 2017); 23-я международная конференция по параллельной, распределенной и сетевой обработке PDP-2015 (Турку, Финляндия, 2015); 11-й международный симпозиум по интеллектуальным распределенным вычислениям IDC-2017 (Белград, Сербия, 2017); XX международная конференция по мягким вычислениям и измерениям SCM-2017 (Санкт-Петербург, 2017); 23-я и 24-я общероссийская научно-техническая конференция «Методы и технические средства обеспечения безопасности информации» (Санкт-Петербург, 2014-2015гг); IX и X Санкт-Петербургская межрегиональная конференция «Информационная безопасность регионов России (ИБРР-2015,2017)» (Санкт-Петербург, 2015,2017гг.); часть 7-й и 9-й Российской мультиконференции по проблемам управления – конференция «Информационные технологии в управлении (ИТУ-2014,2016)» (Санкт-Петербург, 2014,2016гг.); XIV и XV Санкт-Петербургская международная конференция «Региональная информатика (РИ-2014,2016)» (Санкт-Петербург, 2014, 2016гг.); VI международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО-2017)» (Санкт-Петербург, 2017); всероссийский форум «Система распределенных ситуационных центров как основа цифровой трансформации государственного управления (СРСЦ-2017)» (Санкт-Петербург, 2017), 4-я и 6-я международная конференция по практической безопасности «Positive Hack Days (PHD-4,6)» (Москва, 2015, 2017гг.).

Публикации. По материалам диссертационной работы опубликовано более 30 работ, в том числе 8 – в рецензируемых изданиях из перечня ВАК («Информационно-управляющие системы», «Безопасность информационных технологий», «Проблемы

информационной безопасности. Компьютерные системы», «Известия высших учебных заведений. Приборостроение», «Труды СПИИРАН», «Вопросы кибербезопасности»), 8 – в изданиях, индексируемых в международных базах Scopus и Web Of Science, и 7 свидетельств о государственной регистрации программ для ЭВМ.

Структура и объем научно-квалификационной работы. Данная работа включает введение, три главы, заключение, список литературы (74 наименования). Объем работы – 107 страниц машинописного текста; включает 21 рисунок и 7 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность и важность темы научно-квалификационной работы, определена цель и сформулированы задачи исследования, выполнение которых необходимо для ее достижения. Показаны научная новизна и практическая значимость работы. Дано краткое описание предложенного подхода корреляции данных безопасности в условиях неопределенности ЦИ КФС, а также представлены основные результаты разработки предлагаемых моделей и комплексной методики в научно-исследовательских проектах.

В первой главе детально описываются сферы применения КФС как сложных систем управления, а также подробно рассматриваются проблемы обеспечения безопасности систем данного класса. Рассматривается место и роль процесса корреляции в SIEM-системах. Приводится общая схема следования потоков исходных данных для процесса корреляции, классы и виды информации безопасности. Указываются основные отличия между данными с условно-статичным и динамичным содержимым, а также причины подобного разделения. Подробно рассматриваются этапы процесса корреляции, а также отдельные методы корреляции и их общая классификация. Выдвигаются требования к комплексной методике корреляции данных безопасности по аспектам оперативности, масштабируемости и ресурсопотребления, а также основных задач, возлагаемых на процесс корреляции информации безопасности. Формулируются требования к исходным данным, отображающим ограничения предлагаемой методики корреляции. В завершении главы приводится формальная постановка задачи научно-квалификационной работы.

Во второй главе рассматриваются теоретические положения, описывающие модель неопределенной ЦИ, на базе которой производится разработка моделей корреляции данных с условно-статичным и динамичным содержимым. Основу анализируемой ЦИ составляют множества информационных объектов, а также их типов и взаимодействий между ними. Определение множества типов событий направлено на преодоление структурной и конфигурационной неопределенностей ЦИ

КФС, тогда как определение множества взаимодействий между типами информационных объектов ликвидирует функциональную неопределенность.

