СБОРНИК ТЕХНОЛОГИЙ

ИНСТИТУТ СИСТЕМНОГО ПРОГРАММИРОВАНИЯ ИМ. В.П. ИВАННИКОВА РАН

75 ПОСВЯЩАЕТСЯ 75-ЛЕТИЮ ОТЕЧЕСТВЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

ИСП РАН

СБОРНИК ТЕХНОЛОГИЙ

2023

Институт системного программирования
им. В.П. Иванникова РАН

109004, Москва, улица А. Солженицына, дом 25.
По всем вопросам: scsec@ispras.ru

ОГЛАВЛЕНИЕ

2023. 75-ЛЕТИЕ ОТЕЧЕСТВЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

АРУТЮН АВЕТИСЯН

академик РАН,
директор ИСП РАН

В сборнике 2023 года представлены 27 технологий, которые распределены по тематическим блокам. В разделе «ИСП РАН: экосистема инноваций» приведено описание модели развития института, а также перечислены актуальные направления работ; далее идут годовые итоги деятельности научно-исследовательских центров на базе ИСП РАН. Но вначале — о главной теме этого года.

В этом году мы отмечаем 75-летие отечественных информационных технологий. Этот термин появился в конце 1950-х, но сами технологии возникли раньше. В России знаковая точка отсчёта – это 4 декабря 1948 года. Именно в этот день член-корреспондент Академии наук СССР И.С. Брук и инженер-конструктор Б.И. Рамеев представили в профильный Государственный комитет Совета министров СССР заявку на изобретение автоматической цифровой вычислительной машины. Авторское свидетельство №10475 стало первым документом, который зафиксировал начало работ по созданию отечественных ЭВМ.

29 июня 1948 года был основан Институт точной механики и вычислительной техники Академии наук (ИТМиВТ), который впоследствии возглавил выдающийся учёный С.А. Лебедев, а 19 декабря того же года – Специальное конструкторское бюро №245. В стране был создан целый ряд машин: МЭСМ, «Стрела», «М-100», «Днепр» и другие, среди них и БЭСМ-6 — первая отечественная суперЭВМ, чья скорость работы составляла 1 млн операций в секунду. Экземпляр БЭСМ-6 представлен в Музее науки в Лондоне как один из лучших суперкомпьютеров своей эпохи.

Параллельно развивалось образование: уже в 1949 году на механико-математическом факультете МГУ им. М.В. Ломоносова открылась кафедра вычислительной математики, началась подготовка первых системных программистов. В 1970-е годы по инициативе академика А.Н. Тихонова по всей стране стали открываться факультеты вычислительной математики и кибернетики. Появле-

ние и развитие российского сегмента интернета тоже неразрывно связано с наукой. Первая отечественная сеть была создана в Курчатовском институте и подключена к глобальному интернету в 1990 году.

Результаты проведённой работы и сформировавшиеся научные школы обеспечили долгосрочное развитие отрасли ИТ, в том числе и в постсоветское время. Примерами такого успешного развития стали Институт системного программирования им. В.П. Иванникова РАН, выросший из научной школы С.А. Лебедева; всемирно известные компании – «Яндекс», «Лаборатория Касперского» и другие.

Отрасль ИТ в России остаётся конкурентоспособной на мировом уровне и сейчас. Она постоянно растёт: например, в 2022 году объём реализованных товаров и услуг российских компаний-разработчиков ПО увеличился на 27% по сравнению с 2021 годом.

Наш институт тоже продолжает гармоничное развитие. В 2023 году мы стали Центром исследования безопасности системного программного обеспечения; сформирован консорциум компаний, которые вместе с нами проверяют ядро Linux и ключевые системные компоненты (OpenSSL, Qemu + libvirt, NodeJS и др.). Более 250 исправлений ошибок уже внесены только в ядро Linux.

Важнейшая область развития — это искусственный интеллект. Исследования и разработки ведутся по трём направлениям. Во-первых, сейчас с помощью ИИ расширяются возможности наших инструментов анализа программ на безопасность. Например, в статическом анализаторе Svace ведутся работы по применению машинного обучения для фильтрации ложных срабатываний, а также исследования по использованию больших языковых моделей — в частности, для повышения точности анализа. В безопасном компиляторе SafeC с помощью технологий ИИ автоматизируется проверка соответствия кода классам безопасности. В комплексе динамического анализа Crusher ИИ используется для повышения качества мутаций и понимания документации.

Во-вторых, в Исследовательском центре доверенного искусственного интеллекта (ИЦДИИ) на базе ИСП РАН мы создаём средства разработки ИИ-решений, в частности — программные инструменты для противодействия новым угрозам. Созданы доверенные версии фреймворков машинного обучения TensorFlow и PyTorch, которые уже внедрены в «Kaspersky Machine Learning for Anomaly Detection» v. 3.0. В основные ветки фреймворков внесены более 60 предложенных нами исправлений. Разрабатываются инструменты оценки устойчивости обученных моделей к атакам, инструменты повышения доверия к предобученным моделям и другие. Кроме того, создаётся доверенная версия платформы для построения интеллектуальных информационно-аналитических систем Talisman, которая объединяет более 50 моделей машинного обучения.

В-третьих, ИИ активно используется в наших междисциплинарных проектах. Платформа Talisman внедрена в процесс образовательной деятельности МГИМО МИД России и используется для интеллектуального анализа данных в области международных отношений. Совместный проект ИСП РАН и МГИМО в 2023 году получил премию «Гравитация» в специальной номинации «Открытие года». Развивается направление цифровой медицины. Создана и развёрнута в Сеченовском университете облачная платформа НЦМУ «Цифровой биодизайн и персонализированное здравоохранение», предоставляющая сервисы по сбору, разметке и анализу больших медицинских данных. Идёт дальнейшая работа по развитию и внедрению нейросетевой модели классификации 12-канальных ЭКГ; сейчас разработка проходит регистрацию в качестве медицинского изделия. Продолжается разработка системы DocMarking, предназначенной для повышения уровня информационной безопасности. Это система внедрения цифровых меток в текстовые документы, чья работа базируется на результатах исследований в областях стеганографии, цифровой обработки изображений и машинного обучения.

Актуальность междисциплинарных исследований нашла отражение и в тематике двух из пяти кандидатских диссертаций, защищенных в ИСП РАН в 2023 году: по анализу текстов в целях обнаружения межъязыковых заимствований, а также по построению программного конвейера для выравнивания последовательностей в приложениях биоинформатики.

Наш институт — это организационная структура вокруг научной школы, которая растёт и развивается уже много лет. Мы продолжаем традиции наших учителей, а великая история отечественной науки вдохновляет нас на новые успехи и достижения.

ИСП РАН: ЭКОСИСТЕМА ИННОВАЦИЙ

Деятельность ИСП РАН нацелена на внедрение результатов фундаментальных исследований в индустрию. Бизнес-модель института состоит из трёх тесно связанных активностей, которые в совокупности дают синергетический эффект:

- проектно-ориентированные фундаментальные и прикладные исследования в области системного программирования, нацеленные на создание новых технологий (по контрактам с российскими и зарубежными компаниями, Минобрнауки РФ, программам РАН, грантам РНФ и ФПИ и т.п.);
- внедрение новых технологий в компаниях-партнёрах и формирование инновационных продуктов после получения обратной связи от индустрии;
- обучение студентов и аспирантов на основе разработанных технологий (с обязательным участием в исследовательских и промышленных проектах института).

Такая модель хорошо известна и применяется в исследовательских лабораториях ведущих университетов (Stanford, MIT, Berkeley, Carnegie Mellon) и индустриальных гигантов (IBM, Intel), а также в государственных исследовательских центрах (INRIA, Fraunhofer). При условии эффективной реализации данная модель позволяет решить проблему разрыва между наукой и промышленностью, а также организовать подготовку кадров высшей квалификации.

ФУНДАМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Фундаментальные исследования и проведение экспериментальных работ — необходимые элементы деятельности института, позволяющие двигаться в русле новейших тенденций в мире IT, а также генерировать собственные идеи для проектов с бизнес-партнёрами. ИСП РАН ведёт большое число научных и образовательных программ и сотрудничает с ведущими российскими и зарубежными научными и университетскими центрами.

Это позволяет обеспечивать высокий уровень результатов исследований, а репутация в академических и университетских кругах открывает перспективу внедрения отечественных технологий на международных рынках.

В рамках научной деятельности ИСП РАН осуществляет выпуск собственного издания «Труды ИСП РАН» (индексируется в РИНЦ, включён в Russian Science Citation Index (RSCI)). Институт отвечает также за выпуск и редактуру журнала РАН «Программирование». Оба издания входят в перечень ВАК.

ВНЕДРЕНИЕ

ИСП РАН внедряет результаты своих исследований через крупные промышленные и исследовательские организации, которые одновременно используют новые технологии института и продвигают их в широкую практику. Большая часть работ по контрактам ведётся с долговременными партнёрами, в числе главных — Samsung, «Лаборатория Касперского», «Код Безопасности», «Открытая мобильная платформа», «СберТех», АО «НПО РусБИТех», ГосНИИАС, «Базальт СПО». В настоящее время технологии института используются более чем в 100 компаниях.

НАУЧНОЕ СОТРУДНИЧЕСТВО

Одна из форм организации долгосрочного сотрудничества в ИСП РАН — это совместные лаборатории. При наличии постоянного финансирования они позволяют гибко планировать имеющиеся ресурсы, а также наращивать компетенции во вновь образующихся направлениях системного программирования и организовывать подготовку молодых специалистов с компетенциями в интересующих партнёров областях.

С 2009 г. в институте работает совместная лаборатории с Samsung (нацелена на анализ программ, в том числе на обеспечение безопасности в контексте ОС Android и Tizen, а также на исследования в области применения методов искусственного интеллекта и анализа данных для задач программной инженерии). В 2019 г. были открыты совместные лаборатории с Huawei. Создана и успешно функционирует лаборатория для решения задач механики сплошных сред, реализующая исследовательские проекты в интересах промышленных предприятий. С 2021 г. в ИСП РАН работает Лаборатория интеллектуального цифрового прогнозирования и медиадаты.

На базе платформы Lingvodoc в институте работает лингвистическая лаборатория по документации исчезающих языков; исследования ведутся совместно с Институтом языкознания РАН, Томским государственным университетом, а также с другими вузами и НИИ. В 2023 г. четыре потока учёных из нескольких городов России прошли курсы дополнительного образования по направлению «Использование возможностей платформы Lingvodoc в работе лингвистов» (в том числе в Башкирском государственном университете и в РУДН). На открытых

заседаниях Всероссийского форума молодых учителей родного языка в Санкт-Петербурге и в Саранске были проведены мастер-классы по использованию обучающей платформы на занятиях в рамках сотрудничества с Институтом стратегии развития образования.

ЦЕНТРЫ

Важная миссия института заключается в создании и модерировании сообществ. Развиваются три центра:

- Научный центр мирового уровня (НЦМУ) «Цифровой биодизайн и персонализированное здравоохранение» (в консорциуме с Сеченовским университетом, ИБМХ, НовГУ и ИКТИ РАН);
- Исследовательский центр доверенного искусственного интеллекта; в сотрудничестве с Минэкономразвития, академическим сообществом (МФТИ, Сколтех, МНОЦ МГУ, Мехмат МГУ, Университет Иннополис, ННГУ, ИП РАН, МСЦ РАН) и индустрией (АО «Лаборатория Касперского», ЗАО «ЕС-Лизинг», «Интерпроком», «Технопром»);
- Центр исследования безопасности системного программного обеспечения (совместно с ФСТЭК России и при активном участии ведущих отечественных IT-компаний), в том числе ядра Linux и критичных компонентов с открытым исходным кодом. В 2023 г. вокруг Центра создан Консорциум организаций для совместных исследований.

ИНТЕЛЛЕКТУАЛЬНАЯ СОБСТВЕННОСТЬ

В бизнес-модели ИСП РАН права на интеллектуальную собственность остаются у института или передаются сообществу разработчиков свободного программного обеспечения (СПО) в рамках специальных соглашений. С учётом специфики данной модели была разработана оригинальная лицензия, базирующаяся не на получении роялти, а на прямом финансировании со стороны заказчика дальнейших исследований и разработок, направленных на развитие технологии. Заказчику передаются неисключительные права по использованию, при этом исключительные остаются за институтом. В отдельных ситуациях решение по управлению правами принимается индивидуально с учётом перспектив долгосрочного развития. Пример такого исключения — контракт с ФПИ, по которому все права передаются заказчику.

СВОБОДНОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ (СПО)

Один из важнейших компонентов созданной экосистемы — широкое использование СПО, без которого невозможно представить себе современное системное программирование. СПО рассматривается как:

- инструмент, предоставляющий легитимный свободный доступ ко всем современным технологиям, включая готовые к использованию программные продукты и открытые стандарты;
- возможность вести инновационное развитие без аутсорсинга благодаря взаимодействию с глобальным рынком продуктов и услуг;

- мощный образовательный ресурс: среда и инфраструктура международных СПО-проектов могут использоваться для подготовки специалистов. Научная деятельность подразумевает открытость результата и «видимость» его автора, что часто приходит в противоречие с корпоративной политикой IT-компаний. Для ИСП РАН открытость результатов исследований — это одновременно и стимул к работе, и инструмент продвижения технологий института. Открытость приводит к тому, что каждый молодой исследователь «виден» в международном сообществе IT-специалистов. Его вклад и репутация — это его капитал, и институт делает всё возможное, чтобы этот капитал рос максимально быстро.

ОБРАЗОВАНИЕ

Краеугольный камень экосистемы инноваций ИСП РАН — образовательная деятельность, которая осуществляется по нескольким направлениям:

- Интеграция ИСП РАН с ведущими вузами. Кафедры системного программирования, на которых работают сотрудники института, открыты в МГУ им. М.В. Ломоносова, МФТИ и НИУ ВШЭ. С первого курса студентам читаются лекции по дисциплинам системного программирования, а также проводятся семинары. На третьем курсе, после распределения на к]афедры системного программирования, студенты продолжают слушать лекции, посещают спецсеминары, знакомятся с исследовательской тематикой по научным направлениям института, начинают участвовать в проектах и получать специальную стипендию. К моменту выпуска многие учащиеся имеют научные публикации и уже являются специалистами по системному программированию.

  Сотрудники ИСП РАН проводят обновление учебных курсов и модернизацию программ, развивая кибербезопасность как научную специальность. В частности, в 2021 г. под научным руководством института началась модернизация бакалаврской программы «Программная инженерия» на факультете компьютерных наук НИУ ВШЭ. В 2022 г. заключено соглашение с Чувашским государственным университетом и ООО «Кейсистемс-безопасность» (Чебоксары). Стороны совместно принимают участие в разработке учебных дисциплин факультета информатики и вычислительной техники. В ЧГУ откроется лаборатория системного программирования и безопасной разработки ПО.

  Расширяется сотрудничество с МГТУ им. Н.Э. Баумана. На кафедре ИУ-10 запущен специализированный курс «Система сертификации средств защиты информации», действует семинар для студентов, заинтересованных в системном подходе к различным аспектам кибербезопасности; планируется оптимизация учебных программ специалитета, чтобы курсы, посвященные кибербезопасности и анализу кода, изучались уже в первые годы обучения. ИСП РАН сотрудничает с Саратовским государственным университетом: студенты СГУ пишут дипломы и статьи под руководством сотрудников инсти-

тута, а также участвуют в проектах, в частности, в разработке безопасного компилятора. В 2022 г. заключено соглашение о сотрудничестве между ИСП РАН и Московским энергетическим институтом (МЭИ).

Ведутся работы и по другим направлениям. В 2022 г. совместно с МГИМО МИД России запущена магистерская программа «Анализ данных и динамика международных процессов» – для подготовки специалистов в области анализа данных, использования искусственного интеллекта и моделирования социально-экономических процессов. Ведутся работы в рамках совместной Лаборатории интеллектуального анализа данных. В 2023 г. совместный проект ИСП РАН и МГИМО МИД России получил премию «Гравитация» в специальной номинации «Открытие года». Проект посвящен внедрению платформы Talisman, созданной в ИСП РАН, в процесс образовательной деятельности МГИМО.

Кроме того, ИСП РАН начал совместный проект с Российско-Армянским университетом (РАУ) и компанией «Антиплагиат», посвящённый разработке методов автоматического обнаружения заимствований в текстовых документах на разных языках. В работе используются глубокие нейросетевые модели, которые помогают точнее проводить детальный анализ текстов; предполагается также разработать универсальные методы анализа.
В 2023 г. на базе Института динамики систем и теории имени В. М. Матросова СО РАН создана совместная с ИСП РАН научно-исследовательская группа. В числе основных направлений деятельности – разработка программных средств анализа электронных документов и обработки естественного языка. Участники группы — научные сотрудники ИДСТУ СО РАН и стипендиаты совместной программы ИСП РАН и Института математики и информационных технологий ИГУ.

В 2022 г. ИСП РАН стал партнёром Московского авиационного института МАИ по федеральному проекту «Передовая инженерная школа», который посвящён разработке нового поколения летательных аппаратов. Планируется открытие совместной лаборатории и запуск новой магистерской программы по теме БПЛА. Вместе с МАИ и другими организациями институт вошёл в консорциум «Новые аэрокосмические рынки» в рамках программы «Приоритет-2030». В рамках той же программы в 2023 г. сотрудники ИСП РАН приняли участие в образовательном проекте «Цифровая кафедра МАИ» по направлениям «Искусственный интеллект в дистанционном зондировании Земли» и «Цифровое моделирование и суперкомпьютерные технологии». Кроме того, представители ИСП РАН руководили студентами-дипломниками в МАИ, а также принимали экзамены.

С 2017 г. ИСП РАН сотрудничает с IT Академией Samsung. В частности, сотрудники института входят в жюри Межвузовского конкурса проектов, который проводится ежегодно для демонстрации лучших практик и результатов учебной деятельности, реализованной в вузах-партнёрах Академии.

- Стипендиальная программа. В рамках поддержки образовательных процессов институт запустил стипендиальную программу, которая охватывает студентов ряда образовательных организаций, в числе которых МГУ им. М.В. Ломоносова, МФТИ, НИУ ВШЭ, Новгородский государственный университет им. Ярослава Мудрого, Российско-Армянский университет и др.

