Сведения о выполненных работах в 2018 году
по проекту **«Надежность, безопасность и доверие в системах, используемых в качестве сервисов: масштабируемые решения для эффективного анализа и менеджмента»,** поддержанному Российским научным фондом

Соглашение № 16-49-03012

Руководитель д-р техн. наук Евтушенко Нина Владимировна

Как отмечалось в отчетах 2016, 2017 гг., понятие системы как сервиса (SaS) является, вообще говоря, новой концепцией, которая направлена на расширение существующих «облачных» технологий для любых типов систем с целью их использования в качестве (электронных) услуг. В 2018 г. мы продолжили исследования по разработке масштабируемых методов для обеспечения критериев безопасности, надежности и доверия SaS согласно техническому заданию по проекту; кроме того, в соответствии с планом исследований, мы посвятили данный этап разработке программных и/или аппаратных реализаций для предоставления сервисных систем по тестированию / оценке качества / доверия.

Рабочий пакет WP1. В рамках рабочего пакета WP1, посвященного эффективным манипуляциям с конечно автоматными моделями с целью синтеза умозрительных экспериментов с классическими и неклассическими автоматами, в 2018 году научная группа ТГУ продолжила развитие масштабируемых методов идентификации состояний конечных автоматов на основе последовательностных схем с использованием различных формул с кванторами (совместно с Тайваньским партнером), развитие эффективных адаптивных стратегий для этой цели, а также проанализировала влияние частичности и адаптивности на сложность различающих, установочных и синхронизирующих экспериментов с классическими автоматами, поскольку описания большинства сервисных систем определены лишь частично, и частичность и адаптивность можно рассматривать как две силы, работающие в противоположных направлениях; в некоторых случаях адаптивность может упростить построение умозрительных экспериментов с классическими и неклассическими автоматами. Предложенные подходы к построению установочных и различающих (адаптивных) последовательностей были опробованы на элементах сервисных систем, в частности, на системе итальянской железнодорожной компании, позволяющей пассажирам получать компенсации в случае задержки поездов.

Рабочий пакет WP2. В рамках развития концепта «все как сервис» (англ. – anything as a service) представляют интерес программные разработки, которые предоставляют пользователю возможность запрашивать услуги по синтезу тестов для различных типов спецификаций. Мы реализовали ряд сервисов по тестированию дискретных систем; были выбраны следующие спецификации для включения в сервисы по синтезу тестов: классические конечные автоматы, расширенные автоматы, последовательностные и комбинационные логические схемы. Сервисы предварительно опробованы как сотрудниками научной группы, так и студентами ТГУ. Сервис по построению тестов для конечных детерминированных автоматов FSMTestOnline.ru включает построение тестов с гарантированной полнотой на основе моделей «белого и черного ящиков». Для модели "белого ящика" генерируются все

одиночные мутации функций переходов и выходов; построенные последовательности объединяются в единый тест. Для модели «черного ящика» предъявленный автомат-спецификация проверяется на наличие необходимых свойств, таких как детерминированность, связность, и др.; тесты строятся обходом графа переходов и HSI-методом. Сервис по синтезу тестов для расширенных автоматов также включает построение тестов с гарантированной полнотой на основе моделей «белого и черного ящиков». Кроме того, пакет по расширенным автоматам позволяет вывести диаграмму переходов расширенного автомата, заданного пользователем, если заданный автомат не слишком большой, и сохранить данный автомат для построения тестов. Сервис по синтезу тестов на основе логических схем lctester.online в некотором смысле дополняет предыдущий, расширяя возможности пользователя при задании формальных спецификаций, и основополагающей моделью в данном случае выступает комбинационная или логическая схема. Тесты строятся на основе модели «белого ящика», при этом рассматриваются три типа возможных ошибок: одиночные константные неисправности, неисправности перемычек и трудно обнаружимые неисправности, изменяющие значения одного бита на одном входном наборе функционального элемента. Различающая последовательность для каждого из мутантов синтезируется с использованием средств верификации системы ABC (интегрирована в сервис). Данный сервис по синтезу тестов на основе логических схем может быть очевидным образом использован для синтеза тестов относительно модели классического детерминированного автомата. В этом случае необходимо обратиться к подходящему логическому синтезу, в случае, когда задание на синтез предоставляется в виде конечного автомата. Такая опция также предусмотрена в данном сервисе.

