ОТЗЫВ официального оппонента

на диссертацию на соискание ученой степени кандидата физико-математических наук Терёхиной Ирины Юрьевны

на тему: «Методы выявления аномалий в условиях смеси технологических процессов, сопровождающих наблюдаемый объект», по специальности 2.3.6 «Методы и системы защиты информации, информационная безопасность»

Актуальность темы диссертации

Диссертационная работа посвящена решению задачи поиска аномалий в поведении объектов на основании наблюдения за таким поведением. В таком самом общем виде это классическая задача пассивного тестирования или мониторинга, актуальность которой хорошо известна и не нуждается в подтверждении. Особенно актуальной эта задача становится для обеспечения информационной безопасности, поскольку информационная атака как раз и вызывает такое аномальное поведение, выявляя которое мы можем эту атаку обнаружить.

Поведение одного объекта в диссертации названо процессом, а его наблюдение — трассой: конечной последовательностью атомарных наблюдений. Заданное множество таких трасс называется логом. В случае нескольких процессов трасса содержит смесь атомарных наблюдений от разных процессов, что, конечно, усложняет задачу поиска аномалий. Именно эта усложнённая задача в основном и рассматривается в диссертации.

Основное содержание работы

Для решения поставленной задачи необходимы два этапа: 1) построение модели правильного поведения, 2) разработка алгоритмов поиска аномалий как отклонений от поведения, разрешаемого такой моделью.

Первая глава диссертации содержит обзор различных моделей и методов поиска аномалий. Вторая и третья главы посвящены поиску аномалий на основе двух моделей: сетей Петри и ациклических ориентированных графов.

Автор отмечает, что в контексте задач обеспечения информационной безопасности сети Петри ранее не рассматривались. Основной результат второй главы отрицательный: продемонстрировано отсутствие возможности однозначного восстановления модели процесса, представленной в терминах простейших сетей Петри, если на них наложены естественные условия корректности. Этот факт исключает возможность применения данного математического аппарата для решения задачи поиска аномалий без перехода к более сложному классу сетей Петри.

Иначе обстоит дело с ациклическими ориентированными графами. Автор отмечает, что ранее эта модель использовалась только при условии функционирования одного процесса, а в диссертации как раз сделан упор на смесь нескольких процессов. В диссертации показана возможность использования этой математической модели как для решения задачи построения модели процесса, так и для решения задачи поиска аномалий по построенной модели. Получены оценки сложности построения модели и выявления аномалий по построенной модели.

Научная новизна, обоснованность и достоверность научных положений, выводов и рекомендаций, сформулированных в диссертации

Как видно из вышеизложенного содержания диссертации новизна этой работы сосредоточена в двух пунктах:

1) Предложены решения и получены оценки временной сложности построения модели в виде ациклического ориентированного графа, а также алгоритмов выявления аномалий по этой модели. Предложено обобщение для случая функционирования нескольких процессов.

2) Продемонстрировано отсутствие возможности однозначного восстановления модели процесса, представленной в терминах простейших сетей Петри, если на них наложены естественные условия корректности, что исключает возможность применения данного математического аппарата для решения задачи поиска аномалий без перехода к более сложному классу сетей Петри.

Можно отметить, что главы 2 и 3 в основном опираются на ранее опубликованные работы: Van der Aalst`a и др. «Workflow Mining: Discovering process models from event logs» (номер 16 в списке литературы) и Agrawal и др. «Mining process models from workflow logs» (номер 11 в списке литературы). Этот бэкграунд, конечно, не умаляет новизны и ценности полученных результатов. Ньютону принадлежит парафраз «Если я видел дальше других, то потому, что стоял на плечах гигантов». Эйнштейн, создавая специальную теорию относительности, использовал преобразования Лоренца. Также и здесь, в диссертации (хотя и в меньшем масштабе): результаты этих статей модифицированы и интерпретированы для задачи поиска аномалий.

Достоверность полученных результатов подтверждается соответствующими математическими утверждениями (не считая следствий): в главе 2 теоремы 2.1 и 2.2, одна из которых заимствована, а одна доказана в тексте диссертации, Утверждения 2.1-2.4, два из которых заимствованы, а два доказаны в тексте диссертации, и леммы 2.1-2.9, все доказаны в тексте; в главе 3 теоремы 3.1-3.10, все доказаны в тексте диссертации, леммы 3.1-3.3, все заимствованы). Таким образом, из 28 утверждений (не считая следствий) в тексте диссертации доказано 78,6%, а заимствовано 21,4%.

Практическая ценность результатов

Практическая значимость полученных в диссертации результатов состоит в возможности их применения для анализа защищенности реальных систем. Это касается 3-й главы, в которой получены положительные

результаты: предложены алгоритмы и доказана оценка их сложности. Но имеет практическую ценность также и отрицательный результат 2-й главы, утверждающий нежелательность использования простейших сетей Петри как модели для поиска аномалий, поскольку это позволяет экономить научные ресурсы, направляя их по более перспективным направлениям.