- Собственная аспирантура ИСП РАН, предусматривающая одновременно накопление практического опыта и изучение новых технологий. Аспиранты активно вовлекаются в процессы обучения: ведут семинарские и практические занятия со студентами, руководят подготовкой курсовых и дипломных работ. Накопив такой опыт, выпускник аспирантуры, как правило, становится руководителем небольшой исследовательской группы.

- Развитие сети лабораторий системного программирования. Активно работают лаборатории в Ереване, в Великом Новгороде, в Орле; организована лаборатория в РЭУ им. Г.В. Плеханова. Лаборатории привлекают к работе успешных студентов и аспирантов, которые занимаются разработкой перспективных технологий в тесном сотрудничестве с индустрией.

КОНФЕРЕНЦИИ

Ежегодно институт проводит ряд мероприятий:

Международная Открытая конференция ИСП РАН им. В.П. Иванникова: https://www.isprasopen.ru/

Научно-практическая конференция OS DAY (совместно с другими организаторами): https://www.osday.ru/

Международная конференция «Иванниковские чтения»: https://www.ivannikov-ws.org/

Международная научно-практическая конференция «Анализ данных в медицине» (совместно с другими организаторами): https://digital-med.ru/

Весенняя конференция молодых учёных в области программной инженерии: http://syrcose.ispras.ru/

Круглый стол «Системное программирование как ключевое направление противодействия киберугрозам» (Международный военно-технический форум «Армия»)

2023. НАУЧНЫЙ ЦЕНТР МИРОВОГО УРОВНЯ (НЦМУ) «ЦИФРОВОЙ БИОДИЗАЙН И ПЕРСОНАЛИЗИРОВАННОЕ ЗДРАВООХРАНЕНИЕ»

В КОНСОРЦИУМЕ С СЕЧЕНОВСКИМ УНИВЕРСИТЕТОМ, ИБМХ, НОВГУ И ИКТИ РАН

В ЧИСЛЕ ГЛАВНЫХ ДОСТИЖЕНИЙ 2023 ГОДА:

СОЗДАНА И ВВЕДЕНА НА УРОВНЕ СЕЧЕНОВСКОГО УНИВЕРСИТЕТА ОБЛАЧНАЯ ПЛАТФОРМА НЦМУ «ЦИФРОВОЙ БИОДИЗАЙН И ПЕРСОНАЛИЗИРОВАННОЕ ЗДРАВООХРАНЕНИЕ».

Разработанная ИСП РАН Платформа НЦМУ обеспечивает предоставление следующих сервисов:

- базовых облачных сервисов (например, виртуальных серверов и блочных устройств по запросу);
- сервисов по сбору, хранению и анализу больших медицинских данных;
- сервисов по разметке медицинских данных и по применению алгоритмов машинного обучения для решения задач биомедицинского домена;
- сервисов поддержки процессов проведения совместных исследований.

Платформа НЦМУ реализована на базе облачной среды Asperitas (ИСП РАН). В 2023 г. в Платформу включен функционал по формированию научной базы знаний в рамках медицинского исследования с использованием технологий сбора и анализа больших данных, реализованный на базе информационно-аналитической системы Talisman (ИСП РАН). Тестирование облачных сервисов осуществлялось на примере web-лабораторий по анализу данных электрокардиограмм и гистологических изображений. К концу 2024 г. планируется обеспечить полномасштабную опытную эксплуатацию с возможностью подключения внешних участников.

Платформа НЦМУ может быть развёрнута на базе облачной инфраструктуры ИСП РАН или сторонней облачной инфраструктуры для биомедицинских задач, решаемых в рамках НЦМУ, или адаптирована под задачи других медицинских доменов.

ИДЁТ ДАЛЬНЕЙШАЯ РАБОТА ПО РАЗВИТИЮ И ВНЕДРЕНИЮ НЕЙРОСЕТЕВОЙ МОДЕЛИ КЛАССИФИКАЦИИ 12-КАНАЛЬНЫХ ЭКГ.

Нейросетевая модель классификации 12-канальных ЭКГ обучена на данных из разных регионов (Республика Татарстан, Москва, Великий Новгород), интегрирована в режиме опытной эксплуатации в систему «Единый Кардиолог» и апробирована на данных ЭКГ Республики Татарстан. Подписано соглашение о сотрудничестве со Станцией скорой и неотложной медицинской помощи им. А.С. Пучкова. Проанализировано несколько десятков тысяч ЭКГ, полученных от Станции. Качество предсказания моделей сравнимо с 10-секундными 12-канальными ЭКГ. Подписаны протоколы тестирования. В настоящее время разработка проходит регистрацию в качестве медицинского изделия.

РАНЕЕ В РАМКАХ НЦМУ:

РАЗРАБОТАН МАКЕТ СИСТЕМЫ РАЗМЕТКИ 12-КАНАЛЬНЫХ ЭКГ (HTTP://ECG1.ISPRAS.RU).

Качественная стандартизированная разметка по заранее определённому списку патологий помогает достичь высокой степени согласия между экспертами. Макет системы разметки подготовлен для интеграции в облачную платформу ИСП РАН Asperitas для прозрачного масштабирования мощностей по хранению и анализу ЭКГ, но может быть интегрирован и в стороннюю облачную экосистему.

ОБУЧЕНА НЕЙРОСЕТЕВАЯ МОДЕЛЬ ДЕТЕКЦИИ КЛЕТОЧНЫХ ЯДЕР ENDONET НА ГИСТОЛОГИЧЕСКИХ ПРЕПАРАТАХ.

Нейросеть обучена на размеченном гистологическом наборе EndoNuke, собранном совместно с партнёрами (РУДН, ГКБ №31, НМИЦ АГП им. В.И. Кулакова, НовГУ и НИИ морфологии человека). Модель детекции ядер встроена в открытую программную платформу для анализа биоизображений QuPath с использованием технологии ИСП РАН Fanlight. Модифицированная платформа QuPath и система разметки изображений с открытым кодом CVAT подготовлены для интеграции в облачную платформу ИСП РАН Asperitas.

2023. ЦЕНТР ИССЛЕДОВАНИЯ БЕЗОПАСНОСТИ СИСТЕМНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

СОВМЕСТНО С ФСТЭК РОССИИ
В ПАРТНЁРСТВЕ С ИНДУСТРИЕЙ
PORTAL.LINUXTESTING.RU

ГЛАВНЫЕ ДОСТИЖЕНИЯ 2023 ГОДА:

ОРГАНИЗАЦИОННЫЕ:

- создан Консорциум по поддержке Технологического центра исследования безопасности ядра Linux, к которому присоединились 33 организации;
- развёрнута инфраструктура для проведения исследования критичных компонентов системного программного обеспечения с открытым исходным кодом;
- на их основе сформирован единый Центр исследования безопасности системного программного обеспечения.

МЕТОДИЧЕСКИЕ:

- подготовлены методики проведения исследования ядра Linux, OpenSSL, NGinx, Qemu, libvirt, podman, .NET6 Runtime, ASP .NET Core;
- подготовлены рекомендации по конфигурированию ядра с целью повышения его безопасности;
- подготовлены рекомендации по настройке доверенной загрузки ядра.

ТЕХНОЛОГИЧЕСКИЕ:

- началось сопровождение второй ветки ядра Linux, основанной на стабильной версии 6.1;
- проведён анализ 17 тысяч предупреждений инструмента статического анализа Svace (разработан в ИСП РАН);
- подготовлены более 250 исправлений, которые уже приняты в основную ветку ядра;
- подготовлены исправления, которые приняты в основные ветки компонентов OpenSSL, Qemu, libvirt, CPython, Lua, .NET6 Runtime.

ПАРТНЁРЫ ЦЕНТРА:

- АО «Аладдин Р.Д.»
- ООО «Айдеко»
- ООО «АНКАД»
- ООО «Базальт СПО»
- АО «Байкал электроникс»
- ООО «БЕЛЛСОФТ»
- ЗАО «ЗЭТ»
- АО «ИВК»
- ООО «Инферит»
- АО «ИнфоТеКС»
- ООО «ИТБ»
- ООО «Код Безопасности»
- ООО «Конфидент»
- АО НТЦ «Модуль»
- АО «МЦСТ»
- АО «НППКТ»
- ООО «Открытая мобильная платформа»
- ООО «ПиЭлСи Технолоджи»
- АО «РАСУ»
- ООО «РЕД СОФТ»
- ООО «НТЦ ИТ РОСА»
- ООО «РусБИТех-Астра»
- ФГУП «РФЯЦ-ВНИИЭФ»
- АО МВП «Свемел»
- ООО «ТехАргос»
- ООО «Фактор-ТС»
- АО «ФИНТЕХ»
- ООО «Юзергейт»
- АО «НПО «Эшелон»
- ООО «ЯНДЕКС.ОБЛАКО»

ОБРАЗОВАТЕЛЬНЫЕ ПАРТНЁРЫ:

- МГУ им. М.В. Ломоносова
- МФТИ
- НИУ ВШЭ
- Вологодский государственный университет
- НИУ «МЭИ»
- МГТУ им. Н.Э. Баумана
- Воронежский государственный университет
- Чувашский государственный университет им. И.Н. Ульянова

2023. ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР ДОВЕРЕННОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В СОТРУДНИЧЕСТВЕ С МИНЭКОНОМРАЗВИТИЯ РОССИИ,

С АКАДЕМИЧЕСКИМ СООБЩЕСТВОМ (МФТИ, СКОЛТЕХ, МНОЦ МГУ, МЕХМАТ МГУ, УНИВЕРСИТЕТ ИННОПОЛИС, ННГУ, ИП РАН, МСЦ РАН),

С ИНДУСТРИЕЙ (АО «ЛАБОРАТОРИЯ КАСПЕРСКОГО», ЗАО «ЕС-ЛИЗИНГ», «ИНТЕРПРОКОМ», «ТЕХНОПРОМ»).

ГЛАВНЫЕ ДОСТИЖЕНИЯ 2023 ГОДА:

- Создана отчуждаемая методика разработки доверенных фреймворков машинного обучения (TensorFlow, PyTorch).
- Сформирован коллектив, который может оперативно исправлять уязвимости в базовом ПО для машинного обучения, обеспечивая технологическую независимость в области создания систем искусственного интеллекта.
- Более 60 исправлений уже приняты в основные версии фреймворков.
- Доверенные версии фреймворков, которые апробированы на решениях индустриальных партнёров ИЦДИИ, внедрены в «Kaspersky Machine Learning for Anomaly Detection» v. 3.0. Ведётся инициативная сертификация для нового заказчика – АО «КТ — Беспилотные системы».
- Создана доверенная версия платформы для построения интеллектуальных информационно-аналитических систем Talisman, которая объединяет более 50 моделей машинного обучения и соответствует критериям доверия к системам, использующим технологии ИИ (критерии разработаны в рамках Программы ИЦДИИ). Доверенная версия проходит апробацию у индустриальных партнёров: ЗАО «ЕС-Лизинг» и «Интерпроком».
- Создаётся облачная платформа для анализа и разработки доверенных систем, использующих технологии ИИ. Платформа объединяет программные инструменты и методики для противодействия принципиально новым угрозам, возникающим на всех этапах жизненного цикла соответствующих технологий:
  - доверенные фреймворки и библиотеки машинного обучения;
  - инструменты проверки наличия аномалий в наборах данных;
  - инструменты оценки устойчивости обученных моделей к атакам;
  - инструменты для повышения доверия к предобученным моделям;
  - методы защиты моделей от атак на этапе эксплуатации;
  - методы объяснения моделей;
  - методы обнаружения дрейфа данных;
  - методы выявления предвзятости моделей.

1 АНАЛИЗ ПРОГРАММ И КИБЕР-БЕЗОПАСНОСТЬ

ASTRAVER TOOLSET: СИСТЕМА ВЕРИФИКАЦИИ КЛЮЧЕВЫХ КОМПОНЕНТОВ

AstraVer Toolset — система дедуктивной верификации ключевых компонентов. Позволяет разрабатывать и верифицировать модели политик безопасности, а также проводить доказательство корректности компонентов на языке C. Необходимый инструмент достижения целей семейств доверия ADV_SPM и ADV_FSP, определённых в ГОСТ Р ИСО/МЭК 15408-3-2013.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

AstraVer Toolset — комплекс инструментов, предназначенный для промышленного использования и основанный на многолетних научных исследованиях. Объединяет два подхода к верификации: на уровне моделей и на уровне кода. Решает те же задачи, что и аналогичные инструменты (Microsoft VCC, Frama-C WP), однако благодаря специфической доработке обладает технологической уникальностью: возможностью верификации ключевых компонентов системы безопасности ядра Linux. Выложен в открытый доступ (http://linuxtesting.ru/astraver).

AstraVer Toolset — это:

- Комплексный подход к верификации, начиная с формализации требований верхнего уровня и до анализа поведения кода.
- Моделирование функциональных требований (формализация функциональных требований к системе, доказательство внутренней согласованности требований и недостижимости небезопасных состояний).
- Тестирование реализации на соответствие функциональным требованиям с использованием формальной модели требований для проверки корректности наблюдаемого поведения в целях оценки качества тестирования и генерации тестов.
- Верификация ключевых компонентов на языке C (формализация требований к ключевым компонентам, доказательство корректности работы компонента на всех возможных входных данных).

- Поддержка индустриального кода (нестандартные расширения компилятора GCC, арифметические операции с побитовой точностью, адресная арифметика (включая поддержку конструкции container_of), функциональные указатели, приведение целочисленных типов к указательным).
- Решение важнейших задач профилей защиты:
  - формальное моделирование политики безопасности;
  - формальное доказательство внутренней непротиворечивости модели политики безопасности и недостижимости небезопасных состояний;
  - разработка полуформальной или формальной функциональной спецификации;
  - формальное или полуформальное доказательство соответствия между моделью политики безопасности и функциональной спецификацией;
  - формальное или полуформальное доказательство соответствия между различными представлениями целевого ПО, такими как функциональная спецификация, проект ПО и его реализация.
- Возможность доработки комплекса под конкретного заказчика (в плане поддержки верификации компонентов на языке С).

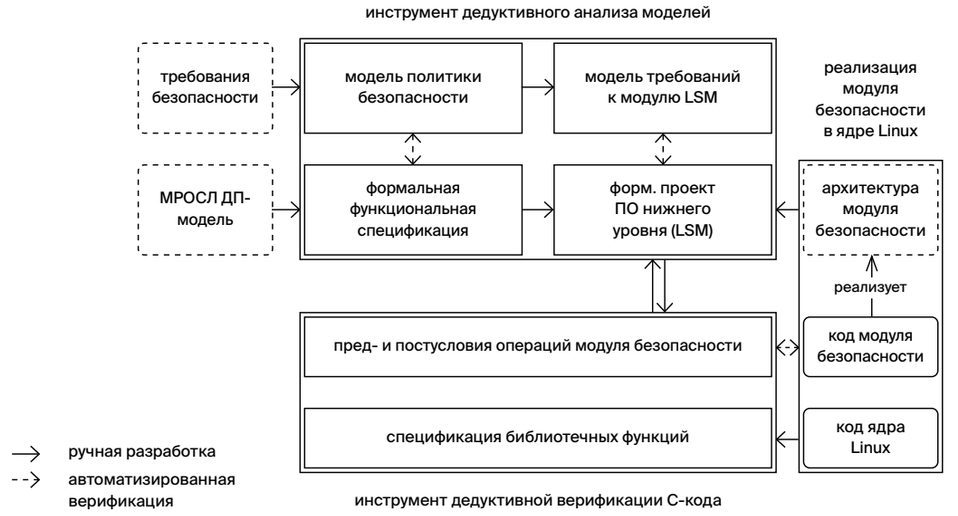
ДЛЯ КОГО ПРЕДНАЗНАЧЕН ASTRAVER TOOLSET?

- Компании, нацеленные на разработку ПО с высокой степенью надёжности и безопасности — как информационной, так и функциональной (ПО для самолётов, АЭС и др.).
- Компании, которые нуждаются в сертификации разрабатываемого ПО в соответствии с ГОСТ Р ИСО/МЭК 15408.
- Испытательные лаборатории средств защиты информации в соответствии с требованиями безопасности.

ОПЫТ ВНЕДРЕНИЯ

Система AstraVer Toolset применялась при разработке средств защиты информации ОС Astra Linux Special Edition (АО «НПО РусБИТех»), которая успешно прошла сертификацию на соответствие требованиям безопасности информации ФСТЭК России к операционным системам по профилю защиты «2А». В основу отечественной разработки была положена МРОСЛ-ДП модель безопасности, а реализация её новых возможностей в ОС Astra Linux Special Edition продолжает верифицироваться с помощью AstraVer Toolset.

СХЕМА РАБОТЫ



KLEVER: СИСТЕМА ВЕРИФИКАЦИИ МОДЕЛЕЙ ПРОМЫШЛЕННОГО ПО

Klever — система верификации моделей, которые генерируются автоматически на основе исходного кода промышленного программного обеспечения, разрабатываемого на языке программирования Си. Klever позволяет специфицировать различные требования по безопасности и надёжности, а также проверять их выполнение автоматически с заданным уровнем точности.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Klever базируется на результатах передовых научных исследований в области автоматического построения и верификации моделей программ. В основе системы верификации лежат методы для покомпонентной верификации, моделирования окружения и спецификации требований, что делает возможным применение формальных методов к исходному коду промышленного программного обеспечения размером в сотни тысяч и миллионы строк кода на языке программирования Си. Klever является проектом с открытым исходным кодом (https://forge.ispras.ru/projects/klever).

Klever — это:

- Высокоточный консервативный анализ исходного кода промышленного программного обеспечения, позволяющий выявлять все возможные ошибки искомых видов и доказывать корректность программ при явно заданных предположениях.
- Масштабируемость. Покомпонентная верификация программ позволяет применять к большому объёму исходного кода наиболее точные методы анализа, такие как верификация моделей и символьное выполнение.
- Возможность адаптации под конкретные нужды заказчиков. Разработка спецификаций моделей окружения целевых программ, а также спецификаций для обнаружения нарушений, специфичных для программ требований, в дополнение к проверке общих правил безопасного программирования на языке Си.
- Детализированное представление выявленных ошибок. При обнаружении ошибок система верификации выдаёт подробные трассы ошибок, включающие конкретные значения переменных и аргументов вызываемых функций.
- Удобный многопользовательский веб-интерфейс для подготовки и запуска верификации, а также для выполнения экспертной оценки результатов верификации.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН KLEVER?