В рамках рабочего проекта WP 2 мы также рассмотрели новые модели неисправности и предложили методы синтеза тестов с гарантированной полнотой на их основе. В этом году мы предложили так называемый гибридный подход к построению проверяющего теста для цифровой схемы. На первом шаге по заданной логической схеме строится тест, обнаруживающий одиночные константные неисправности, неисправности «перемычек» и трудно обнаружимые неисправности, изменяющие выходное значение вентиля на одном входном наборе. На втором шаге, если по логической схеме можно построить автомат, то тест достраивается до обхода графа переходов построенного автомата или любого другого «автоматного» теста с гарантированной полнотой; как известно, такой тест обнаруживает большое количество функциональных ошибок в реализации системы. Если построить конечный автомат по логической схеме не представляется возможным, то устанавливается модель неисправности, относительно которой построенный тест является полным.

Часть работ по пакету WP 2 была посвящена исследованию новых моделей неисправности для дискретных (сервисных) систем. Поскольку частичность является характерной особенностью спецификаций таковых, например, сетевых протоколов, при реализации системы неопределенные переходы доопределяются некоторым образом, и нужно проверить, что при таком доопределении не возникает осцилляция. Для описания возможных реализаций используются мутационные автоматы, содержащие на неопределенных переходах возможные реализации этих переходов.

При композиции возникает переход в состояние ОСЦИЛЛЯЦИЯ, если осцилляция возможна для реализаций, переводящих композицию в данное состояние. Полный проверяющий тест должен содержать входные последовательности, которые «покрывают» каждый переход композиции в состояние ОСЦИЛЛЯЦИЯ, и построение такого теста не является тривиальной задачей для недетерминированного автомата, каковым является композиция мутационных автоматов. Мы уточнили понятие проверяющего теста на осцилляции, добавляя в тестовые последовательности, помимо входных воздействий, время ожидания выходных реакций, и, соответственно, уточнили понятие композиции для оценки таймаута при ожидании выходного символа. Мы также предложили новый метод построения тестов с гарантированной полнотой для инициальных автоматов с таймаутами. В отличие от классических автоматов, минимальная форма инициального автомата с таймаутами не является единственной (с точностью до изоморфизма), и, более того, минимальные формы одного автомата могут иметь различное число состояний. Соответственно, мы предложили метод синтеза проверяющего теста для автоматов с таймаутами относительно модели неисправности, где спецификацией является конечно автоматная абстракция, а область неисправности содержит каждый автомат с таймаутами, такой, что минимальная форма его конечно автоматной абстракции имеет не более m > 1 состояний.

Продолжено исследование сервисов, направленных на эффективную обработку, передачу и хранение информации. Предложены модели неисправности для тестирования архитектуры SDN и методы синтеза тестов с гарантированной полнотой на их основе. Установлено следующее. а) Набор параметризованных путей, в котором для каждой упорядоченной пары (h1, h2) различных хостов h1 и h2 множество TS содержит параметризованный путь, который начинается в h1 и заканчивается в h2, является hc-полным тестом относительно <=, FD>, где = - отношение равенства, а область неисправности FD содержит возможные реализации виртуальных путей. б) Набор параметризованных путей, в котором для каждой пары двух соседних коммутаторов (s1, s2) множество TS содержит параметризованный путь, имеющий ребро (s1, s2), является sc-полным тестом относительно <=, FD>. в) Набор параметризованных путей, в котором для каждого коммутатора s_i каждая пара его соседних узлов n_j и n_k и каждого хоста d существует путь из класса (s_i, n_j, n_k, d)-эквивалентности, является sf-полным тестом относительно <=, FD>.