Замечания по диссертационной работе

1) На взгляд оппонента обзор в первой главе страдает некоторой туманностью и расплывчатостью используемых понятий, определений и утверждений. Поскольку обзор, как и положено, опирается на ранее опубликованные работы различных исследователей, то этот туман и расплывчатость можно преодолеть, заглядывая в эти работы. Однако это требует от читателя дополнительных усилий, без которых остаются только догадки. Впрочем, следует отметить, что такова особенность большинства обзоров в диссертационных работах.

2) На шести страницах обзора (стр. 33-38) рассматриваются методы поиска аномалий без использования модели, а именно три метода. Но в самом начале указано, что в диссертации такой способ решения задачи поиска аномалий (без модели) рассматриваться не будет. Тогда зачем столь детально описывать эти методы? Для понимания сути того, что автором сделано, это ничего не даёт.

3) В главе 1 текст на стр. 39-40 озаглавлен «Определение того, что считать аномалией». К сожалению, никакого определения оппонент в этом тексте не увидел. Сформулировано несколько содержательных предположений и приведён пример их формального определения. Но в этом примере отношения «объясняет» и «имеет больше смысла» не определены, а только сказано, что они заданы и введены их обозначения. Справедливости ради нужно отметить, что этот недостаток видит и сам автор, начиная этот текст словами: «Следует отметить сложность в формальном определении того, что можно назвать аномалией. Часто авторы исходят из интуитивных

предположений». Если это такое рамочное определение, то можно было бы предположить, что в дальнейшем оно уточняется и применяется. Но это не так.

4) В главе 1 на стр. 40-41 расположен текст озаглавленный «Оценка качества поиска аномалий», и даже приведено несколько метрик качества. Но, к сожалению, далее, в главе 3, где как раз и предлагаются алгоритмы поиска аномалий, эти методы оценки никак не применяются для этих алгоритмов.

5) В выводах главы 1 утверждается, что в главе 3 использованы подходы «наивный, пороговый, итеративный, семплирующий» Однако на самом деле в главе 3 нет даже упоминаний этих подходов.

6) Глава 2 демонстрирует отсутствие возможности однозначного восстановления модели процесса, представленной в терминах простейших сетей Петри, если на них наложены естественные условия корректности. С этим можно согласиться, но «демонстрация отсутствия возможности» — не то же самое, что доказательство несуществования. Хотя следует отметить, что доказательства несуществования подчас очень трудные и, может быть, в данном случае это было бы излишней формализацией.

7) В Выводах главы 3 некоторые оценки памяти, требуемой алгоритмам, указаны как один из результатов работы. Однако эти оценки почему-то не указаны в числе основных результатов работы в Заключении. Возможно, автор не считает их основными? Хотя для практического применения алгоритмов такие оценки были бы полезны.

8) В подразделе «Научная новизна» Введения есть такие слова: «Показано, что критерий простоты описания математической модели не может быть решающим в выборе подходящей модели». Мы все, конечно, понимаем, что хорошо, когда модель, определение, утверждение, доказательство или теория простые, красивые и т.п., но всё же это никакой не критерий истины. К тому же оппонент, честно говоря, не заметил в тексте диссертации доказательства того, что «критерий простоты описания

математической модели не может быть решающим в выборе подходящей модели».

Заключение

Диссертация И. Ю. Терёхиной является завершенной научно-квалификационной работой, выполненной автором самостоятельно на достаточно высоком научном уровне, и вносит значимый вклад в теорию и практику поиска аномалий в поведении наблюдаемых объектов в контексте информационной безопасности.

Работа актуальна и имеет практическую ценность. Полученные автором результаты достоверны, выводы и заключения обоснованы.

Работа написана в основном доходчиво, грамотно и аккуратно оформлена. По каждой главе и работе в целом сделаны выводы.

Содержание диссертации соответствует специальности 2.3.6 «Методы и системы защиты информации, информационная безопасность».

Автореферат соответствует основному содержанию диссертации. Указанные выше замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В. Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 2.3.6 «Методы и системы защиты информации, информационная безопасность» (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В.Ломоносова, а также оформлена согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Московского государственного университета имени М.В.Ломоносова.

Таким образом, соискатель Терёхина Ирина Юрьевна заслуживает присуждения ученой степени кандидата физико-математических наук по

специальности 2.3.6 «Методы и системы защиты информации, информационная безопасность».

Официальный оппонент
доктор физико-математических наук,
главный научный сотрудник Института системного программирования РАН (ИСП РАН) им. В.П. Иванникова,
БУРДОНОВ Игорь Борисович

20.09.2024

Специальность, по которой официальным оппонентом защищена диссертация:
05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Адрес места работы:

109004, (Москва) г. Москва, ул. Александра Солженицына, д. 25, Федеральное государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ИСП РАН), отдел «Технологии программирования»
Тел.: +7(495) 912-44-25; e-mail: info-isp@ispras.ru

Подпись сотрудника ИСП РАН
И.Б. БУРДОНОВА удостоверяю:
руководитель/кадровый работник — И.О. Фамилия
дата

ученый секретарь ИСП РАН
8.10.2024 Емельянов О.А.