ОТЗЫВ

официального оппонента на диссертационную работу Смирнова Д.В. «Методы поиска инсайдера в Big Data», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.19 — «Методы и системы защиты информации, информационная безопасность».

Актуальность темы диссертационной работы.

Развитие средств и систем автоматизации в различных областях человеческой деятельности, характерное для текущего момента развития общества и государства, создает новые вызовы. Одна из сложнейших проблем современности состоит в поиске разумного компромисса между простотой выполнения операций, связанных с персональными данными, и требуемым уровнем защищенности этих данных. Регулярно приходит информация об организованных атаках на информационные ресурсы, государств и частных компаний, приводящих к существенному финансовому или репутационному ущербу. Поэтому критически важные для деятельности государства и крупного бизнеса информационные системы имеют многоуровневую систему защиты от внешних угроз.

Помимо защиты внешнего периметра информационных систем, специалисты выделяют большой класс задач защиты информации от внутреннего нарушителя. Работа автоматизированной системы невозможна без наличия подчас весьма многочисленного обслуживающего персонала. Эти люди объективно имеют доступ к ресурсам системы и, пользуясь предоставленными им в силу производственной необходимости правами, могут использовать их для доступа к защищаемой информации, например, личной и финансовой информации клиентов сервисной государственной структуры или крупного банка. Криминальная деятельность внутренних злоумышленников или инсайдеров представляет собой существенную угрозу для больших информационных систем.

Для распределенных систем, обрабатывающих миллионы транзакций в сутки, проблема своевременного выявления вредоносной активности инсайдера является не только практически важной, но и достаточно трудной, поскольку необходимо выявить ничтожную часть подозрительных транзакций в гигантском потоке операций, выполняемых параллельно на большом количестве территориально распределенных серверов и баз данных.

Диссертационная работа Смирнова Д. В. посвящена разработке методов поиска инсайдеров в системах распределенной обработки данных, характеризующихся несколькими тысячами серверов, петабайтными объемами данных и числом пользователей в несколько тысяч. Количество записей о регистрации событий (логи) превышает 10^9 единиц в сутки. Данное направление

исследований является актуальным и востребованным для практических приложений.

В результате проведенного диссертационного исследования были получены *новые научные результаты*, наиболее существенными из которых следует признать:

1. Разработка оригинальной методики аналитической обработки больших объемов данных логов с целью выявления признаков вредоносной активности инсайдеров, включая распределенные операции, выполняемые участниками сговора лиц.
2. Теоретические модели, основанные на методах статистического анализа данных, интеллектуального анализа данных и машинного обучения, ориентированные на выявление аномалий в больших наборах данных с сильной несбалансированностью классов, идентифицирующих признаки вредоносной активности инсайдеров.
3. Алгоритмы, реализующие процедуры выявления аномалий в регистрируемых данных о событиях в системе и представления их в форме, ориентированной на использование рекомендаций работниками профильных служб безопасности. Программные реализации разработанных алгоритмов.

Достоверность полученных результатов и возможность их использования на практике подтверждается теоретическими обоснованиями и результатами внедрения в форме промышленного решения – программного комплекса, внедренного в крупной коммерческой организации.

Основные результаты диссертации *опубликованы* в 7 печатных работах, 6 научных статей в изданиях, рекомендованных ВАК России для публикаций результатов диссертационных работ, 4 статьи – в изданиях, индексированных в базах SCOPUS.

Опубликованные работы отражают основные научные и практические результаты диссертации.

Результаты диссертационной работы обсуждались на совместном семинаре 53 и 16. отделов ФИЦ ИУ РАН и международных конференциях и семинаре в подразделении, реализовавшем промышленную систему поиска признаков инсайдера в крупном коммерческом банке.

Практическая ценность диссертационной работы состоит в том, что ее результаты могут быть использованы при проектировании и модернизации подсистем защиты информации, как в проектируемых информационных системах промышленного уровня, так и в информационных системах, находящихся в эксплуатации.

По результатам работы получены:

- Свидетельство о государственной регистрации программы для ЭВМ № 2021614494 «Аналитическая панель доступов к данным» (дата государственной регистрации 25.03.2021);

– Свидетельство о государственной регистрации программы для ЭВМ № 2021613506 «Поисковая система доступа к данным» (дата государственной регистрации 19.04.2021).