Рабочий пакет WP3. В этом году мы рассматривали построение тестов относительно робастности и безопасности, которые во многих случаях строятся для композиций классических и неклассических автоматов. Были расширены возможности построения композиций расширенных автоматов без перехода к конечно автоматным абстракциям, исследованы проблемы, возникающие при построении параллельных композиций для временных автоматов, и предложены подходы к их решению. На основе автоматной модели предложен метод синтеза тестов относительно робастности; проведены исследования по анализу качества тестов, построенных по автоматным моделям, для обнаружения тупиков и осцилляций в композициях программных реализаций телекоммуникационных протоколов. Эксперименты проводились со студенческими реализациями различных протоколов: клиентское и серверное приложения протокола разрабатывались различными студентами, и их

совместная работа проверялась на наличие осцилляций (рабочий пакет WP 2). Если осцилляции в композиции отсутствовали, то в одну из компонент вносились достаточно простые мутации, и получившаяся композиция снова проверялась на наличие осцилляций. Все осцилляции, возникающие при таких мутациях, были обнаружены подходящими тестами. Более того, тест на осцилляции обнаружил также некорректную работу с памятью, т.е. тесты, построенные для проверки взаимодействия, в данном случае позволили обнаружить проблемы безопасности/надежности реализации.

Продолжены исследования в области поиска уязвимостей для веб сервисов на основе решения (полу-) автоматных неравенств и уравнений. Для определения множества входных последовательностей, которые могут привести к успешной атаке, Тайваньские коллеги используют функции PreConcatSuffix и PreConcatPrefix, аналогичные нахождению левого и правого частного и общего решения полуавтоматного неравенства относительно операции конкатенации. Установлено, что подход на основе правого/левого частного дает возможность выявить возможные уязвимости, однако может привести к ложным сигналам о наличии уязвимостей (ложные срабатывания). На основе полученных результатов предложены два подхода к поиску более точного (с точки зрения последующей санитизации) решения. Мы рассмотрели некоторые распространенные атаки на веб сервисы, а именно, SQL-инъекции и XSS-атаки, и проиллюстрировали, каким образом можно обнаружить уязвимости на основе нахождения правого/левого частного, т.е. на основе решения полуавтоматных неравенств. Удалось, в частности, обнаружить уязвимость на сервисе электронного расписания ТГУ. Получен ряд результатов в области санитизации пользовательских данных (работы неотъемлемо сопряжены с анализом уязвимостей веб сервисов). Для распознавания входных данных, представляющих различные атаки, мы попробовали использовать технологию машинного обучения, например, нейронную сеть, которая по введенным пользователем данным распознает, являются ли данные безопасными, и если нет, то при введении данных в какое поле заданная атака может быть реализована. Эксперименты по созданию и обучению подходящей нейронной сети проводились на основе SQL-инъекций и XSS-атаки; для создания нейронной сети использовался инструмент «Keras over TensorFlow». Для обучения нейронной сети была сгенерирована обучающая выборка: миллион различных строк с XSS-атаками, миллион различных строк с SQL-инъекцией вида «;‘OR 1= 1’», миллион различных строк с SQL-инъекцией общего вида. Каждая строка с атакой маскировалась случайными текстовыми данными. Были сгенерированы строки, которые «похожи» на атаки, но таковыми не являются (вместо тега <SCRIPT>, указывался несуществующий тег <ASCRIPT> и т.п.). Таких псевдо атак было сгенерировано три миллиона (по одному миллиону для каждого вида атаки). Были сгенерированы просто текстовые строки (вновь один миллион), длина строк в каждом случае была 100 символов. После обучения сети на 10% такой искусственно сгенерированной выборки нейронная сеть классифицировала угрозы с точностью 99,97%. Рабочие пакеты WP 4 и WP 5. На втором этапе выполнения проекта было принято решение о совмещении масштабируемых решений для предсказания уровня качества сервисных систем и уровня доверия к таким системам (четвертый и пятый рабочие пакеты), а также был предложен подход к прогнозированию обеих величин на основе специально синтезированных логических