- Компании, разрабатывающие программное обеспечение с высоким уровнем надёжности и безопасности.
- Испытательные лаборатории.

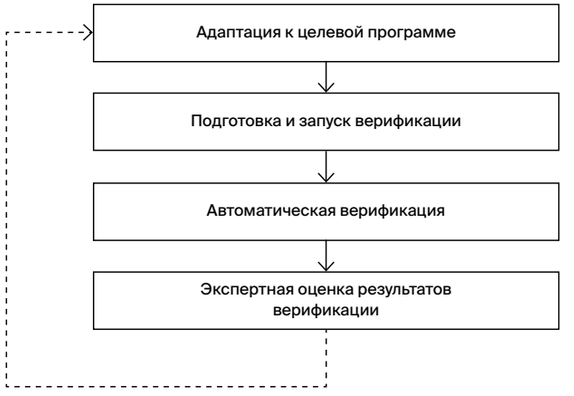
ОПЫТ ВНЕДРЕНИЯ

В основном система верификации Klever применяется для тщательной проверки ядер и драйверов различных операционных систем. Для демонстрации возможностей системы Klever были выполнены работы по верификации драйверов устройств ядра Linux. В результате удалось обнаружить более 400 ошибок следующих типов: выход за границу буфера, разыменование нулевого указателя, использование неинициализированной памяти, повторное или некорректное освобождение памяти, утечка памяти, состояние гонки, взаимная блокировка, некорректный вызов функции в зависимости от контекста, некорректная инициализация структур данных ядра Linux и т.д. Данные ошибки были признаны разработчиками ядра Linux.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

Ubuntu 18.04/20.04, не менее 4-х процессорных ядер x86-64, 16 ГБ оперативной памяти и 100 ГБ свободного места на диске.

СХЕМА РАБОТЫ



MASIW: НАБОР ИНСТРУМЕНТОВ ДЛЯ ПРОЕКТИРОВАНИЯ ОТВЕТСТВЕННЫХ СИСТЕМ

MASIW — набор инструментов для разработки программно-аппаратных комплексов ответственных систем в сфере авиации, медицины и др. Создан для инженеров-конструкторов комплексов бортового оборудования для авиационных судов, разрабатываемого с применением интегрированной модульной авионики (ИМА). Оперативно адаптируется под другие предметные области.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

MASIW — технология для оптимизации разработки сложных программно-аппаратных комплексов, а также их верификации. Позволяет провести предварительную оценку качества изделия до появления опытного образца, а также анализ на отказоустойчивость. Снижает риск появления ошибок и дефектов. Разрабатывается совместно с ФГУП «ГосНИИАС». Несмотря на наличие инструмента OSATE на момент начала разработки, на сегодняшний день MASIW превосходит его по функциональности в плане верификации, а также статического и динамического анализа.

MASIW — это:

- Создание, редактирование и управление моделями на языке AADL:
  - создание/редактирование моделей посредством текстового или графического редактора;
  - поддержка командной разработки с возможностью отслеживания и внесения изменений для отдельных элементов модели;
  - поддержка переиспользования AADL-моделей сторонних разработчиков.
- Анализ моделей:
  - анализ структуры программно-аппаратного комплекса (достаточности аппаратных ресурсов, согласованности интерфейсов и т. п.);
  - проверка разрабатываемого программно-аппаратного комплекса на соответствие требованиям;
  - анализ характеристик передачи данных в сети AFDX (времени доставки сообщений от отправителя к получателю, глубины очередей передающих портов и т. п.);
  - построение дерева неисправностей и его численный анализ для определения вероятности отказного события верхнего уровня;
  - анализ видов и последствий отказов на основе архитектурной модели комплекса бортового оборудования, включая построение таблицы видов и последствий отказов;
  - симуляция модели программно-аппаратного комплекса с генерацией пользовательских отчётов по результатам работы симулятора, в том числе совместная симуляция работы прикладных разделов под управлением ОС РВ в эмуляторе QEMU и универсального симулятора AADL моделей.
- Синтез моделей:
  - распределение функциональных приложений по вычислительным модулям с учётом ограничений ресурсов аппаратной платформы и с учётом дополнительных ограничений, касающихся вопросов надёжности и безопасности программно-аппаратного комплекса;
  - генерация распределения вычислительного времени процессора между функциональными приложениями (циклограмма расписания запуска приложений для ARINC-653 совместимых ОС РВ).
- Генерация конфигурационных данных:
  - разработка специализированных инструментов конфигурационных данных на основе предоставляемого программного интерфейса (API);
  - генерация конфигурационных файлов для компонентов КБО.
- Возможность расширения набора инструментов путём создания собственных модулей (благодаря модульной архитектуре в основе технологии).

СХЕМА РАБОТЫ

MICROTESK: ГЕНЕРАТОР ТЕСТОВЫХ ПРОГРАММ

MicroTESK — реконфигурируемая и расширяемая среда генерации тестовых программ для функциональной верификации микропроцессоров. Позволяет автоматически конструировать генераторы тестовых программ для целевых архитектур микропроцессоров на основе их формальных спецификаций. MicroTESK применим для широкого спектра архитектур (RISC, CISC, VLIW, DSP). Поддерживает онлайн-генерацию тестовых программ.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

MicroTESK — комплекс технологий для промышленного использования, включающий в себя базовую среду моделирования (строит модели микропроцессоров на основе формальных спецификаций) и среду генерации (строит тестовые программы на основе шаблонов). По решаемым задачам близок к мировым аналогам (Genesys Pro и RAVEN), однако отличается от них повышенной производительностью и удобством использования. Распространяется по лицензии Apache 2.0. MicroTESK доступен на сайте ИСП РАН: https://forge.ispras.ru/projects/microtesk. Описание технологии — на сайте http://www.microtesk.org.

MicroTESK — это:

- Использование формальных спецификаций в качестве источников знаний о конфигурации верифицируемого микропроцессора:
  - спецификации архитектуры на nML (регистры, память и режимы адресации, логика инструкций, текстовый/бинарный формат инструкций);
  - дополнительные спецификации подсистемы памяти на mmuSL (свойства буферов памяти (TLB, L1 и L2), логика трансляции адресов и логика операций чтения и записи);
  - потенциальная возможность перехода к формальной верификации, а также к генерации набора инструментов для разрабатываемого микропроцессора (дизассемблер, эмулятор и др.).
- Генерация тестовых программ на основе объектно-ориентированных тестовых шаблонов:
  - тестовые шаблоны на языке Ruby (за счёт чего шаблоны наглядны и удобны в поддержке);
  - возможность одновременного использования различных техник генерации наборов инструкций и тестовых данных (случайная генерация, комбинаторная генерация, генерация на основе разрешения ограничений и др.);
  - масштабируемость среды генерации (возможность разрабатывать сложные шаблоны при небольших затратах за счёт повторного использования).
- Широкий набор поддерживаемых архитектур микропроцессоров:
  - поддерживаются особенности различных классов архитектур на уровне среды разработки генераторов (RISC, CISC, VLIW, DSP);
  - разработаны генераторы тестовых программ на основе MicroTESK для таких архитектур, как RISC-V, ARM, MIPS, PowerPC;
  - поддерживается многоядерность целевой микропроцессорной архитектуры.
- Оперативная настройка среды под новые архитектуры с минимальными затратами и автоматическое извлечение информации о тестовых ситуациях (благодаря формальным спецификациям).
- Удобный язык разработки тестовых шаблонов, позволяющий быстро описывать сложные сценарии верификации.
- Поддержка онлайн-генерации тестовых программ для проведения пост-производственной верификации целевого микропроцессора. Онлайн-генерация осуществляется исполняемым генератором, входящим в состав MicroTESK. Генератор строит тестовые последовательности по формальным спецификациям, способен модифицировать данные последовательности с помощью функционально-эквивалентных замен, а также позволяет многократно исполнять тестовые последовательности на целевом микропроцессоре.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

ОС Windows или ОС на базе ядра GNU/Linux, Java 11.

ОПЫТ ВНЕДРЕНИЯ

MicroTESK разрабатывается с 2007 года. Использовался в российских и международных проектах по разработке современных индустриальных микропроцессоров (в частности, в промышленных проектах по верификации микропроцессоров ARMv8, MIPS64 и RISC-V).

СХЕМА РАБОТЫ

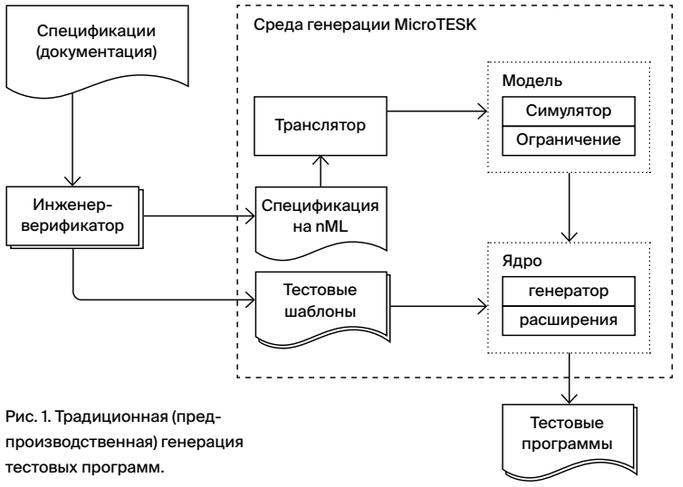


Рис. 1. Традиционная (предпроизводственная) генерация тестовых программ.

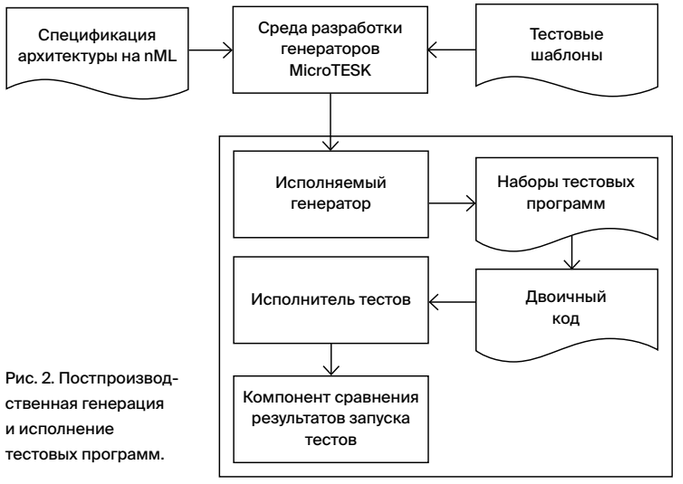


Рис. 2. Постпроизводственная генерация и исполнение тестовых программ.

SAFEC: БЕЗОПАСНЫЙ КОМПИЛЯТОР

Безопасный компилятор SAFEC предотвращает появление уязвимостей в программе при выполнении агрессивных оптимизаций кода (например, в результате использования конструкций, проявляющих неопределённое поведение). Минимально ограничивает выполняемые оптимизации, что позволяет избежать значительного падения производительности по сравнению с полным отключением оптимизаций.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Безопасный компилятор разработан на основе промышленного компилятора GCC и может быть использован в качестве его замены (например, при сборке дистрибутива ОС на базе Linux). Сохраняет эффективность генерируемого кода и даёт готовую безопасную сборку ПО.

SAFEC — это:

- Доработанные оптимизации компилятора для консервативной обработки конструкций с неопределённым поведением и безопасного доопределения семантики программы.
- Принудительная инициализация неинициализированных автоматических переменных.
- Выдача предупреждений компиляции при обнаружении конструкций с неопределенным поведением.
- Добавление динамических проверок для отдельных классов конструкций в целях исключения появления неопределённого поведения по время работы программы.
- Диверсификация генерируемого кода во время выполнения программы.
- Отсутствие необходимости модифицировать исходный код и файлы системы сборки, что позволяет максимально упростить пользование компилятором.
- Распределение функций безопасности по трём классам, определяющим компромисс между безопасностью генерируемого кода и его производительностью. Самый низкий класс — III, самый высокий — I.

SAFEC осуществляет следующие действия:

Согласно III классу:

- Предотвращение возникновения целочисленного переполнения, доступа к объектам через указатель несовместимого типа, разыменования нулевого указателя, замены функций ввода/вывода и работы с памятью стандартной библиотеки на эквивалентные машинные инструкции.

— Обнаружение деления на ноль, некорректности битовых сдвигов, выхода за границы кадра стека, операций загрузки и/или записи в массив за пределами выделенной для него памяти. Обнаружение переменных автоматического класса памяти, которые хранятся на машинном регистре при вызове функций.

Согласно II классу:
— Анализ аргументов битовых сдвигов; избыточности операций работы с памятью; выравнивания данных при работе с векторными инструкциями; адресной арифметики при оптимизациях, связанных с изменением порядка операций обращения к памяти.
— Инициализация не инициализированных явно переменных автоматического класса памяти нулевыми значениями.
— Остановка процесса трансляции при выдаче определённых предупреждений.

Согласно I классу:
— Выполнение уникального распределения в памяти машинного кода функций в процессе динамической компоновки программы или при выполнении статической компиляции. Данная функциональность ранее была реализована в Инструменте диверсификации и в настоящее время интегрирована в SAFEC.
— Добавление в программу машинного кода, вызывающего её аварийный останов при обнаружении ситуаций с неопределённым поведением во время выполнения программы:

1 Операции с целыми или вещественными типами:
  — загрузка небулевого значения в переменную булевого типа;
  — преобразование значений вещественной переменной, приводящее к целочисленному или вещественному переполнению;
  — операция сдвига, в которой величина сдвига отрицательна либо не меньше ширины типа сдвигаемого значения;
  — операция со знаковыми целыми переменными, результат которой не может быть представлен в нужном типе;
  — целочисленное деление или взятие остатка, в котором значение делителя равно нулю.
2 Операции с указателями и массивами:
  — загрузка или запись по некорректно выравненному или нулевому указателю;
  — загрузка или запись в массив по адресу, находящемуся за пределами выделенной для него памяти;
  — передача нулевого указателя в качестве параметра функции, помеченного атрибутом nonnull;
  — выполнение операции адресной арифметики, приводящей к целочисленному переполнению;
  — возврат нулевого значения из функции, объявление которой содержит атрибут returns_nonnull;
  — выделение в автоматической памяти массива переменной длины с некорректным размером.
3 Операции с функциями:
  — вызов по указателю функции, тип которой не соответствует объявленному типу указателя;
  — выход из функции, имеющей возвращаемое значение, без выполнения операции возврата значения;
  — вызов встроенной функции компилятора с недопустимыми значениями аргументов;
  — выполнение точки в программе, отмеченной как недостижимая.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН SAFEC?

— Разработчики ОС.
— Компании, разрабатывающие ПО с высокой степенью надёжности и безопасности.

ОПЫТ ВНЕДРЕНИЯ

Безопасный компилятор поставляется в ряд российских компаний и организаций в дополнение к комплексу ИСП Crusher.

ПОДДЕРЖИВАЕМЫЕ ОС

ОС семейства Linux x86 (32/64), armv7, arm64, RISC-V 64; Windows (MinGW).

SVACE: СТАТИЧЕСКИЙ АНАЛИЗАТОР ИСХОДНОГО КОДА

Svace — необходимый инструмент жизненного цикла разработки безопасного ПО, основной статический анализатор компании Samsung. Обнаруживает более 50 классов критических ошибок в исходном коде. Поддерживает языки C, C++, C#, Java, Kotlin, Go; в бета-версии — язык Python. Включён в Единый реестр российского ПО (№4047). Поставляется с web-интерфейсом просмотра предупреждений Svacer (Svace History Server).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Svace — постоянно развивающийся инновационный продукт, основанный на многолетних исследованиях. Объединяет ключевые качества иностранных аналогов (Synopsis Coverity Static Analysis, Perforce Klocwork Static Code Analysis, Fortify Static Code Analyzer) с уникальным использованием открытых промышленных компиляторов в целях максимальной поддержки новых стандартов языков программирования.

Svace — это:

- Высокое качество анализа:
  - точное представление исходного кода (благодаря интеграции с любой системой сборки);
  - символьное выполнение: полное покрытие всех путей с учетом связей между функциями для поиска сложных ошибок;
  - учёт контекстов вызовов при межпроцедурном анализе, анализ потока данных, анализ чувствительных данных, анализ статистики вызовов;
  - высокий процент истинных срабатываний (60-90%).
- Масштабируемость и высокая скорость:
  - параллельный анализ с использованием всех доступных процессорных ядер;
  - возможность анализировать системы из десятков миллионов строк кода (анализ мобильной ОС Tizen 7 из 57 миллионов строк занимает 7-8 часов основным движком Svace и 9-10 часов всеми движками);
  - поддержка не только полного, но и инкрементального анализа системы (подразумевает быструю повторную проверку недавно изменённого кода).
- Ускоренная кастомизация (конфигурация существующих детекторов, а также написание индивидуальных, доступных только данному заказчику; программный интерфейс для разработки пользовательских детекторов-плагинов);
- Ускоренная адаптация к работе с новым окружением (добавление новых компиляторов в течение 1-2 недель, в сложных случаях — до 2 месяцев).
- Полная совместимость с нормативными документами и требованиями регуляторов (ФСТЭК РФ).
- Возможность использования для реализации обеспечительных мер ГОСТ Р 56939-2016 и процедур «Методики выявления уязвимостей и недекларированных возможностей в программном обеспечении» ФСТЭК России (при необходимости сертификации ПО для использования на территории России).
- Svacer — комплексный пользовательский интерфейс просмотра предупреждений, сервер хранения и управления результатами анализа. Поддерживает многопользовательский режим и различные фильтры данных.