Описанная модель корреляции данных с условно-статичным содержимым разработана с помощью онтологического подхода, поскольку автоматизированное выявление семантических связей между подобной полуструктурированной информацией практически не выполнима. В свою очередь, модель корреляции данных с динамичным содержимым включает множество журналов безопасности, их типов событий и типов свойств событий. Такое представление данных позволило не только определять элементы модели ЦИ в рамках этапа адаптации, но и связать два класса информации (условно-статичной и динамичной) за счет комплексной методики корреляции событий и информации безопасности. В основе методики лежит структурный анализ журналов событий безопасности, позволяющий идентифицировать типы активов ЦИ КФС, их взаимосвязи, а также выделять отдельные активы для проведения поведенческого анализа каждого из них. Условно-статичная информация в данном случае выполняет роль частных характеристик информационных объектов ЦИ КФС.

В третьей главе подробно описывается опытный стенд и наборы данных, за счет которых производятся эксперименты по соответствию полученных результатов исследований предъявляемым требованиям. Рассматриваются показательные случаи результатов адаптации комплексной методики корреляции данных безопасности к неопределенной ЦИ КФС, а также их причины и варианты преодоления коллизий различного рода. Оценивается возможность применения предлагаемого подхода корреляции к ЦИ условно-неограниченного размера. Полученные результаты подтверждают выполнение поставленной научной задачи, их теоретическую и практическую значимость, а также наличие инновационной составляющей подхода по отношению к существующим решениям.

ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ

Комплекс предложенных в научно-квалификационной работе моделей и методики корреляции данных в условиях неопределенности, и их практическая реализация, представляют собой решение актуальной задачи разработки модельно-методического аппарата для мониторинга и управления безопасностью в КФС. Их применение вносит значительный вклад в развитие систем управления информацией и событиями безопасности. При решении данной задачи были получены следующие результаты:

1. Разработана модель корреляции данных с условно-статичным содержимым на основе применения онтологического подхода связывания разнотипной информации безопасности. Подробный анализ источников описания различных типов данных

позволил определить взаимоотношения как между разными типами данных, так и между однотипными данными из различных несогласованных друг с другом источников.

2. Предложена модель корреляции данных безопасности с динамичным содержимым для осуществления этапа адаптации процесса корреляции к условиям неопределенности ЦИ КФС. Выделение основных элементов ЦИ производится на основании ряда гипотез, справедливость которых подтверждается соответствующими экспериментами.

3. Разработана комплексная методика корреляции данных на основе структурного анализа множества журналов событий, позволяющая выполнять автоматизированную идентификацию неопределенной ЦИ КФС и последующий анализ поведения ее объектов.

Результаты соответствуют п. 8 Паспорта специальностей ВАК (технические науки) «Теоретико-множественный и теоретико-информационный анализ сложных систем.» по специальности 05.13.01. Дальнейшим направлением исследований может являться расширение области применения предлагаемого подхода корреляции к другим видам систем, например, социо-кибер-физическим, а также к прикладным задачам, не связанным с безопасностью.

СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ НАУЧНОГО ИССЛЕДОВАНИЯ

Публикации в рецензируемых изданиях из списка ВАК:

Федорченко, А. В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 1 / А.В. Федоченко, Д.С. Левшун, А.А. Чечулин, И.В. Котенко // Труды СПИИРАН. - 2016. - Вып. 4 (47). - С. 5-27.

Федорченко, А. В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 2 / А.В. Федоченко, Д.С. Левшун, А.А. Чечулин, И.В. Котенко // Труды СПИИРАН. - 2016. - Вып. 6 (49). - С. 208-225.

Федорченко, А. В. Исследование открытых баз уязвимостей и оценка возможности их применения в системах анализа защищенности компьютерных сетей / А.В. Федорченко, А.А. Чечулин, И.В. Котенко // Информационно-управляющие системы. – 2014. - № 5. - С. 72-79.

Федорченко, А. В. Аналитический обзор открытых баз уязвимостей программно-аппаратного обеспечения / А.В. Федорченко, А.А. Чечулин, И.В. Котенко // Проблемы информационной безопасности. Компьютерные системы. - 2014. – № 3. - С. 131-135.

Федорченко, А. В. Построение интегрированной базы уязвимостей / А.В. Федорченко, А.А. Чечулин, И.В. Котенко // Известия высших учебных заведений. Приборостроение. – 2014. - Т. 57, № 11. - С. 62-67.