схем. Соответственно, на текущем этапе выполнения проекта мы, в первую очередь, остановились на программно-аппаратной реализации предложенного подхода. Пусть логические схемы LC1 и LC2 используются для предсказания значения QoE / предсказания уровня доверия (K) по двум заранее выбранным «эталонным» технологиям машинного обучения M1 и M2, причем алгоритм M1 детерминирован в сторону фиксированного ответа K из множества D1, в то время как алгоритм M2 достаточно хорошо классифицирует данные из множества D2. Значение K предлагается параллельно вычислять на обеих схемах LC1 и LC2, а далее на выходе схемы, вычисляющей значение K для композитного сервиса, воспользоваться подходящим мультиплексором для «снятия» требуемых выходов. Для аппаратной реализации модели машинного обучения в виде логической схемы подходят программируемые логические интегральные схемы (ПЛИС). Процесс программирования (и перепрограммирования) логических схем занимает считанные минуты с использованием программной среды Altera Quartus II и отладочной платы Altera Cyclone V GX Starter Kit, имеющей в своем составе ПЛИС Cyclone V GX 5cgxfc5c6f27c7n. Такая программно-аппаратная реализация для композитной логической схемы, описанной выше, была построена в 2018 г. В дальнейшем планируется интеграция подходящей FPGA реализации в телекоммуникационную сеть для «Интернета вещей» (IoT), где узлами выступают подходящие сенсоры, и распознавание ведется на сети из элементов с ограниченными вычислительными возможностями, т.е. устройств с небольшими ресурсами и вычислительной мощностью, которые создаются для специальных задач, например, IoT. Для подтверждения эффективности нашего подхода мы провели предварительные эксперименты, показавшие превосходство (в терминах скорости исполнения и потребления ресурсов) реализаций логических схем перед программными реализациями ряда методов машинного обучения, таких как искусственные нейронные сети (ANN или ИНС) и системы опорных векторов (SVM). Производительность логической схемы превосходит модель SVM в 424 раза. Производительность логической схемы (реализованной аппаратно) оказалась выше производительности нейронной сети примерно в 1000 раз. Интересным также представляется вопрос использования «неклассических» моделей машинного обучения, в частности, одна из компонент LC1 или LC2 может быть синтезирована напрямую, т.е. без обращения к операции моделирования эталонной модели обучения. Согласно проведенным экспериментам, наибольшее влияние на качество обучения рассмотренных самообучающихся моделей (логическая схема и ИНС) оказывает именно объём обучающей выборки. Вторым параметром выборки, на который необходимо обращать внимание, является среднее расстояние между опорными векторами выборки. Необходимо стараться «кластеризировать» опорные векторы, располагать их плотными группами. В этом случае можно рассчитывать на эффективный синтез логических схем-компонент, участвующих в вычислении результирующего значения QoE / уровня доверия. Тайваньский партнер также оценивает возможности эффективной реализации нейронной сети подходящей логической схемой. Акцент делается на ИНС, у которых узлы скрытых слоев соединены по принципу «каждый с каждым»; рассматривается случай двоичных входных / выходных наборов. При условии, что каждый из нейронов имеет подходящую аппаратную реализацию, ставится цель оптимизировать число связей в результирующей схеме. На последнем семинаре, проведенном в Национальном

Университете Тайваня (октябрь 2018 г.), были начаты совместные обсуждения по предложению комбинированного (оконного или window) подхода на основе имеющегося задела двух научных групп. В частности, были поставлены следующие задачи: 1) определить свойства выделяемого фрагмента и условия «вырезания» нейронной подсети для дальнейшего моделирования; 2) определить композицию Verilog описания или AIG графа с подходящей аппаратной реализацией нейронной сети, полученной после оптимизации внутренних связей; 3) исследовать «независимость» двух и более фрагментов нейронной сети, которые могут быть промоделированы и оптимизированы параллельно и др.