Svacer — это:

- Разметка и отчёты:
  - широкие возможности сравнения и разметки результатов, пользовательские фильтры детекторов, навигация по коду и трассе предупреждения;
  - выгрузка отчётов в формате PDF, CSV, JSON;
  - аннотация результатов работы пользовательскими атрибутами и файлами;
  - возможность разметки результатов в исходном коде в виде специальных комментариев (режим редактирования), ведение истории работы с комментариями;
  - гибкий пользовательский интерфейс с поддержкой вкладок.
- Совместная работа и управление:
  - богатая ролевая модель, позволяющая гибко настраивать права пользователей и организаций по работе с данными;
  - групповые операции работы с проектами, пользователями, разметкой и др.;
  - поддержка разметки различных версий одного проекта, перенос разметки между версиями, ветками и др.;
  - поддержка LDAP для аутентификации пользователей;
  - доступ ко всем данным через API;
  - импорт и экспорт собранных данных: разметки, исходного кода, комментариев, детекторов (в том числе пользовательских).
- Интеграция в CI/CD:
  - интеграция с Visual Studio Code и предоставление интерфейса командной строки для интеграции в типовые CI/CD-процессы.
  - Поддержка работы в контейнерах.
- Поддержка открытого формата SARIF: позволяет импортировать результаты работы других анализаторов, а также экспортировать результаты вместе с разметкой, комментариями, исходным кодом.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН SVACE?

- Компании, нацеленные на разработку ПО с высокой степенью надёжности и безопасности.
- Компании, которые нуждаются в сертификации разрабатываемого ПО.
- Сертификационные лаборатории.

ОПЫТ ВНЕДРЕНИЯ

Svace — основной анализатор Samsung с 2015 года. Применяется для проверки собственного ПО компании на базе ОС Android и исходного кода ОС Tizen, которая используется в смартфонах, информационно-развлекательных системах и бытовой технике Samsung. С 2017 года Svace проверяет все изменения, присланные для рецензирования и включения в ОС Tizen. С 2020 года Svace применяется также в компании Huawei.

В России Svace используется более чем в 100 компаниях и лабораториях, в том числе в ОАО «РусБИТех», АО «Лаборатория Касперского», Postgres Professional, ООО «Код Безопасности», МВП «СВЕМЕЛ» и др.

ПОДДЕРЖИВАЕМЫЕ ПЛАТФОРМЫ И АРХИТЕКТУРЫ

- Платформы, на которых работает анализатор Svace: Linux/x64 (версия ядра 3.10+, версия glibc 2.17+), Linux/ARM 64 (Ubuntu 18.04), Windows (начиная с Windows 7 SP 1 с обновлением KB2533623) и WSL (версий 1 и 2); macOS/x64 (версия 10.10+, язык C# не поддерживается); архитектура x86 (только перехват сборки).
- Архитектуры, для которых анализируется исходный код: для C/C++ — Intel x86/x86-64, ARM/ARM64, MIPS/MIPS64, Power PC/Power PC 64, RISC-V 32/64, SPARC/SPARC64, Hexagon (генерация кода через Clang); Эльбрус, AEON, TriCore, HIDSP, OpenRISC (генерация кода через одну из предыдущих архитектур); для Go — Intel x86-64 на ОС Linux; для C#, Java, Kotlin — платформы, на которых работает анализатор.
- Платформы и архитектуры, на которых работает Svacer: архитектура x86-64; ОС Linux (версия ядра 3.10+, версия glibc 2.17+), ОС Windows (начиная с Windows 10) и WSL (версий 1 и 2); macOS на x86-64 (начиная с версии 10.12 Sierra).

ПОДДЕРЖИВАЕМЫЕ КОМПИЛЯТОРЫ

Для C/C++ (версий до C++20): GCC (GNU Compiler Collection), Clang (LLVM compiler), Microsoft Visual C++ Compiler, RealView/ARM Compilation Tools (ARMCC), Intel C++ Compiler, Elbrus C/C++ Compiler, Wind River Diab Compiler, Keil CA51 Compiler Kit, NEC/Renesas CA850, CC78K0(R) C Compilers, C/C++ Compiler for the Renesas M16C Series and R8C Family, Panasonic MN10300 Series C Compiler, C compiler for Toshiba TLCS-870 and T900 Family, Samsung CalmSHINE16 Compilation Tools, Texas Instruments TMS320C6* Optimizing Compiler, Digital Mars C and C++ Compiler, Green Hills compiler for ARM, TASKING C compiler for TriCore, CEVA Toolbox for CEVA DSP cores, IAR C/C++ Compiler for ARM / Renesas RL78 MCU, CodeWarrior Development Studio for StarCore DSPs, Open Watcom C/C++ compiler, Freescale CodeWarrior, Cadence Tensilica Xtensa C/C++ Compiler.

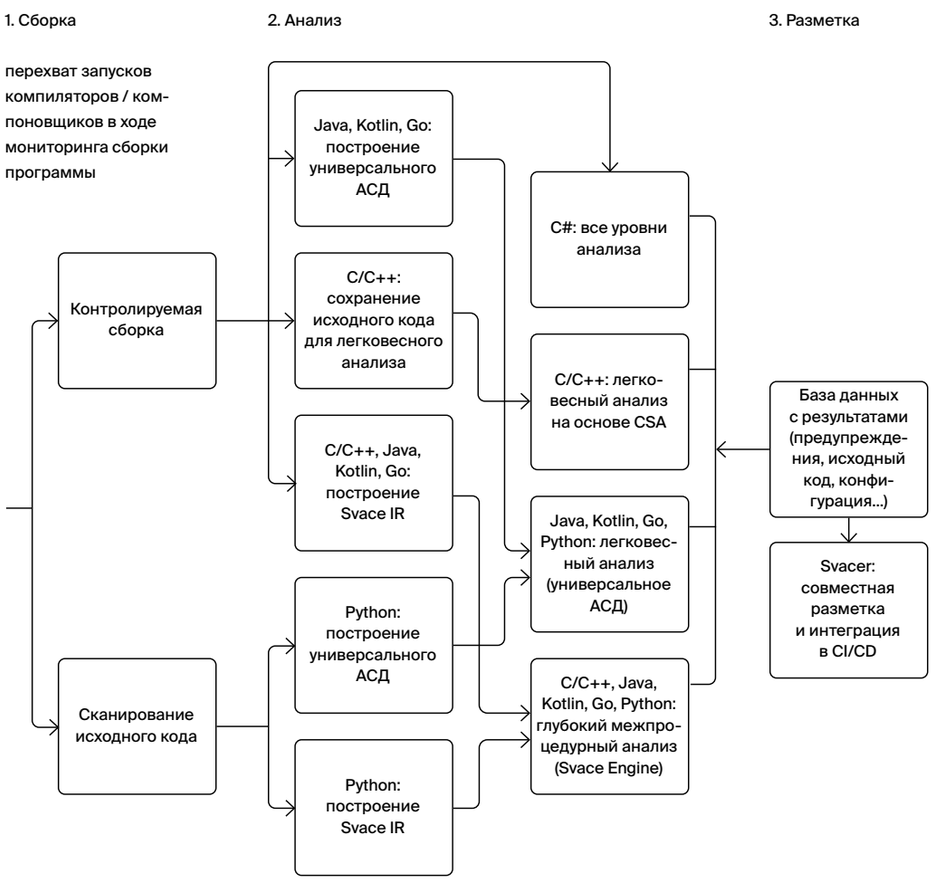
Для C# (версий до C# 11): Roslyn, Mono.

Для Java (версий до Java 17): OpenJDK Javac Compiler, Eclipse Java compiler.

Для Kotlin: Kotlin 1.8.

Для Go: Go 1.21.

СХЕМА РАБОТЫ



сканирование папок с исходным кодом программ на интерпретируемых языках

создание внутренних представлений для легковесного анализа (универсальный АСД) и глубокого анализа (Svace IR)

- легковесный анализ синтаксических деревьев программы;
- межпроцедурный анализ (контекстно-чувствительный и чувствительный к путям на основе символьного выполнения);
- анализ помеченных данных (пользователь может задать источники и приёмники помеченных данных, в том числе помеченные аргументы функций и поля структур).

- подсветка синтаксиса и навигация по коду;
- просмотр предупреждений (с разметкой на истинные и ложные);
- сравнение результатов анализа и сокрытие ложных предупреждений, найденных на предыдущих запусках
- групповые операции работы с проектами, пользователями, разметкой
- доступ ко всем данным через API
- импорт и экспорт собранных данных

TESTOS: ОКРУЖЕНИЕ ДЛЯ ТЕСТИРОВАНИЯ ПО

TestOS — окружение для модульного тестирования программного обеспечения на целевой аппаратуре. Позволяет проводить отработку программного обеспечения для ответственного применения в целях выполнения сертификационных и иных мероприятий на архитектурах AArch64, ARM, PowerPC, MIPS, RISC-V и x86.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

TestOS позволяет заменить такие средства верификации критических систем, как LDRA, являясь более гибким средством с активной поддержкой отечественных изделий.

С помощью TestOS обеспечивается выполнение тестов на целевой аппаратуре и производится генерация отчётов с трассой по каждому тесту, с информацией по составу и статусу прохождения тестового плана и с покрытием кода тестируемой системы как для одного теста, так и для всего тестового плана. Для разработки модульных тестов для функций на языке C (поддерживается C18 с GNU-расширениями GCC и Clang) предоставляется среда генерации заглушек и обёрток с удобным механизмом написания тестовых сценариев. Построение отчётов осуществляется в формате HTML и TXT. Доступна отладка кода на целевом вычислителе как с использованием, так и без использования JTAG.

С использованием плагинов поддерживаются:

- сбор покрытия функций, операторов и ветвей с помощью GCOV и LLVM Coverage;
- сбор покрытия по MC/DC с помощью COVERest;
- проведение статического анализа с помощью статических анализаторов:
  - Clang Tidy;
  - Clang Static Analyzer;
  - Svace.
- динамическое инструментирование кода с помощью санитайзеров LLVM:
  - AddressSanitizer (выявление ошибок работы с памятью);
  - MemorySanitizer (выявление ошибок доступа к неинициализированной памяти);
  - UndefinedBehaviorSanitizer (выявление ошибок арифметики, операций с плавающей точкой и иных проблем, вызванных неопределённым поведением).

СИСТЕМНЫЕ ТРЕБОВАНИЯ

GNU/Linux дистрибутив на архитектуре x86_64 (например, Ubuntu 22.04), а также Apple macOS 10.12 или новее в качестве инструментальной машины.

Целевой вычислитель с ОЗУ не менее 2 МБ на архитектурах:

- AArch64 (Cortex-A53, Cortex-A55);
- ARM (Cortex-A7, Cortex-A9, Cortex-M4), в частности, процессоры i.MX6 или STM32F429;
- PowerPC (e500mc, e500v2, 476FP), в частности, процессоры p1010 или p3041;
- MIPS (MIPS Release 1, MIPS Release 2 / MIPS32, КОМДИВ), в частности, процессор 1892ВМ15АФ;
- RISC-V (RV32 IMA);
- x86 (Intel Prescott и новее).

При необходимости окружение адаптируется к оборудованию заказчика.

ОПЫТ ВНЕДРЕНИЯ

TestOS разрабатывается с 2019 года. Успешно применяется для модульного тестирования программного обеспечения для аэрокосмической отрасли.

БЛЕСНА: ИНСТРУМЕНТ ДИНАМИЧЕСКОГО АНАЛИЗА ПОМЕЧЕННЫХ ДАННЫХ

Блесна — специализированный инструмент, предназначенный для поиска утечек в памяти чувствительных данных, таких как пользовательские пароли и ключи шифрования. Применим для анализа всего программного стека от загрузчика до прикладного ПО. Включён в Единый реестр российского ПО (№13095).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Блесна проводит точный динамический анализ помеченных данных, проводимый комплексно, для объекта оценки и его программной среды функционирования.

Возможности инструмента опираются на многолетний опыт разработчиков компиляторов и специалистов по информационной безопасности. В отличие от аналогичных научно-исследовательских технологий в области анализа бинарного кода инструмент доработан до промышленного использования.

Блесна — это:

- Автоматизация экспертизы исполняемого кода в части утечек данных с минимальными требованиями к квалификации пользователей.
- Упрощённая подготовка к анализу по типовым сценариям.
- Для выявленной утечки предъявляется цепочка команд, реализующая утечку с указанием необходимой диагностики: имён функций и модулей, стека вызовов.
- Полносистемный анализ бинарного кода: исследуется весь стек развёрнутого ПО, что позволяет выявлять утечки данных, проходящие через границы виртуальных адресных пространств процессов.
- Глубокий анализ:
  - для анализа достаточно наличия лишь исполняемого бинарного кода и описания сигнатур анализируемых функций;
  - точный анализ потоков данных, учитывающий особенности аппаратуры (конвейер команд, прерывания, трансляция виртуальных адресов, DMA).
- Высокая производительность:
  - параллельный анализ с высокими показателями масштабируемости на многоядерных рабочих станциях;
  - возможность анализа длительных сценариев работы анализируемой системы.

ДЛЯ КОГО ПРЕДНАЗНАЧЕНА БЛЕСНА?

- Государственные учреждения, в обязанности которых входит экспертиза безопасности программного обеспечения.
- Отечественные компании-разработчики безопасного программного обеспечения.
- Испытательные лаборатории и органы по сертификации

УСЛУГИ С ИСПОЛЬЗОВАНИЕМ БЛЕСНЫ

- Исследования программного обеспечения при проведении сертификации на высокие уровни доверия.
- Целевое обучение и повышение квалификации.

ПОДДЕРЖИВАЕМЫЕ ПЛАТФОРМЫ И АРХИТЕКТУРЫ

- Системные требования инструмента Блесна: ОС Linux x86-64, ОЗУ не менее 16 Гбайт, рекомендуется не менее 2 Тбайт дискового пространства.
- Целевые процессорные архитектуры: x86/x86-64, ARM v7.
- Целевые ОС: семейство Windows, семейство Linux, поддерживается возможность работы с неопознанной ОС и с кодом, работающим вне ОС.

ИНСТРУМЕНТ ДИВЕРСИФИКАЦИИ: КОМПЛЕКС ЗАЩИТЫ ОТ ЭКСПЛУАТАЦИИ УЯЗВИМОСТЕЙ

Инструмент диверсификации — комплекс технологий по противодействию массовой эксплуатации уязвимостей, возникающих в результате ошибок или закладок. Если злоумышленник смог атаковать одно из устройств с одинаковым ПО, остальные останутся под защитой благодаря изменениям, внесённым в код.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Инструмент диверсификации защищает систему от массовой эксплуатации уязвимостей с помощью различных методов диверсификации кода и позволяет собирать код полного дистрибутива ОС.

Инструмент диверсификации — это:

- Тонкая настройка баланса между степенью запутывания и уровнем производительности (при применении с целью защиты от обратного анализа). Минимальное замедление работы — в 1,2 раза, максимальное — в 8 раз.
- Полная автоматизация (не требуется специальная подготовка исходного кода программы и дополнительные усилия со стороны билд-инженеров заказчика).
- Использование набора открытых компиляторов GCC, который позволяет корректно собирать код полного дистрибутива ОС.
- Использование оригинального метода обеспечения целостности потока управления (CFI), который успешно противодействует большинству атак с повторным использованием кода (ROP, JOP, ret-to-plt и др.). Реализация метода CFI на базе компилятора GCC показала среднее замедление на наборе тестов SPEC CPU2006 около 2%, что заметно ниже, чем у традиционных методов.
- Два метода диверсификации:
  - Динамическая диверсификация кода при запуске программы. Применяется, когда заказчику обязательно нужен один и тот же код на всех устройствах (например, из-за обязательной сертификации). Этот метод позволяет перемещать до 98% кода с небольшим увеличением его объёма и ухудшением производительности примерно на 1,5%. Преимущества Инструмента диверсификации по сравнению с аналогичными продуктами:
    - перемешивание до функции (в отличие от технологий ASLR и Pagerando, которые перемещают только крупные блоки кода);
    - перемешивание функций во всей системе, кроме ядра, а также отсутствие конфликта с антивирусами (преимущества перед аналогичной технологией Selfrando, разработанной для Tor Browser);
  - Статическая диверсификация кода. Каждый раз при компиляции в зависимости от заданного ключа получается новый исполняемый файл. Преимущества данного метода:
    - не увеличивается объём бинарного кода (в частности, важно для интернета вещей);
    - ухудшение производительности стремится к нулю;
    - благодаря работе внутри компилятора, а не постфактум в компоновщике, можно применять расширенный набор диверсифицирующих преобразований и более гибко его настраивать;
    - метод обеспечения целостности потока управления (CFI).
- Бесконфликтное совмещение с другими средствами защиты ПО (в том числе с системным механизмом ASLR).

ДЛЯ КОГО ПРЕДНАЗНАЧЕН ИНСТРУМЕНТ ДИВЕРСИФИКАЦИИ

- Разработчики специализированных дистрибутивов операционных систем.
- Разработчики прикладного ПО.

ОПЫТ ВНЕДРЕНИЯ

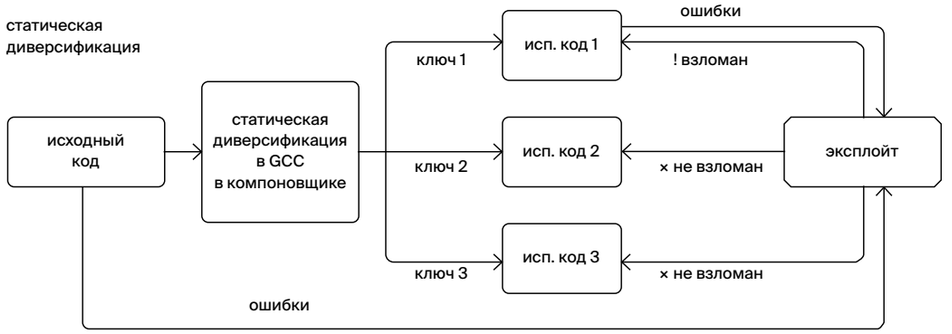
Инструмент диверсификации внедрен в ОС «Циркон», которую используют МИД и Пограничная служба ФСБ России. В настоящее время Инструмент диверсификации реализуется как часть уровня 1 безопасного компилятора SAFEC и поставляется вместе с ним.