Федорченко, А. В. Разработка сервиса доступа и управления интегрированной базой уязвимостей / А.В. Федорченко, А.А. Чечулин, И.В. Котенко // Безопасность информационных технологий. – 2015. – № 4. - С. 26-32.

Котенко, И. В. Технологии больших данных для корреляции событий безопасности на основе учета типов связей / И.В. Котенко, А.В. Федорченко, И.Б. Саенко, А.Г. Кушнеревич // Вопросы кибербезопасности. – 2017. № 5(23). - С. 2-16.

Федорченко, А. В. Корреляция информации в SIEM-системах на основе графа связей типов событий / А.В. Федорченко, И.В. Котенко // Информационно-управляющие системы. - 2018. № 1. - С. 58-67.

Публикации из баз данных Web Of Science и Scopus:

Fedorchenko, A. Design of integrated vulnerabilities database for computer networks security analysis / A.Fedorchenko, I. Kotenko, A. Chechulin // Proceedings of the 23th Euromicro International Conference on “Parallel, Distributed, and Network-Based Processing”. - 2015. - P. 559-566.

Fedorchenko, A. Integrated repository of security information for network security evaluation / A.Fedorchenko, I. Kotenko, A. Chechulin // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. – 2015. - Vol. 6, No. 2. - P. 41-57.

Fedorchenko, A. The ontological approach application for construction of the hybrid security repository / A. Fedorchenko, I. Kotenko, E. Doynikova, A. Chechulin // Proceedings of the XX International Conference on “Soft Computing and Measurements”. – 2017. - P. 525-528.

Kotenko, I. Ontological hybrid storage for security data / I. Kotenko, A. Chechulin, E. Doynikova, A. Fedorchenko // Proceedings of the 11th International Symposium on “Intelligent Distributed Computing”. - 2017. - P. 159-171.

Fedorchenko, A. Correlation of security events based on the analysis of structures of event types / A. Fedorchenko, I. Kotenko, D. El Baz // Proceedings of the 9th IEEE International Conference on “Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications”. – 2017. - P. 270-276.

Kotenko, I. Parallelization of security event correlation based on accounting of event type links / I. Kotenko, A. Fedorchenko, I. Saenko, A. Kushnerevich // Proceedings if the 26th Euromicro International Conference on “Parallel, Distributed, and Network-Based Processing”. - 2018. - P. 462-469.

Свидетельства о государственной регистрации программ для ЭВМ:

Федорченко, А. В. Сервисы доступа и управления интегрированной базой уязвимостей для систем мониторинга и управления безопасностью информационно-телекоммуникационных систем. Свидетельство № 2015615366 от 15.05.2015 / А. В. Федорченко, И. В. Котенко.

Федорченко, А. В. Компонент анализа статистики и оценки качественных параметров интегрированной базы уязвимостей. Свидетельство № 2015662208 от 18.11.2015 / А. В. Федорченко, А. А. Чечулин, И. В. Котенко.

Федорченко, А. В. Интегрированная база уязвимостей для систем мониторинга и управления безопасностью информационно-телекоммуникационных систем. Свидетельство № 2015621655 от 17.11.2015 / А. В. Федорченко, А. А. Чечулин.

Федорченко, А. В. Компонент экспертной оценки качества визуализации неформализованных данных разнородной структуры. Свидетельство № 2016663861 от 19.12.2016 / А. В. Федорченко, А. А. Чечулин.

Федорченко, А. В. Агент сбора событий безопасности ОС Windows с функцией выборочной анонимизации передаваемой информации. Свидетельство № 2017619728 от 01.09.2017 / А. В. Федорченко, И. В. Котенко, И. Б. Саенко.

Федорченко, А. В. Компонент анализа полуструктурированных баз данных для построения гибридного хранилища информации безопасности. Свидетельство № 2017663404 от 01.12.2017 / А. В. Федорченко, А. А. Чечулин, Е. В. Дойникова.

Дойникова, Е. В. Компонент нормализации данных из внешних источников для построения гибридного хранилища информации безопасности. Свидетельство № 2017663405 от 01.12.2017 / Е. В. Дойникова, А. А. Чечулин, А. В. Федорченко.