Общая характеристика работы.

Во *введении* формально отражено выполнение требований, предъявляемых к содержанию диссертаций: представлены аргументы, обосновывающие актуальность темы диссертации, сформулированы цель и задачи работы, кратко сформулирована научная новизна и практическая ценность задач, решаемых в диссертации. Представлены основные результаты работы.

В *первой главе* приведен обзор методов решения сходных задач, представленных в научной литературе, включая краткий обзор методов социальной инженерии. Отмечено, что основными методами, используемыми для решения сходных задач являются:

– анализ локальных уровней выброса (LoF),
– метод опорных векторов,
– байесовский индуктивный вывод,
– методы кластеризации методом k-средних,
– случайный лес решений и др.

Показано, что специфика используемых методов состоит в предположении об однородности данных, что исключает учет дополнительных факторов из других пространств признаков, таких как характеристики личности сотрудников, попадающих в круг подозреваемых и другие важные факторы. Делается вывод о том, что анализ инсайдерской активности нужно проводить по всему спектру сведений: записей событий из эксплуатируемых технических систем и систем контроля (логов), кадровой информации о персональных характеристиках сотрудников, информации служб собственной безопасности, сведений из баз данных правоохранительных органов и других государственных структур и т.д.

Во *второй главе* рассмотрены методы выявлении аномалий в поведении сотрудников, а также методы, позволяющие интегрировать гетерогенные данными с целью выявления компрометирующих сведений, объединяя данные, выявленные в различных информационных пространствах, в единый результат.

Характерной особенностью предлагаемого метода является постановка задачи анализа информации в каждом признаковом пространстве как оценка характеристик соответствующего вероятностного процесса. Рассматриваемый подход основан на запретах и полузапретах вероятностных мер в различных информационных пространствах. С последовательностями событий, зафиксированных в таких информационных пространствах, связываются булевские переменные. Производные признаки, описывающие ситуации, подозрительные на криминальную активность инсайдеров, представляются как

булевские формулы, построенные по наблюдаемым признакам во всем множестве анализируемых информационных пространств.

В *третьей главе* представлены методы, позволяющие сегментировать анализируемые гетерогенные данные на однородные фрагменты (сегменты), где возможно применение традиционных методов математический статистики и вероятностных методов выявления аномалий.

Автором представлен оригинальный подход к проблеме выявления организованной группы нарушителей информационной безопасности. Сущность предлагаемого решения состоит кластеризации исходного множества данных малых выборок, описывающих функционал информационных технологий организации. Представлен алгоритм выявления множества клиентов, которые представляют интерес для инсайдеров, использующих сговор. Получены интересные асимптотические результаты.

Также отмечаю как заслуживающую внимания представленную в третьей главе идею оценки правильности (правдоподобности) действий потенциальных инсайдеров на основе оценок вероятностей случайного появления найденных закономерностей в некоторых вероятностных моделях. Показано, при каких соотношениях параметров возможно эффективное выявление корреляционных связей между событиями, указывающих на признаки вредоносной активности инсайдеров.

Предложены два варианта управления параметрами системы обнаружения, которые позволяют получать содержательную информацию о подозрительных действиях потенциальных инсайдеров.

Первый вариант предполагает разделение периода наблюдений на промежутки, в течение которых корреляция характеристик, указывающая на подозрительные действиях потенциальных инсайдеров, может проявиться в наибольшей степени.

Второй вариант предполагает сокращение множества пользователей, которые потенциально проявили признаки инсайдерской активности.

В *четвертой главе* представлена разработанная автором методика оперативного анализа большого потока данных, содержащих признаки криминальных действий инсайдеров.

Выделен перечень основных функций программно-технических инструментов. Определены как критически значимые характеристики: скорость индексации первичных данных, и их переиндексации, поддерживаемые API, взаимосвязи текущих размеров базы и скорости поиска, поддерживаемые типы входных документов и др.