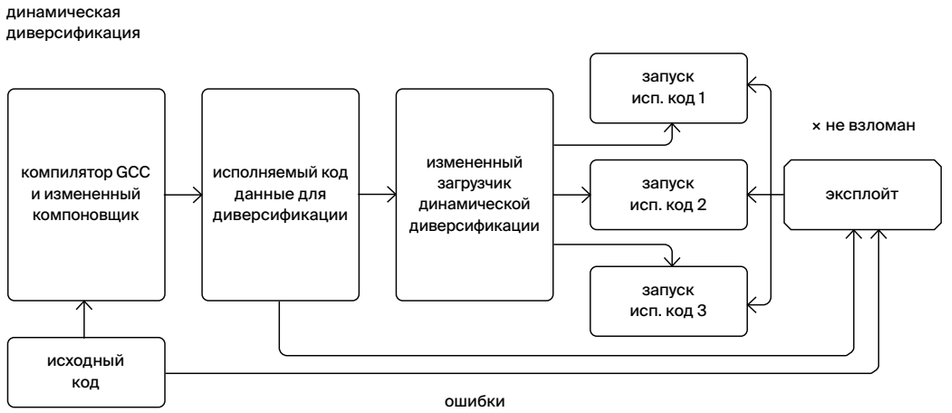
СИСТЕМНЫЕ ТРЕБОВАНИЯ

Инструмент диверсификации — универсальный продукт, который можно адаптировать под любые системные требования. В настоящее время основная версия работает в ОС на базе ядра Linux (начиная с версии 2.6) с поддержкой архитектуры Intel x86/x86-64.

СХЕМА РАБОТЫ

обычная сборка





ПЛАТФОРМА ДЛЯ АНАЛИЗА ПРОГРАММ НА ОСНОВЕ ЭМУЛЯТОРА QEMU

Платформа ИСП РАН для анализа программ построена на базе открытого эмулятора QEMU, который используется при необходимости кроссплатформенной разработки. Поддерживает механизмы обратной отладки и интроспекции, а также режим полносистемной эмуляции для отладки низкоуровневого ПО.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

QEMU поддерживает более 10 архитектур процессоров (i386 и x86-64, ARM и Thumb, MIPS, PowerPC и др.). Реализует отладку по сетевому протоколу GDB RSP и совместим с IDA Pro, GDB и средами разработки. В режиме полносистемной эмуляции подходит для отладки низкоуровневого ПО, такого как загрузчик и ОС. Исходный код QEMU систематически проверяется двумя статическими анализаторами (Coverity и Svace), что делает анализ потенциально вредоносного ПО в эмуляторе более безопасным. Эмулятор с поддержкой обратной отладки и интроспекции доступен на GitHub (https://github.com/ispras/swat), как и набор инструментов автоматизации (https://github.com/ispras/qdt, https://github.com/ispras/i3s).

Платформа ИСП РАН на основе QEMU — это:

- Запись и воспроизведение работы виртуальной машины:
  - При каждом воспроизведении виртуальная машина ведёт себя одинаково и точно так же, как при записи. Все воздействия извне зафиксированы и повторяются самим эмулятором. Это упрощает отладку ошибок, связанных с параллельной работой (состояние гонки, взаимные блокировки).
  - На базе воспроизведения реализована GDB-совместимая обратная отладка, которая заключается в откате к предыдущим снимкам состояния виртуальной машины и поиске предпоследнего срабатывания точки останова или предыдущей инструкции.
  - Записывается минимум информации, что позволяет вести длительную запись, необходимую для отладки редко повторяющихся ошибок.
  - Низкое относительное замедление, вносимое записью, позволяет контролировать ПО, взаимодействующее с удалённой неконтролируемой системой в реальном времени.

- Получение высокоуровневой информации о работе гостевой ОС (интроспекция ВМ) без внесения каких-либо изменений в ядро ОС или установки программ мониторинга:
  - Возможность получить последовательность совершаемых системных вызовов, обращений к именованным функциям в динамических библиотеках, список работающих процессов, список открытых файлов и загруженных в память модулей.
  - Поддержка любого образа виртуальной машины на основе Linux, в том числе — образов встраиваемого ПО различных устройств.
  - Отладка с помощью встроенного в эмулятор сервера WinDbg, что позволяет отображать информацию о гостевом ПО в терминах абстракций ядра Windows. При этом не требуется включение отладочного режима работы гостевой ОС.
- Ускорение разработки расширений для QEMU:
  - Сокращение времени на подготовку средств динамического анализа для образцов кода, требующих специализированной аппаратуры.
  - Автоматизированное добавление процессорных архитектур с использованием генератора декодеров машинных команд и Си-подобного языка описания семантики инструкций.
  - Система автоматического первичного тестирования виртуальной машины. Для работы системы требуются только утилиты GNU Binutils и компилятор языка Си.
  - Автоматизированная разработка моделей устройств.
  - Генерация виртуальной машины (в форме исходного кода модуля QEMU) как из существующих, так и из новых устройств по описанию на языке Python с использованием графического интерфейса пользователя со схематичным изображением системы.
  - API для автоматизации процесса отладки на языке Python по протоколу GDB RSP: отладка гостевого кода, кода эмулятора и обоих одновременно.
- Удобство практического использования:
  - Свободное расширение возможностей QEMU благодаря открытому исходному коду и собственным инструментам ускоренной разработки ИСП РАН.
  - Анализ бинарного кода без внедрения программ в гостевую систему.
  - Модульная структура механизма интроспекции с возможностью расширения за счёт новых плагинов.
  - Удобное API для самостоятельной разработки плагинов интроспекции.
  - Возможность адаптации под конкретные нужды пользователя.
  - Поддержка актуальных версий QEMU с новой периферией и процессорными ядрами.

ДЛЯ КОГО ПРЕДНАЗНАЧЕНА ПЛАТФОРМА НА БАЗЕ QEMU?

- Разработчики загрузчиков, драйверов, ОС и другого системного ПО.
- DevOps-команды (воспроизводимость ошибок, кросс-разработка, масштабирование тестирования в облачной среде).
- Аналитики потенциально вредоносного ПО.
- Специалисты по сертификации ПО.

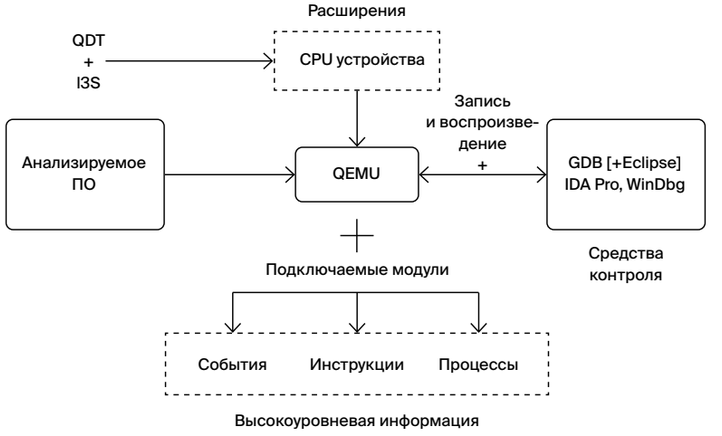
ПОДДЕРЖИВАЕМЫЕ ГОСТЕВЫЕ СРЕДЫ

Эмулируемые платформы: i386, x86-64, ARM, MIPS, PowerPC и другие.
Гостевые системы, поддерживаемые интроспекцией: Windows XP (x86), Windows 10 (x86-64) и Linux 2.x-5.x на платформах x86, x86-64, ARM, AArch64.

ОПЫТ ВНЕДРЕНИЯ

Реализованный механизм воспроизведения принят мировым сообществом разработчиков QEMU и включен в версию 3.1.

СХЕМА РАБОТЫ



ИСП CRUSHER: КОМПЛЕКС ДИНАМИЧЕСКОГО И СТАТИЧЕСКОГО АНАЛИЗА БИНАРНОГО КОДА

ИСП Crusher — программный комплекс, комбинирующий несколько методов динамического и статического анализа, в частности, фаззинг (ИСП Fuzzer) и символьное выполнение (в качестве одного из движков может выступать Sydr). В ближайшее время в комплекс планируется включить ещё одну технологию ИСП РАН: BinSide. ИСП Crusher позволяет построить процесс разработки в соответствии с ГОСТ Р 56939-2016 и «Методикой выявления уязвимостей и недекларированных возможностей в программном обеспечении» ФСТЭК России. Включён в Единый реестр российского ПО (№10468).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

В основе комплекса ИСП Crusher находится ИСП Fuzzer — инструмент проведения фаззинг-тестирования, необходимый на всех этапах разработки, тестирования и эксплуатации ПО. Обнаруживает ошибки как при наличии, так и при отсутствии исходного кода. Решает те же задачи, что и мировые аналоги (Synopsys Codenomicon, beSTORM, Peach Fuzzer), однако более удобен для российских компаний в условиях процесса импортозамещения.

ИСП Fuzzer — это:

Фаззинг широкого класса ПО:
- пользовательские приложения, ядро и библиотеки;
- приложения на различных языках программирования: C/C++, Java, Python, C#;
- фаззинг нейронных сетей. Выявляет случаи ошибочных предсказаний нейронной сети при искажении корректно классифицируемых входных данных, опираясь на анализ карты активации нейронов во время работы нейросетей. Это позволяет повысить качество и безопасность систем ИИ; благодаря такому функционалу можно:
  - находить ошибки в работе сетей для ситуаций, которые изначально не попали в набор данных для обучения;
  - находить возможные закладки и НДВ.
- фаззинг-тестирование через различные источники входных данных: файл, аргументы командной строки, стандартный поток ввода, аргументы переменных окружений, сеть, прямая запись в память;
- возможность проведения анализа серверного и клиентского ПО, работающего по протоколам с состояниями и без состояний;
- фаззинг протоколов с использованием модифицированного клиента: отпадает необходимость написания клиента или его спецификации с нуля для фаззинга сервера (поддерживается и зеркальная схема – с модифицированным сервером);
- широкие возможности для фаззинга ПО встраиваемых устройств с использованием частичной эмуляции и символьного исполнения;
- фаззинг браузеров: управление браузером — через Selenium, обратная связь по покрытию — через Frida.
- фаззинг ПО, требующего изоляции, в докер-режиме — каждый экземпляр фаззера работает в отдельном докер-контейнере;
- фаззинг приложений в rootfs в chroot-режиме.

Фаззинг на больших мощностях:
- поддержка многопоточного анализа как на одной машине, так и на распределённых;
- возможность распределения корпуса входных данных между процессами фаззеров для повышения эффективности их работы;
- поддержка дифференциального фаззинга.

Поддержка большого набора типов инструментации:
- статическая (в основном, для C/C++): с помощью GCC/LLVM;
- статическая инструментация байт-кода Python;
- динамическая (в основном, для ELF, PE): DynamoRIO, Qemu (user-mode), TinyInst;
- на основе частичной эмуляции;
- на основе полносистемной эмуляции;
- с использованием Nyx-снапшотов и снапшот-API;
- Java-приложений;
- C#-приложений;
- удалённая инструментация (позволяет выполнять фаззинг приложения, работающего на удалённом устройстве).

Возможность интеграции с рядом необходимых инструментов жизненного цикла разработки безопасного ПО, созданных в ИСП РАН:
- использование инструмента динамического символьного выполнения Sydr для повышения эффективности фаззинг-тестирования;

— возможность получать входные данные, на которых проявляются ошибки, размеченные инструментом статического анализа BinSide в автоматизированном режиме;
— отображение трассы последовательности функций, приводящих к аварийному завершению, в интерфейсе статического анализатора Svace;
— использование генератора данных по формальным грамматикам ANTLR для формирования корпуса входных данных.

Интеграция с другими инструментами динамического анализа:
— со сторонними фаззерами: позволяет запускать в рамках одной фаззинг-сессии набор различных фаззеров, между которыми обеспечена синхронизация, что увеличивает эффективность тестирования;
— с инструментами динамического символьного выполнения SymCC и Angr: позволяет получать новые входные данные для увеличения покрытия кода целевого ПО;
— совместная работа с дизассемблером IDA PRO (сохранение покрытия для плагина Lighthouse, которое отображает покрытые базовые блоки в ПО, а также вывод процента покрытых базовых блоков);
— использование фаззера Radamsa для генерации новых данных.

Дополнительный анализ полученных входных данных:
— оценка критичности найденных аварийных завершений;
— возможность запуска систем динамического анализа на новых входных данных: Valgrind, DrMemory, QASan;
— создание профиля покрытия по исходному коду.

Широкие возможности по встраиванию пользовательских расширений:
— возможность добавления пользовательских обработчиков, которые будут автоматически запускаться на новых входных данных;
— возможность добавления пользовательских мутационных преобразований (для генерации новых входных данных и увеличения эффективности тестирования);
— наличие модулей пред- и постобработки входных данных для осуществления константных преобразований над данными перед их отправкой в анализируемое ПО;
— поддержка пользовательских плагинов отправки данных по сети (плагины позволяют осуществлять взаимодействие с клиентским или серверным ПО и отправлять мутированные данные);
— поддержка пользовательских скриптов на языке Python для модификации опций (позволяет избежать конфликтов при одновременном запуске множества процессов фаззинга);

— поддержка пользовательских плагинов на языке Python для управления окружением запуска целевого ПО (позволяет сохранять идентичное окружение на каждом запуске);
— поддержка пользовательских плагинов инструментации (позволяет задавать произвольные правила классификации входных данных на основе поведения целевого ПО: определение нормального и аварийного завершения, зависания);
— возможность описывать сценарии для фаззинга ПО с пользовательским интерфейсом.

Лёгкая расширяемость и добавление новых методов в рамках существующей инфраструктуры; оперативная адаптация под новые задачи.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН ИСП CRUSHER?

— Компании, нацеленные на разработку ПО с высоким уровнем надёжности и безопасности.
— Компании, отвечающие за аудит или сертификацию ПО.

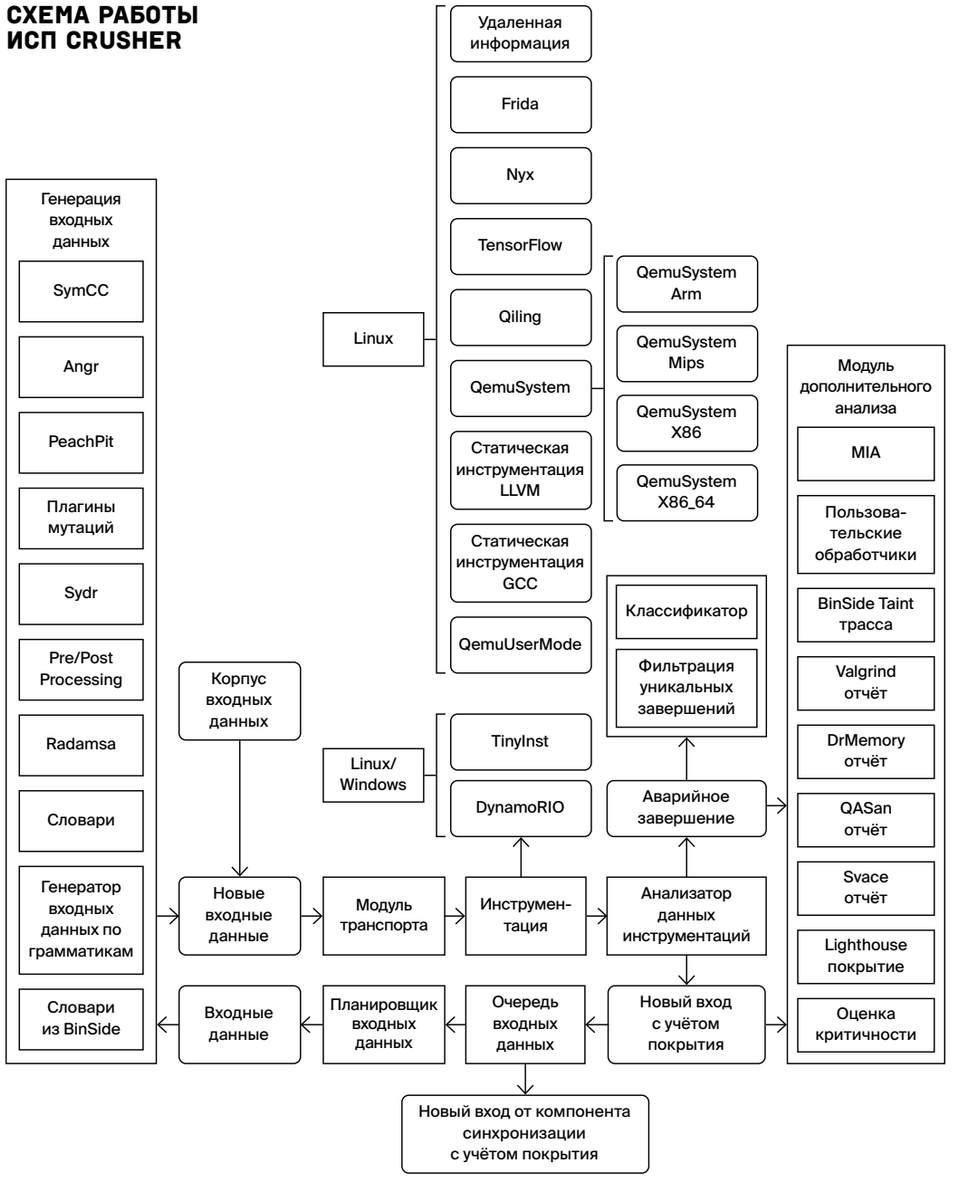
СИСТЕМНЫЕ ТРЕБОВАНИЯ

Поддержка фаззинга на Linux и Windows. Поддержка фаззинга ПО для архитектур x86_64, ARM, MIPS. Возможность проводить фаззинг-тестирование встроенных устройств (контроллеры, устройства интернета вещей), а также сервисов и COM-объектов ОС Windows.

ОПЫТ ВНЕДРЕНИЯ

ИСП Crusher в разной комплектации используется более чем в 70 компаниях и лабораториях, в том числе в ОАО «РусБИТех», Postgres Professional, ООО «Код Безопасности», МВП «СВЕМЕЛ» и др.