Результаты, полученные в процессе выполнения работ по проекту, опубликованы в 18 работах, в том числе, 1 монография, 8 – в журналах и периодических изданиях (2 входят в международные базы цитирования Scopus и Web of Science и 6 – в Перечень ВАК), 9 – в трудах международных конференций (8 входят в международные базы цитирования Scopus и Web of Science и 1 совместная публикация с Тайваньским партнером), сделано более 15 докладов на конференциях и семинарах различного уровня, в том числе, международных. Таким образом, в 2018 г. был получен ряд научных результатов мирового уровня; разработанные методы и алгоритмы были опробованы на элементах сервисных систем, проведенные эксперименты подтвердили их эффективность. Полученные результаты и возможные направления дальнейшей работы обсуждались на семинарах в Тайбэе, в частности, с целью продолжения совместных исследований и возможности подачи заявки на грант РНФ-MOST 2019 года.

В рамках рабочих пакетов WP 1 и WP 2 **Тайваньский партнер** продолжает исследования в области анализа масштабируемых представлений для классических задач анализа (полу-) автоматов и формальных языков (с целью дальнейшего использования при проверке функциональных и нефункциональных требований к заданному сервису). В частности, в научной группе проф. Чианга был реализован специальный решатель для анализа строк и выполнения различных строковых операций (конкатенация, пересечение и др.). Примечательно, что основополагающим представлением для конечных полуавтоматов выступают логические схемы для подходящих характеристических функций, описывающих отношения переходов. Данный решатель может быть расширен на случай полуавтоматов с бесконечным множеством состояний, однако в этом случае вместо SAT и QBF задач для проверки существования и генерации контр-примеров партнеры предлагают обратиться к символьной формальной верификации.

Исследования Тайваньского партнера в рамках рабочего пакета WP 3 можно условно разделить на две части: 1) анализ уязвимостей в мобильных приложениях Apple; 2) исследование безопасности смарт-контрактов, используемых в технологии блокчейн.

Первая группа задач и полученные результаты в основном сводятся к эффективным реализациям технологий и алгоритмов, предложенных партнером на предшествующем этапе выполнения проекта. В частности, разработано соответствующее программное обеспечение. Задача идентификации вредоносных строк в данном случае решается на основе деассемблирования исходного кода с

дальнейшим статическим анализом графа управления. Следует отметить, что не все приложения поддаются подобному анализу. Причиной, в первую очередь, является вычислительная сложность перечисленных задач. Экспериментальные результаты позволили проанализировать более сотни приложений с подходящей генерацией вредоносных строк. Отмечается, что было также смоделировано семейство сценириев, при котором пользователь может получить доступ к конфиденциальным приложениям Apple. В результате злоумышленник имеет возможность управлять чужим устройством для личных целей, осуществлять звонки, передавать сообщения и др. По данным нашего партнера компания Apple не осведомлена об обнаруженных им уязвимостях в ее программном обеспечении. Поэтому наши коллеги планируют установить научное и техническое сотрудничество с данной компанией для внедрения полученных результатов.

Вторая группа задач и соответствующих исследований связана с реализациями специальных алгоритмов (протоколов), используемых на платформе Ethereum. Для анализа реализаций смарт-контрактов вновь применяется подход, основанный на статическом анализе графа управления для бинарного файла. В настоящий момент партнером проверяется наличие следующих двух уязвимостей в смарт-контрактах: а) переполнение стека и б) достижимость критического значения цены контракта. Программная реализация (с возможным расширением списка проверяемых свойств) находится в разработке.

Рабочие пакеты WP 4 и WP 5 связаны с масштабируемыми представлениями для задач машинного обучения и их использованием для предсказания уровня качества и доверия к предъявленному сервису. В этом случае Тайваньский партнер оценивает возможности эффективной реализации нейронной сети логической схемой. В предположении, что уже реализован функциональный элемент для каждого нейрона в сети, ставится задача уменьшения числа связей в полной сети. Оптимизация производится для бинарной сети на основе анализа миноров булевой матрицы, представляющей веса для каждого из уровней нейронной сети. Выделение миноров с повторяющимися строками или строками, являющимися инверсиями других строк, позволяет минимизировать количество связей в аппаратной реализации нейронной сети.