В качестве практической реализации предлагаемого метода в задаче идентификации признаков вредоносной инсайдерской активности определены параметры целевой ИТ-среды анализа данных и поддержки принятия решений для реального объекта защиты в крупном российском коммерческом банке.

Разработанные методика и программный инструментарий выявления признаков инсайдера реализованы в специализированном комплексе. Инструментальный комплекс порождает несколько десятков терабайт «сырых» данных в сутки.

Отличительной особенностью комплекса является наличие первичного и вторичный поиска.

Первичный поиск характеризуется классом задач, где аналитик выявляет в исходных «сырых» данных те данные, которые отвечают знаниям, накопленным в соответствии с текущим профилем угроз. Первичный поиск аккумулирует опыт экспертов служб безопасности о признаках наблюдавшихся ранее инсайдерских активностей и определяет значения характеристик поиска в «сырых» данных (поля, идентификаторы и т.п.), необходимых для принятия решения.

Вторичный поиск обеспечивает ответы на релевантные целям анализа безопасности запросы уполномоченных сотрудников. В частности, обеспечивается детальный профиль соответствующего сотрудника или подразделения, в том числе, штатный профиль доступов данного сотрудника к ресурсам хранилища Big Data. Автор реализовал такие программно-технические решения для имеющихся ограничений, что время отклика на запрос не превышало 10 сек. на выделенном для этого программно-техническом комплексе.

На основании анализа материала диссертационной работы можно сформулировать следующие недостатки.

1. Раздел 2.2.1. «Общая математическая модель множества информационных пространств» представляется излишним, поскольку вводит ряд понятий высокой степени абстракции, которые не используются в дальнейшем. Автор в дальнейшем (раздел 2.2.2) по сути переформулирует задачу в терминах, которые можно рассматривать как общую математическую модель.
2. Примеры 12-13 (стр.88-91) представляются излишними, поскольку не относятся к теме представленного диссертационного исследования, хотя и расширяют кругозор специалиста по защите информации.
3. Раздел 3.2 «О вероятностных оценках достоверности эмпирических выводов» целесообразно вынести в первую главу, поскольку он содержит данные о результатах исследований по одной из затронутых в диссертации проблем.
4. Раздел 3.2.2 «Модель машинного обучения» по сути несколько изменяет исходную модель. Машинное обучение может быть выполнено и на исходной модели. Изменение исходной модели не представляется оправданным.

Отмеченные недостатки не снижают общей положительной оценки диссертационной работы.

Диссертация и автореферат написаны ясным и достаточно строгим языком, стилем, принятым в научной литературе. Личный вклад автора в совместных работах может быть оценен как существенный и относящийся к сути диссертационного исследования.

Диссертационная работа Смирнова Д.В «Методы поиска инсайдера в Big Data» является законченным самостоятельным исследованием, содержит оригинальные модели и алгоритмы решения новых задач. Представленные в диссертации результаты являются значимыми для науки и инженерной практики.

Диссертационная работа соответствует требованиям, предъявляемым ВАК Российской Федерации к материалам кандидатских диссертаций, а ее автор – Смирнов Дмитрий Владимирович – заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.19 «Методы и системы защиты информации, информационная безопасность».

Официальный оппонент:
доктор технических наук, профессор С.Н. Смирнов

С.Н. Смирнов защитил диссертацию на соискание доктора технических наук по специальности 05.13.19 – “Методы и системы защиты информации; информационная безопасность”. Диплом ДК № 018922 (решение Президиума ВАК России от 25 июля 2003 года № 31д/4).

Приказом Министерства образования и науки Российской Федерации от 29 апреля 2013 №211/нк-1 Смирнову С. Н. присвоено ученое звание профессора по специальной кафедре; аттестат: серия ПР № 043654.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет)», профессор кафедры ИУ-10 (Защита информации), 105005, г. Москва, ул., 2-я Бауманская, д. 5, стр. 1, +7(499)-263-64-07, smirnovsn@bmstu.ru

«7 » декабря 2021 г.

ВЕРНО
ЗАМ. НАЧАЛЬНИКА УПРАВЛЕНИЯ КАДРОВ
МГТУ ИМ. Н.Э. БАУМАНА
А.Г. МАТВЕЕВ