СХЕМА РАБОТЫ ИСП CRUSHER



BINSIDE: СТАТИЧЕСКИЙ АНАЛИЗАТОР БИНАРНОГО КОДА

BinSide — платформа обнаружения дефектов в программе методами статического анализа исполняемого кода. Необходима, когда нет доступа к исходному коду (например, при анализе закрытых библиотек).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

BinSide — платформа для анализа исполняемого кода, разрабатываемая на основе фреймворка BinNavi. Исполняемый файл анализируется в представлении IDA PRO или Ghidra. BinSide ищет дефекты, клоны кода, автоматизирует работу аналитика, уточняет поверхность атаки и оптимизирует динамическое тестирование.

Возможности ядра BinSide:

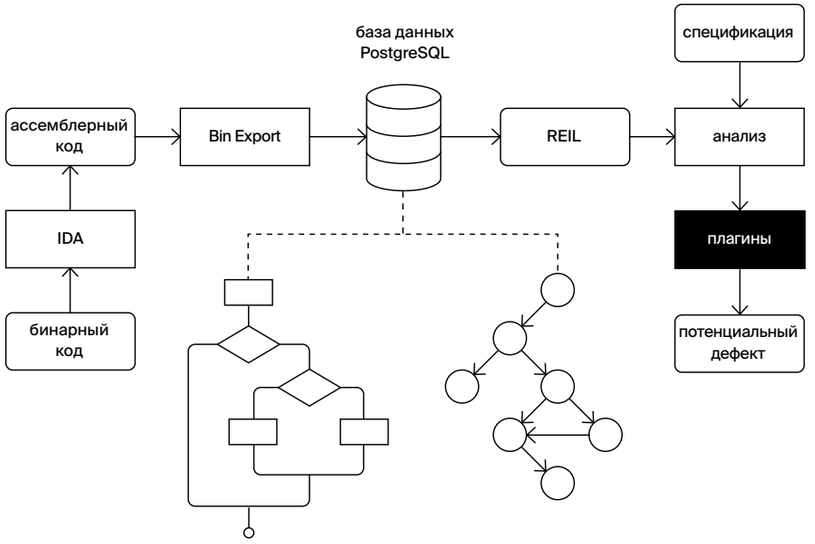
- Лёгкая расширяемость:
  - детекторы ошибок реализованы в виде подключаемых модулей;
  - используется промежуточное представление REIL из 17 инструкций без побочных эффектов (каждая ассемблерная инструкция транслируется в набор из REIL-инструкций);
  - анализ может быть уточнён аналитиком вручную.
- Поддержка анализа исполняемых файлов и библиотек для архитектур x86-64, ARM и MIPS, в том числе драйверов.
- Поиск дефектов следующих типов:
  - CWE-121 (Stack-based Buffer Overflow);
  - CWE-122 (Heap-based Buffer Overflow);
  - CWE-134 (Format String Vulnerability);
  - CWE-415 (Double Free);
  - CWE-416 (Use-After-Free);
  - CWE-77 (Command Injection).
- Выполнение следующих задач:
  - анализ потока данных и управления: восстановление значений и указателей, распространение помеченных данных, определение возможных состояний кучи, определение вычислимых ребёр графа потока управления;
  - межпроцедурный поиск дефектов: поиск дефектов выполняется на основе результатов внутрипроцедурного анализа потока данных и управления, результатов динамического анализа и ручной разметки кода аналитиком. Это особенно полезно при анализе комплексного ПО и встраиваемых систем;
  - анализ всех путей, независимо от покрытия кода.

- Взаимодействие с технологиями ИСП РАН:
  - с инструментом Svacer (при наличии исходного кода);
  - с инструментом LibraryIdentifier (для поиска клонов кода, например, для идентификации библиотек, с кодом которых был скомпонован исполняемый файл);
  - с инструментом Crusher.
- Анализ операционных систем:
- определение заимствования кода в ПО из ОС с открытым исходным кодом;
- установление зависимостей между компонентами ОС и внутри компонентов;
- статический анализ исходного и бинарного кода ОС;
- определение защит исполняемого кода в компонентах ОС;
- определение покрытия кода в компонентах ОС unit-тестами.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН BINSIDE?

- компании, которые нуждаются в тщательной проверке стороннего ПО, в том числе при отсутствии доступа к исходному коду;
- разработчики, которым требуется повысить качество работы инструментов динамического анализа за счёт дополнительных данных, полученных с помощью статического анализа;
- специалисты по обратной инженерии;
- компании, отвечающие за аудит или сертификацию ПО.

СХЕМА РАБОТЫ



CASR: ИНСТРУМЕНТ ФОРМИРОВАНИЯ ОТЧЁТОВ ОБ ОШИБКАХ

GitHub → https://github.com/ispras/casr

Casr — это инструмент, позволяющий автоматически формировать отчёты об аварийных завершениях, возникающих во время эксплуатации и тестирования ПО на ОС Linux. В отчётах содержатся сведения о степени критичности аварийного завершения, а также дополнительная информация, которая помогает установить его причины. CASR — это проект с открытым исходным кодом (https://github.com/ispras/casr).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

CASR позволяет получать отчёты об ошибках разными способами (coredump, GDB, Asan, Ubsan) и обрабатывать исключения от разных языков программирования (Rust, Go, Java, Python). Предоставляет возможность автоматизации анализа результатов фаззинга с выгрузкой в системы управления уязвимостями.

Casr — это:

- Обнаружение критичных аварийных завершений, которые могут привести к перехвату потока управления.
- Классификация аварийных завершений, которая проводится в зависимости от состояния программы на момент завершения (перезапись адреса возврата из функции, разыменование нулевого указателя и др.). Далее аварийные завершения группируются по степени критичности: эксплуатируемые, потенциально эксплуатируемые, отказ в обслуживании.
- Развёрнутый отчёт об ошибке, который содержит информацию о степени критичности аварийного завершения, а также дополнительные данные (версии ОС и пакета, строка запуска программы, стек вызовов, открытые файлы и сетевые соединения, состояние регистров и др.).
- Дедупликация и кластеризация аварийных завершений на основе стека вызовов. Кластеры потенциально содержат схожие отчёты, которые описывают одну и ту же уязвимость.
- Интеграция с современными фаззерами Sydr/AFL++/LibFuzzer (go-fuzz, Atheris, Jazzer).
- Библиотека libcasr для написания собственных инструментов анализа.
- Выгрузка результатов в систему управления уязвимостей DefectDojo, которая позволяет удобно встроить процесс разбора результатов фаззинга в CI.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН CASR?

- Компании, которым необходимо получать информацию об ошибках, возникающих у пользователей, в целях разработки ПО с высокой степенью надёжности и безопасности.
- Компании, нуждающиеся в сертификации разрабатываемого ПО.
- Испытательные лаборатории.

ОПЫТ ВНЕДРЕНИЯ

CASR используется для анализа аварийных завершений в инструменте Sydr (ИСП РАН).

СИСТЕМНЫЕ ТРЕБОВАНИЯ

Поддержка ОС семейства Linux x86 (32/64), aarch64, RISCV64.

СХЕМА РАБОТЫ

NATCH: ИНСТРУМЕНТ ОПРЕДЕЛЕНИЯ ПОВЕРХНОСТИ АТАКИ

Natch — это инструмент для определения поверхности атаки, основанный на полносистемном эмуляторе QEMU. Использует технологии анализа помеченных данных, интроспекции виртуальных машин и детерминированного воспроизведения. Включён в Единый реестр российского ПО (№13673).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Основная функция Natch — получение поверхности атаки, то есть поиск исполняемых файлов, динамических библиотек, а также функций, отвечающих за обработку входных данных (файлов, сетевых пакетов) во время выполнения задачи. Собранные данные визуализируются в графическом интерфейсе SNatch, входящем в поставку инструмента.

Natch построен на базе полносистемного эмулятора QEMU, что позволяет анализировать все компоненты системы, включая ядро ОС и драйверы. Важное преимущество — объединение ключевых возможностей аналогов в одном инструменте.

Построение поверхности атаки может быть встроено в CI/CD для интеграционного и системного тестирования, что позволит повысить эффективность технологий функционального тестирования и фаззинга в жизненном цикле безопасного ПО.

ВОЗМОЖНОСТИ ИНСТРУМЕНТА NATCH

- сбор поверхности атаки: процессов, модулей и функций, которые обрабатывали помеченные данные во время выполнения тестового сценария;
- определение открытых портов, файлов, сокетов и потоков данных через них;
- поддержка анализа для языков C, C++, Go, Python;
- автоматическое скачивание отладочной информации для ядра и системных модулей;
- извлечение отладочной информации из DWARF;
- построение графа потоков помеченных данных через процессы и модули по всей системе;

- сбор лога сетевых пакетов в формате pcap;
- сбор покрытия бинарного кода;
- сбор корпуса данных для фаззинга выбранных функций (для аргументов простых типов).

СХЕМА ПРИМЕНЕНИЯ NATCH

- Natch записывает сценарий работы аналитика в виртуальной машине;
- аналитик помечает входящий сетевой трафик или обращения к определённым файлам;
- Natch воспроизводит записанный сценарий работы и отслеживает распространение помеченных данных, формирует архив логов операций с помеченными данными и системных событий;
- аналитик загружает полученный архив в графический интерфейс SNatch для изучения.

ВОЗМОЖНОСТИ ГРАФИЧЕСКОГО ИНТЕРФЕЙСА SNATCH

- Граф взаимодействия процессов, работавших с помеченными данными. Позволяет отслеживать потоки данных во времени и упорядочивать элементы схемы удобным для аналитика образом.
- Временная диаграмма процессов, работавших в системе.
- Стеки вызовов помеченных функций с разделением по процессам.
- Стеки вызовов для функций из скриптов при наличии их в объекте оценки.
- Flame-диаграмма процессов с цветовым дифференцированием помеченных и не помеченных функций.
- Просмотр дерева всех процессов, работавших в системе, с возможностью фильтрации только помеченных процессов.
- Просмотр ресурсов, использованных процессами во время работы.
- Просмотр операций чтения и записи для всех файлов и сокетов.
- Подсветка процессов, взаимодействующих с помеченными данными или имеющих повышенные привилегии (например, запущенные с правами root).
- Генерация аннотаций для функций для инструмента Futag.
- Передача отфильтрованного сетевого трафика в Wireshark.
- Удобный поиск по графам, а также сохранение истории перемещений.
- Генерация отчёта с основными аналитиками в формате PDF.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН NATCH?

- Отечественные компании-разработчики безопасного программного обеспечения.
- Испытательные лаборатории и органы по сертификации.

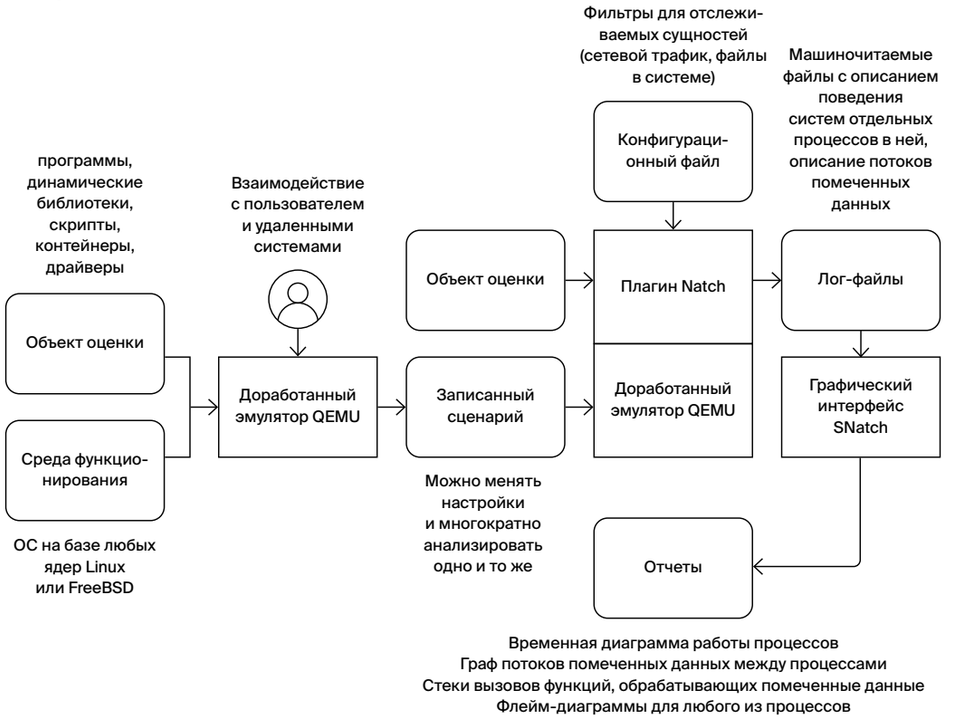
ПОДДЕРЖИВАЕМЫЕ ПЛАТФОРМЫ И АРХИТЕКТУРЫ

- Системные требования инструмента Natch: ОС Linux x86-64, ОЗУ не менее 16 Гбайт, рекомендуется не менее 200 Гбайт дискового пространства.
- Целевые процессорные архитектуры: x86-64.
- Целевые ОС: семейство Linux (любые версии ядер), Windows 7-10, FreeBSD последних версий.

ОПЫТ ВНЕДРЕНИЯ

Инструмент для определения поверхности атаки Natch распространяется в сообществе для пробного использования и тестирования, а также применяется для целевого обучения и повышения квалификации.

СХЕМА РАБОТЫ



SYDR + SYDR-FUZZ: КОМПЛЕКС ГИБРИДНОГО ФАЗЗИНГА И ДИНАМИЧЕСКОГО АНАЛИЗА

GitHub →
https://github.com/ispras/oss-sydr-fuzz

Sydr — инструмент автоматической генерации тестов для сложных программных систем с целью увеличения покрытия кода и обнаружения ошибок. Строит математическую модель программы, позволяя фаззеру открывать новые пути выполнения, которые сложно обнаружить классическими методами генетических мутаций. Разработанные методы развивают технологию символьного выполнения, представленную в созданных ранее в ИСП РАН инструментах Avalanche и Anxiety. Sydr-fuzz — инструмент динамического анализа программ для безопасного цикла разработки ПО, который сочетает в себе возможности инструмента динамического символьного выполнения Sydr и современных фаззеров (libFuzzer и AFL++).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

В отличие от аналогичных открытых инструментов, Sydr проверяет результаты своей работы на корректность и определяет, действительно ли сгенерированные входные данные приводят к инвертированию целевых переходов. Sydr-fuzz предоставляет удобный пайплайн фаззинга:

- гибридный фаззинг с помощью Sydr и libFuzzer/AFL++: sydr-fuzz run;
- минимизация корпуса: sydr-fuzz cmin;
- поиск ошибок (выхода за границу массива, целочисленного переполнения, деления на ноль и других): sydr-fuzz security;
- сбор покрытия: sydr-fuzz cov-report;
- дедупликация, кластеризация и оценка критичности аварийных завершений с использованием Casr: sydr-fuzz casr.

Sydr + Sydr-fuzz — это:

- Гибридный фаззинг Sydr и libFuzzer/AFL++. Возможность гибридного фаззинга проектов на C/C++, Rust, Go.
- Фаззинг с помошью Atheris и Jazzer, полный пайплайн динамического анализа Python/Java проектов.
- Эффективность на уровне мировых аналогов: регулярный бенчмаркинг фаззинга (https://sydr-fuzz.github.io/fuzzbench/).
- Репозиторий с настроенными проектами под фаззинг: 70+ проектов (500 фаззинг-целей) в OSS-Sydr-Fuzz (https://github.com/ispras/oss-sydr-fuzz).
- Достижения: Sydr-fuzz нашел 145 новых ошибок в 28 проектах с открытым исходным кодом (https://github.com/ispras/oss-sydr-fuzz/blob/master/TROPHIES.md), 25 ошибок найдено с помощью предикатов безопасности.
- Инвертирование на конкретном пути выполнения всех условных переходов, которые зависят от входных данных, для открытия новых путей выполнения программы. Реализована возможность параллельного инвертирования.
- Предикаты безопасности. Генерация входных данных, приводящих к проявлению дефекта (деление на ноль, разыменование нулевого указателя, выход за границы массива, целочисленное переполнение и др.).
- Инвертирование косвенных переходов (switch statement). Разработан алгоритм определения таблиц и переходов по вычисляемым адресам.
- Слайсинг формул. Удаление избыточных формул из предиката пути, которые не влияют на инвертируемый условный переход. Решает проблему недостаточной помеченности, а также ускоряет обработку запросов SMT-решателем.
- Обработка символьных указателей, зависящих от входных данных. Это позволяет находить критичные ошибки, когда они возникают из-за влияния пользовательского ввода на вычисление индекса массива. Поддержка символьных указателей требует дополнительного моделирования, которым часто пренебрегают аналогичные инструменты.

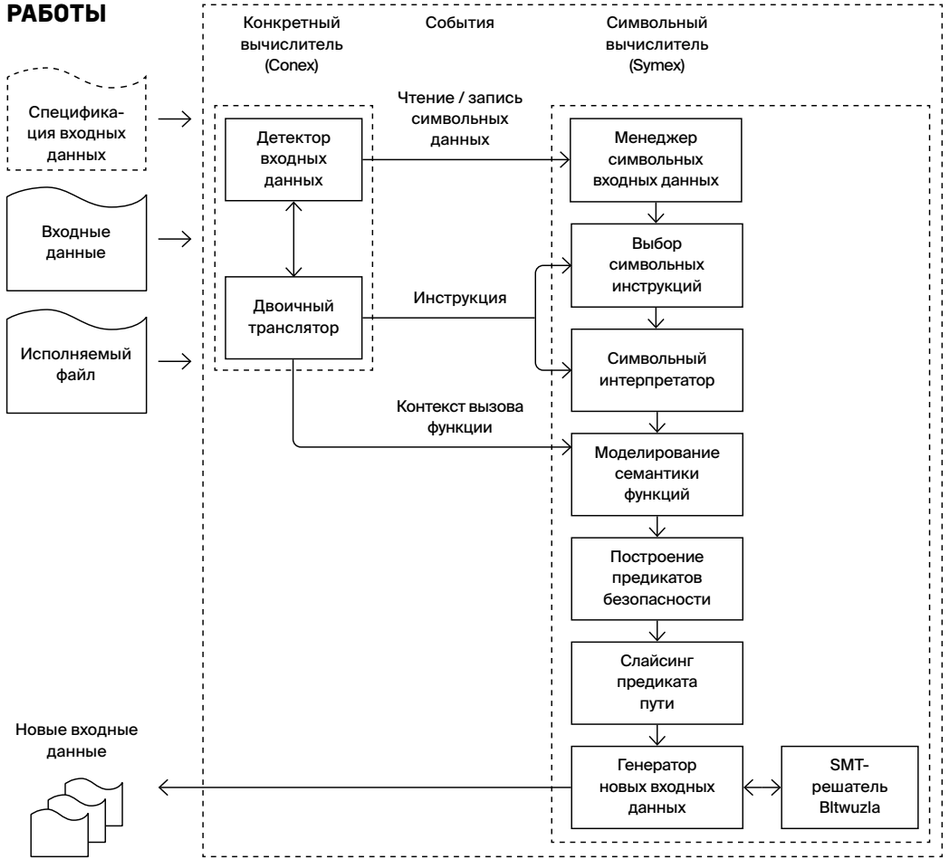
СИСТЕМНЫЕ ТРЕБОВАНИЯ

Запуск на платформах архитектуры x86_64 и Aarch64. Поддержка 64-битных ОС семейства Linux: Ubuntu 18.04/20.04/22.04, Astra 1.7, ALT Workstation 10.0 и аналогов.

ОПЫТ ВНЕДРЕНИЯ

Sydr и Sydr-fuzz входят в состав программного комплекса ИСП Crusher, который в разной комплектации используется более чем в 70 компаниях и лабораториях, в том числе в ОАО «РусБИТех», Postgres Professional, ООО «Код Безопасности», МВП «СВЕМЕЛ» и др. Используется в Исследовательском центре доверенного искусственного интеллекта ИСП РАН.

СХЕМА РАБОТЫ



PROTOSPHERE: СИСТЕМА АНАЛИЗА СЕТЕВОГО ТРАФИКА

Protosphere — система глубокого анализа сетевого трафика (DPI). Может встраиваться как компонент в системы мониторинга, классификации, защиты от вторжений и утечек информации. Регистрирует несоответствия между реализацией протокола и фактическим трафиком. Позволяет быстро добавлять поддержку новых (в том числе закрытых) протоколов благодаря универсальности внутреннего представления.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Protosphere — инновационная система, основанная на научных исследованиях технологий анализа сетевого трафика. Объединяет ключевые особенности иностранных аналогов (Wireshark, Microsoft Message Analyzer, nDPI) с универсальным внутренним представлением, позволяющим быстро расширять возможности анализа.

Protosphere — это:

- Оптимальные возможности ядра системы:
  - универсальная модель представления данных при разборе сетевого трафика;
  - обработка данных, содержащих искажения, потери, перестановки и дублирование пакетов, а также асимметричный трафик;
  - поддержка анализа сжатых и зашифрованных данных;
  - поддержка туннелей произвольной конфигурации;
  - поддержка связанных потоков.
- Поддержка всех этапов анализа сетевой трассы — каждый этап с отдельным компонентом визуализации, все компоненты синхронизированы:
  - локализация одного или нескольких исследуемых сетевых соединений на графе сетевых взаимодействий и в дереве сетевых потоков;
  - детализация выделенных соединений на временной диаграмме;
  - наглядное представление выделенных в сетевых пакетах полей в дереве разбора сетевого потока;
  - выявление несоответствий между реализацией протокола и фактическим трафиком в журнале диагностики;
  - извлечение и анализ данных произвольного уровня (L7+).
- Широкий список поддерживаемых протоколов:
  - Dns, Dhcp, Rip;
  - Tls, Microsoft-Rpc, Postgresql;
  - Ftp, Http, Imap, Smtp, Pop3, BitTorrent;
  - Kerberos, Ntlm;
  - Gre, IpSec, Ppp, OpenVpn, Wireguard;

- Быстрое расширение списка поддерживаемых протоколов:
  - API доступ к результатам разбора;
  - локализация ошибок разбора;
  - декларативное описание форматов сетевых протоколов.
- Поддержка двух режимов работы: на потоке и в отложенном режиме.
- Возможность работы в качестве DPI as Service.
- Продвинутый графический интерфейс, позволяющий выбирать наиболее удобный вариант представления результатов проводимого анализа.
- Возможности по расширению функционала системы:
  - поддержка новых программных интерфейсов;
  - разработка различных механизмов работы с результатами разбора;
  - обавление новых возможностей ядра разбора.
- Адаптация под сетевой канал и доступные вычислительные ресурсы: гибкая система конфигурирования позволяет находить баланс между детализацией/точностью анализа и потребляемыми ресурсами.

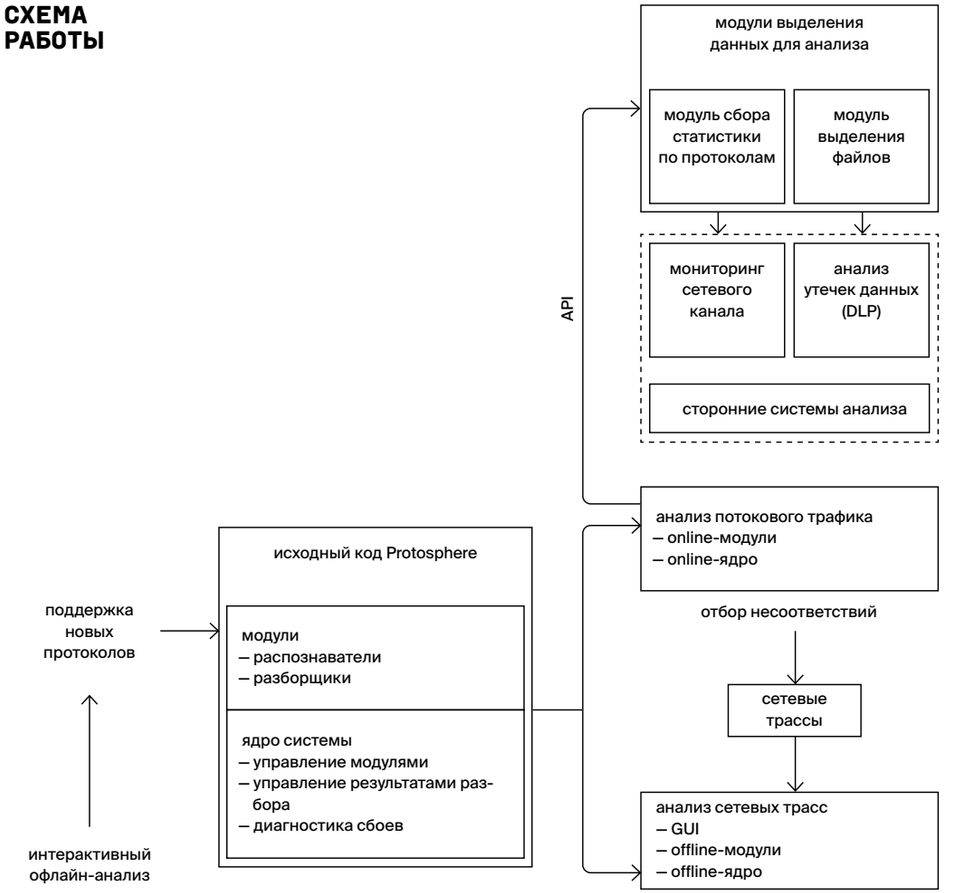
ДЛЯ КОГО ПРЕДНАЗНАЧЕНА PROTOSPHERE?

- Компании, занимающиеся тестированием реализаций сетевых протоколов (в том числе во встраиваемых ОС и сетевой аппаратуре).
- Компании-разработчики средств сетевой безопасности (межсетевых экранов, а также систем обнаружения и предотвращения вторжений, в том числе от атак нулевого дня).
- Компании по производству техники, нуждающейся в повышенном уровне безопасности из-за обязательной сертификации.
- Компании, которым требуется контроль и мониторинг сетевых каналов в режиме реального времени.

ПОДДЕРЖИВАЕМЫЕ ПЛАТФОРМЫ И АРХИТЕКТУРЫ

Архитектуры: Intel x86-64, arm64.
Платформы: ОС Windows, ОС на базе ядра Linux, Apple macOS.

СХЕМА РАБОТЫ



REQUALITY: ИНСТРУМЕНТ УПРАВЛЕНИЯ ТРЕБОВАНИЯМИ

Requality — расширяемый инструмент для управления требованиями. Позволяет как разрабатывать требования к проектируемым системам с нуля, так и формировать каталоги требований путём разметки существующих документов, сохраняя при этом связи между требованиями и текстовыми фрагментами исходных документов. Поддерживает организацию иерархической структуры требований, трассируемость между требованиями разных уровней, возможность совместной работы над требованиями с использованием системы управления версиями GIT.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Уникальной возможностью Requality является возможность формирования каталога требований путём разметки существующих документов, при этом у каждого требования сохраняется связь с одним или несколькими текстовыми фрагментами исходного документа.

По другим функциональным возможностям Requality близок к существующим коммерческим аналогам (IBM DOORS, Jama, Polarion) и превосходит ряд имеющихся свободно распространяемых продуктов (aNimble, ProR, RMTOO). Инструмент и руководство пользователя доступны на сайте проекта: https://requality.ru.

Requality — это:

- Структурирование и хранение каталога требований:
  - каталог требований — это структурированный набор связанных требований и других элементов, хранимый в рамках единого рабочего пространства. В качестве элементов верхнего уровня выступают проекты, в которых хранятся отдельные наборы требований. Такая возможность используется в том числе для отделения требований верхнего уровня от разработанных на их основе требований нижнего уровня;
  - элементы каталога включают в себя различные виды узлов:
    - требования, содержащие описания возможностей разрабатываемого объекта или накладываемые на него ограничения;
    - текстовые узлы, не являющиеся непосредственно требованиями, но обеспечивающие контекст для их понимания (например, описания терминов или примечания);
    - документальные представления требований, на основе которых был разработан каталог;
    - настройки отчётов и результаты их генерации;
    - комментарии.

  Набор элементов каталога может быть расширен путем разработки расширений;
  - идентификация узлов поддерживается несколькими способами, в том числе с использованием уникального в пределах проекта числового идентификатора и составного удобочитаемого иерархического пути;
  - свойства узлов включают в себя как предоставляемые инструментом по умолчанию (описание узла, короткий строковый идентификатор и другие), так и определяемые пользователем параметры, используемые для обозначения характеристик элементов;
  - применение HTML-разметки в тексте требований и в других свойствах позволяет использовать различные средства форматирования текста и обеспечивать использование вспомогательных ресурсов, таких как изображения и таблицы.
- Управление связями, трассируемость и анализ покрытия:
  - задание связей между элементами каталога с произвольным именем. Имена связей позволяют определить отношения различных видов;
  - автоматическое построение связей по терминологии путём перечисления списка терминов, используемых в требовании, и установки соответствующего атрибута у узла, который этот термин определяет;
  - задание связи между фрагментом текста и требованием позволяет, с одной стороны, определить источник возникновения отдельного требования, а с другой — предоставляет возможность автоматизированного переноса подобных связей на новые версии документов;
  - сквозная трассируемость представляет собой возможность проследить для отдельного требования как исходные требования, так и разработанные на его основе, а также изучить контекст элемента каталога, в рамках которого его стоит рассматривать;
  - покрытие представляет собой информацию о степени реализации или полноте тестирования каталога требований. Покрытие оценивается по информации о наличии связи между элементами каталога и внешними элементами либо между элементами внутри каталога. В инструменте поддерживается возможность использования внешней информации о покрытии в форме файла определенного формата, а также обеспечивается расширяемый набор источников данных о покрытии.

- Управление изменениями и поддержка совместной работы:
  - в качестве основной системы для совместной работы над каталогом требований поддерживается GIT. В интерфейсе инструмента доступен упрощённый набор команд для отправки изменений и обновления локальной версии проекта;
  - в интерфейсе инструмента доступен просмотр версий как отдельного узла, так и каталога требований в целом с возможностью сравнения отдельных версий;
  - поддерживается возможность сравнения версий проектов, а также переключения на предыдущие версии.
- Генерация отчётов, в частности:
  - формирование различных представлений каталога требований, в том числе для работы с ним вне инструмента, а также для обмена данными каталога требований с другими инструментами или решения нестандартных задач в рамках разработки;
  - предоставление данных о трассируемости для просмотра информации о связях между элементами каталога требований;
  - сравнение версий каталога в целях контроля прогресса работы над каталогом требований путем изучения различий структуры и свойств требований проекта для выбранных версий каталога;
  - анализ покрытия для рассмотрения статуса отдельных элементов каталога с точки зрения информации о покрытии, полученной из выбранного источника;
  - поддержка пользовательских шаблонов с использованием доступной информации относительно каталога, его версий и информации о покрытии.
- Поддерживается программный интерфейс (API) с возможностью изменения хранимых данных и создания новых проектов. В том числе он может быть использован для обмена данными со сторонними инструментами.
- Доступна возможность разработки расширений для задания новых элементов, источников информации о покрытии или для получения новых функциональных возможностей.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

ОС Windows или ОС на базе ядра GNU/Linux, Java Runtime.

ОПЫТ ВНЕДРЕНИЯ

Requality разрабатывается с 2011 года. Использовался для разработки и управления изменениями требований в проекте по разработке операционной системы реального времени в соответствии с процессами, предписанными КТ-178С, а также для каталогизации требований из различных стандартов (в том числе TTCN и POSIX) с целью проведения последующего тестирования совместимых с ними продуктов на соответствие стандарту.

2 АНАЛИЗ ДАННЫХ

ASPERITAS И ДРУГИЕ ОБЛАЧНЫЕ РЕШЕНИЯ

Asperitas — платформа организации ресурсоёмких вычислений и хранения данных в научных, образовательных и коммерческих целях. Кроме одноимённой облачной среды в состав платформы входят также Michman — оркестратор платформенных распределённых сервисов и Clouni — мультиоблачный оркестратор инфраструктурных ресурсов, основанный на стандарте TOSCA. Кроме того, в число облачных решений ИСП РАН входит Fanlight — платформа для объединения исследований в web-лаборатории и Cotea — системное ПО, предназначенное для программного контроля исполнения сценариев Ansible.

ОБЛАЧНАЯ СРЕДА ASPERITAS

В основе облачной среды Asperitas лежат открытые технологии Openstack и Ceph, которые являются современным стандартом для построения больших приватных облачных систем. Дистрибутив поставляется как готовое решение, включающее всё необходимое для настройки, в том числе TUI-установщик.

Другие преимущества Asperitas:

- Отчуждаемость решений: инфраструктура воссоздаётся в изолированной среде с полным контролем над ней за счёт использования открытых стандартов, свободного ПО, научных разработок ИСП РАН, а также изолированного репозитория исходных кодов.
- Высокий уровень безопасности: платформа построена на уменьшенной кодовой базе для сокращения поверхности атаки, а также использует собственные решения по усилению безопасности.
- Стандартные интерфейсы управления виртуальными сетями и вычислительными ресурсами с использованием систем Keystone, Neutron, Nova.
- Блочное хранение данных и расширяемое объектное хранилище на основе распределённой файловой системы Ceph.
- Возможность адаптации платформы под решение задач конкретной предметной области: решение задач механики сплошных сред, анализ больших данных, биомедицина, анализ уязвимости программ и др.

Облачная среда Asperitas включена в Единый реестр российского ПО (№5921).

МУЛЬТИОБЛАЧНЫЙ IAAS ОРКЕСТРАТОР CLOUNI

GitHub → https://github.com/ispras/clouni

Для расширения возможностей управления инфраструктурными ресурсами в ИСП РАН разрабатывается инструмент Clouni, который позволяет развёртывать кластеры виртуальной инфраструктуры по нормативным шаблонам TOSCA Simple Profile при помощи инструмента управления конфигурациями Ansible. Его основные характеристики:

- Собственный механизм трансляции декларативных шаблонов TOSCA в скрипты Ansible, благодаря которому пользователь избавляется от необходимости описывать процесс развёртывания инфраструктуры.
- Отсутствие зависимости от используемой облачной платформы (поддерживаются Openstack, Amazon AWS и частично Kubernetes).
- Тонкая настройка параметров виртуальных машин, групп безопасности, портов и сетей.

В тесной интеграции с Clouni разрабатывается фреймворк TOMMANO — инструмент управления сетевыми службами в произвольных облачных средах. Его основные характеристики:

- Автоматизированное развёртывание виртуализированных сетевых функций по их декларативному описанию на языке TOSCA в соответствии со стандартом ETSI MANO.
- Набор заготовленных шаблонов сетевых функций: Firewall, NAT, DPI, DNS, DHCP, анализаторы трафика.
- Организация service function chaining на базе программно-конфигурируемой сети, управляемой контроллером OpenDayLight. Это позволяет управлять сложными сетевыми службами, в которых разные типы трафика обрабатываются различными сетевыми функциями.
- Возможность развёртывания сетевых функций как в режиме standalone с настройкой маршрутизации через параметры TOSCA шаблона, так и в составе сервисных цепочек с автоматической маршрутизацией между узлами.

УНИВЕРСАЛЬНЫЙ ОРКЕСТРАТОР MICHMAN

GitHub → https://github.com/ispras/michman

Michman — оркестратор сервисов уровня PaaS для хранения и анализа больших данных, задач машинного обучения, инструментов управления нагрузкой и других. Инструмент позволяет автоматически развёртывать распределённые кластеры в облачной среде с учётом пользовательских требований и настроек. Им предоставляется интерфейс для развёртывания наборов сервисов из заранее настроенных шаблонов и управления их жизненным циклом, в частности:

- кластер для анализа больших данных с системами Apache Spark, Apache Hadoop, настроенными для корректного взаимодействия между собой на произвольном числе вычислительных узлов;
- СУБД различных классов: от классических реляционных до распределённых аналитических;
- системы хранения и обмена файлами, в частности MiniO, Nextcloud, NFS, GlusterFS;
- система управления ресурсами кластера и планировщик задач Slurm с возможностью использования GPU;
- гибко настраиваемая система оркестрации контейнерных приложений Kubernetes, а также инструменты, работающие поверх неё;
- инструменты для разработки моделей искусственного интеллекта, в частности Jupyter, MLflow, Ray.

Ключевые преимущества оркестратора Michman – это гибкость и расширяемость списка поддерживаемых сервисов за счёт использования языка TOSCA и поддержки следующих механизмов:

- Substitution Mapping, который позволяет унифицировано использовать однотипные ресурсы. Например, этот механизм позволяет описывать возможность развёртывания на различных ресурсах (в приватном или публичном облаке, на выделенных серверах либо с использованием контейнеров). Также с его помощью можно описать интеграцию какого-либо приложения с различными СУБД, подключение различных файловых систем и другое.
- Select, с помощью которого пользователь может переиспользовать созданные ранее ресурсы или использовать сторонние (внешний репозиторий, общая сетевая файловая система).

Кроме того, Michman позволяет консистентно сохранять состояние всех компонентов облачного приложения, масштабировать узлы, управлять отдельными частями приложения, производить обновление запущенных сервисов.

ПЛАТФОРМА FANLIGHT

Fanlight — платформа по предоставлению виртуальных рабочих столов (DaaS — Desktop as a Service). Позволяет разворачивать SaaS-инфраструктуру для вычислительных web-лабораторий. Создана в результате участия ИСП РАН в программе «Университетский кластер» и в международном проекте Open Cirrus (учреждён HP, Intel и Yahoo!). Fanlight базируется на контейнерных технологиях — в отличие от основных решений данного класса, основывающихся на виртуальных машинах. Изначально платформа базировалась на технологии Docker Compose. Позднее появилась реализация на основе Kubernetes. Поддерживает только приложения, разработанные для ОС на базе ядра Linux. Включена в Единый реестр российского ПО (№6066).

Другие преимущества Fanlight:

- Высокая эффективность работы с облачными вычислениями благодаря использованию контейнеров:
  - комфортная работа с тяжёлыми инженерными CAD-CAE приложениями, требующими поддержки аппаратного ускорения 3D-графики для сложной визуализации;
  - поддержка выполнения MPI, OpenMP, CUDA приложений за счет доступа к HPC-кластерам, многоядерным процессорам и графическим ускорителям NVIDIA.

- Расширенные вычислительные возможности на уровне PaaS за счёт подключения аппаратных ресурсов (HPC/BigData кластеры, системы хранения, сервера с графическими ускорителями).
- Возможность кастомизации под заданную прикладную область за счёт интеграции специализированных расчётных прикладных пакетов, а также простоты их добавления. В частности, есть опыт внедрения:
  - в области MCC: OpenFOAM, SALOME, Paraview и др.;
  - в области Gas&Oil: tNavigator, Eclipse, Roxar, Tempest и др.
- Работа через любой тонкий клиент (включая мобильные устройства) без вспомогательного ПО.
- Развёртывание на сервере, вычислительной ферме, в облаке (с уровня IaaS), в кластере Kubernetes или в собственном облачном ЦОД. Версия на основе Kubernetes позволяет также использовать различные CRI-движки исполнения контейнеров.

ИНСТРУМЕНТ COTEA

GitHub → https://github.com/ispras/cotea

Cotea — инструмент, позволяющий программно запускать Ansible и контролировать его выполнение (Ansible — одна из самых популярных систем по развёртыванию ПО). Cotea позволяет:

- программно контролировать выполнения Ansible, итерируясь по составным частям Ansible-сценария;
- встраивать Ansible в другие системы;
- отлаживать выполнения Ansible, в том числе и интерактивно. Переход в интерактивный режим происходит в случае ошибки выполнения task (составной части Ansible-сценария). Примеры функций, предоставляемых в интерактивном режиме:
  - перезапустить задание, завершившееся с ошибкой;
  - продолжить выполнения сценария Ansible без неудавшегося задания;
  - добавить новую переменную Ansible в процессе исполнения;
  - добавить новый Ansible task в процессе исполнения.

Интерактивный режим позволяет не начинать выполнение сценария заново в случае возникновения ошибок, что особенно важно при работе с большими сценариями.

На данный момент Cotea используется при развёртывании платформы Asperitas. Grpc-cotea является компонентом Michman и Clouni. Именно grpc-cotea позволяет данным системам оркестрации контролировать процесс развёртывания облачных приложений.

ОПЫТ ВНЕДРЕНИЯ

Вычислительный кластер на базе Asperitas используется для работы ряда технологий ИСП РАН (в частности, для анализа ОС Android с помощью Svace). Реализованы совместный проект с компанией Huawei (анализ больших графов с помощью технологий обработки больших данных) и инфраструктура поддержки жизненного цикла ОС Tizen, позволяющая организовать процесс совместной разработки компонентов ОС и автоматизировать регулярную сборку и тестирование образов. Кроме того, осуществляется ряд работ при участии Минобрнауки РФ. На базе среды Asperitas реализована облачная Платформа НЦМУ «Цифровой биодизайн и персонализированное здравоохранение».

Возможности платформы Fanlight использовались в ряде совместных проектов по развёртыванию web-лабораторий с ФГУП «РФЯЦ-ВНИИЭФ», ООО «РРС-Балтика», ИПМ им. М.В. Келдыша РАН (разработка технологий для увеличения и эффективного использования ресурсного потенциала углеводородного сырья Союзного государства), а также с Лабораторией механики сплошных сред ИСП РАН (https://unicfd.ru).

TALISMAN: ПЛАТФОРМА ДЛЯ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ

Talisman — это комплекс взаимосвязанных программных инструментов для автоматизации типовых задач обработки данных, включая их сбор, интеграцию, анализ, хранение и визуализацию. Обеспечивает быструю разработку специализированных многопользовательских интеллектуальных аналитических систем, объединяющих информацию из внутренних баз данных и открытых источников сети Интернет (в том числе из социальных сетей).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Talisman использует технологии больших данных и передовые методы искусственного интеллекта для извлечения информации из произвольных источников. Позволяет быстро создавать интеллектуальные аналитические системы, используя подходы Low-code и No-code. Постоянно обучается на результатах работы аналитика, не требуя дополнительных трудозатрат.

Talisman — это:

- Широкий набор переиспользуемых компонентов, каждый из которых обладает программным интерфейсом для удобного управления и взаимной интеграции:
  - Компоненты для получения исходных данных. В частности, это программный комплекс сбора данных из сети Интернет: из соцсетей (Вконтакте, Facebook, Twitter, Instagram, Одноклассники, Youtube, LinkedIn и др.), блогов, СМИ, сайтов mediawiki, порталов разработчиков ПО и др. Кроме того, есть система импорта данных из файловых хранилищ и СУБД.
  - Компоненты автоматического анализа данных. Набор инструментов, позволяющих преобразовать входные данные любых форматов и привести их к единому универсальному представлению (в частности, используется разработка ИСП РАН Dedoc). Документы в этом представлении подвергаются анализу с помощью методов машинного обучения. Имеется возможность добавлять собственные обработчики в виде контейнеров с REST API. Управление последовательностью обработки осуществляется системой «Talisman.Поток» (№6045 в Едином реестре российского ПО).
  - Компоненты хранения и индексации. Это группа СУБД и информационно-поисковых систем, где хранятся исходные данные, результаты автоматической обработки, а также результаты работы пользователей.
- Удобный веб-интерфейс, который объединяет все компоненты, предполагающие взаимодействие с пользователями.
- Гибкая модульная архитектура, позволяющая добавлять новые функции в отдельные компоненты без изменения большинства остальных.
- Горизонтально масштабируемая архитектура, позволяющая увеличивать объёмы обрабатываемых и хранимых данных без изменения программной части за счёт добавления аппаратных ресурсов.
- Специализированные подсистемы, которые отвечают за мониторинг состояния компонентов, управление журналом событий, развёртывание, аутентификацию и авторизацию, разграничение прав доступа, а также однонаправленную передачу данных.
- Инструменты и методики обучения моделей машинного обучения, а также переноса имеющихся моделей и алгоритмов на новую предметную область.
- Настраиваемая схема предметной области с возможностью внесения изменений оператором в процессе эксплуатации системы.
- Полная отчуждаемость разрабатываемых систем. Каждая из них может быть развёрнута на площадке заказчика — как на существующем оборудовании, так и в составе программно-аппаратного комплекса.
- Интеграция с внутренними системами потребителя благодаря наличию программного интерфейса для управления всеми компонентами.
- Лицензионная чистота благодаря базированию на собственных разработках ИСП РАН и свободном ПО.

ОБЛАСТИ ПРИМЕНЕНИЯ

Talisman позволяет создавать аналитические системы для решения широкого круга прикладных задач. Примеры применения:

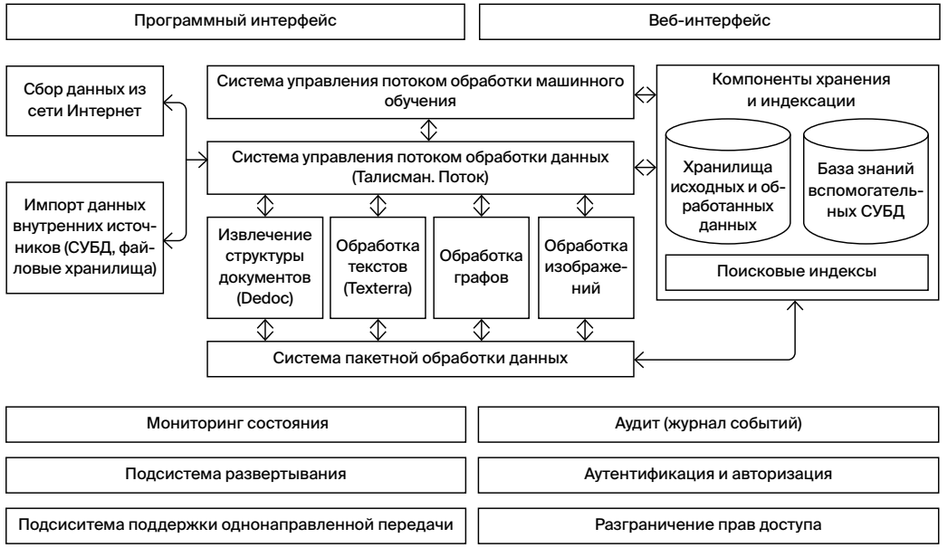
- Автоматизация построения базы знаний по интересующей предметной области и обеспечение постоянного мониторинга новой информации об объектах интереса (аналог Palantir Gotham).
- Проведение конкурентной разведки по открытым данным (OSINT) с целью поиска сведений по объектам интереса (аналог Maltego).

- Мониторинг СМИ с целью решения аналитических задач (аналог LexisNexis).
- Оптимизация управления персоналом: эффективный подбор сотрудников, верификация анкетных данных, выявление некорректного поведения в открытом информационном пространстве (система «Talisman.Биография», №5547 в Едином реестре российского ПО).
- Выявление информационных кампаний, манипулирующих мнением целевой аудитории, а также определение целевой аудитории, на которую направлена кампания.
- Выявление и анализ особенностей инфраструктуры распространения информации (ресурсы, пользователи, боты), а также анализ типичных ролей членов сообществ в коммуникации (первоисточник, лидер мнения, распространитель, модератор, бот, комментатор).
- Управление деловой репутацией людей и организаций: мониторинг релевантных сообщений, выявление проблем, вызывающих недовольство, мониторинг утечек и разглашения внутренней информации.
- Объективная оценка эффективности деятельности, а также тестирование стратегий на целевой аудитории в целях получения обратной связи.
- Управление точками социального напряжения; обнаружение и своевременное предупреждение эскалации конфликтов.

ПОДДЕРЖИВАЕМЫЕ ЯЗЫКИ

Talisman использует современные искусственные нейронные сети для анализа данных. Используемые инструменты позволяют извлекать информацию более чем из 100 естественных языков.

СХЕМА РАБОТЫ



ДОВЕРЕННЫЕ ФРЕЙМВОРКИ МАШИННОГО ОБУЧЕНИЯ

Фреймворк машинного обучения — это среда с набором инструментов для обеспечения быстрой разработки соответствующих программных продуктов. Наиболее популярными открытыми фреймворками считаются TensorFlow и PyTorch. Работа по созданию их доверенных версий ведётся в Исследовательском центре доверенного искусственного интеллекта (ИЦДИИ) ИСП РАН. Цель проекта – найти и устранить ошибки, которые могут вызвать некорректную работу фреймворков, а значит, привести к некорректной работе приложений.

ИНФРАСТРУКТУРА

Для анализа фреймворков развёрнута аппаратно-программная инфраструктура обеспечения доверия, с помощью которой проводятся:

- статический и динамический анализ кода;
- постоянная проверка новых изменений кода;
- синхронизация с оригинальными открытыми версиями

ИСПОЛЬЗУЕМЫЕ ИНСТРУМЕНТЫ (РАЗРАБОТАНЫ В ИСП РАН)

Svace – инструмент статического анализа исходного кода. Обнаруживает более 50 классов критических ошибок. Поддерживает языки C, C++, C#, Java, Kotlin и Go. Поставляется с web-интерфейсом просмотра предупреждений Svacer (Svace History Server).

Sydr+Sydr-fuzz — комплекс гибридного фаззинга и динамического анализа. Включает в себя Sydr – инструмент автоматической генерации тестов для сложных программных систем с целью увеличения покрытия кода и обнаружения ошибок, а также Sydr-fuzz — инструмент динамического анализа программ, который сочетает в себе возможности Sydr и современных фаззеров (libFuzzer и AFL++).

РЕЗУЛЬТАТЫ (ДАННЫЕ НА НОЯБРЬ 2023)

- Обнаружен ряд ошибок, почти все исправления приняты в основные ветки фреймворков:
  - в TensorFlow найдены 27 ошибок (4 с помощью Sydr+Sydr-fuzz, 23 с помощью Svace), 26 исправлений приняты в основную ветку;
  - в PyTorch найдены 39 ошибок (26 с помощью Sydr+Sydr-fuzz, 13 с помощью Svace), 36 исправлений приняты в основную ветку.
- 10 ошибок обнаружены в сторонних проектах (LLVM, oneDNN, miniz, torchvision, openjpeg).
- Разработана методика создания доверенных фреймворков.
- Разработаны доверенные фреймворки, которые апробированы на решениях индустриальных партнёров ИЦДИИ и внедрены в «Kaspersky Machine Learning for Anomaly Detection» v. 3.0.

ДЛЯ КОГО ПРЕДНАЗНАЧЕНЫ ДОВЕРЕННЫЕ ФРЕЙМВОРКИ?

Доверенные фреймворки предназначены для применения в компаниях, разрабатывающих приложения на основе машинного обучения и нацеленных на высокий уровень надёжности и безопасности ПО.

LINGVODOC: ВИРТУАЛЬНАЯ ЛАБОРАТОРИЯ ДЛЯ ДОКУМЕНТАЦИИ ИСЧЕЗАЮЩИХ ЯЗЫКОВ

GitHub → https://github.com/ispras/lingvodoc

Lingvodoc — система для совместной многопользовательской документации исчезающих языков, создания многослойных словарей и научной работы с полученными звуковыми и текстовыми данными. Совместный проект с Институтом языкознания РАН и Томским государственным университетом. Разрабатывается с 2012 года. Сайт — lingvodoc.ispras.ru.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Lingvodoc — кроссплатформенная технология с открытым исходным кодом (https://github.com/ispras/lingvodoc и https://github.com/ispras/lingvodoc-react), основанная на научных исследованиях.

Lingvodoc — это:

- Совместная работа пользователей над пополнением словарных данных (в отличие от аналогичного проекта Starling, где такая работа не предусмотрена).
- Сохранение полной истории действий пользователей.
- Одновременная работа с аудиотекстовыми корпусами и словарями на основе интеграции с программой ELAN, разработанной Институтом психолингвистики Макса Планка (Нидерланды).
- Расставление однонаправленных и двунаправленных связей между лексическими входами внутри словарей, а также между словарями.
- Запись, проигрывание и хранение звуков с разметкой (в форматах WAV, MP3 и FLAC), а также построение формант гласных с последующей визуализацией.
- Продвинутый поиск, который позволяет искать данные в словарях по множеству параметров (в отличие от аналогичного проекта TypeCraft).
- Возможность поиска данных на карте с автоматическим построением изоглосс.
- Возможность бесконфликтной двусторонней отложенной синхронизации.

- Повышенный уровень автоматизации (по сравнению с аналогичным проектом Kielipankki): возможность проводить автоматический этимологический и фонетический анализ.
- Создание словарей любой структуры, как типичных двуслойных (слой лексических входов и слой парадигм), так и многослойных. Кроме того, существует функция импорта для готовых словарных структур.
- Наличие программ, воспроизводящих работу ученых по фонетическому и этимологическому анализу.
- Возможность уточнения классификации языков и диалектов с построениями графиков в форматах 2D и 3D по глоттохронологии, морфологии, этимологико-фонетическим признакам.
- Возможность размещения корпусов текстов в формате Word и словарей в формате Excel.
- Встроенные морфологические анализаторы для языков народов России в формате Aperitum.
- Удобный интерфейс для ручного снятия омонимии после работы морфологического анализатора.
- Работа как с привлечением облачных ресурсов ИСП РАН, так и с развёртыванием локальной версии с изоляцией собственных данных.
- Наличие программы для веб-просмотра и десктопной версии.
- Открытая регистрация (с подтверждением).
- Оперативная доработка технологии с расширением набора функций, а также адаптация под другую научную отрасль.

На основе глоссированных корпусов Lingvodoc и программ анализа создана драфтовая версия обучающей платформы edu.ispras.ru, объединяющая 10 000 заданий на 9 языках. Платформа предоставляет возможность:
- создавать упражнения на языках России для любого возраста и уровня владения (любой учитель может создавать свои упражнения, и ему не придётся самому их проверять);
- выполнять упражнения как на родном языке, так и на иностранном, причём каждому ученику программа может подбирать задания индивидуально в зависимости от его ошибок.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН LINGVODOC?

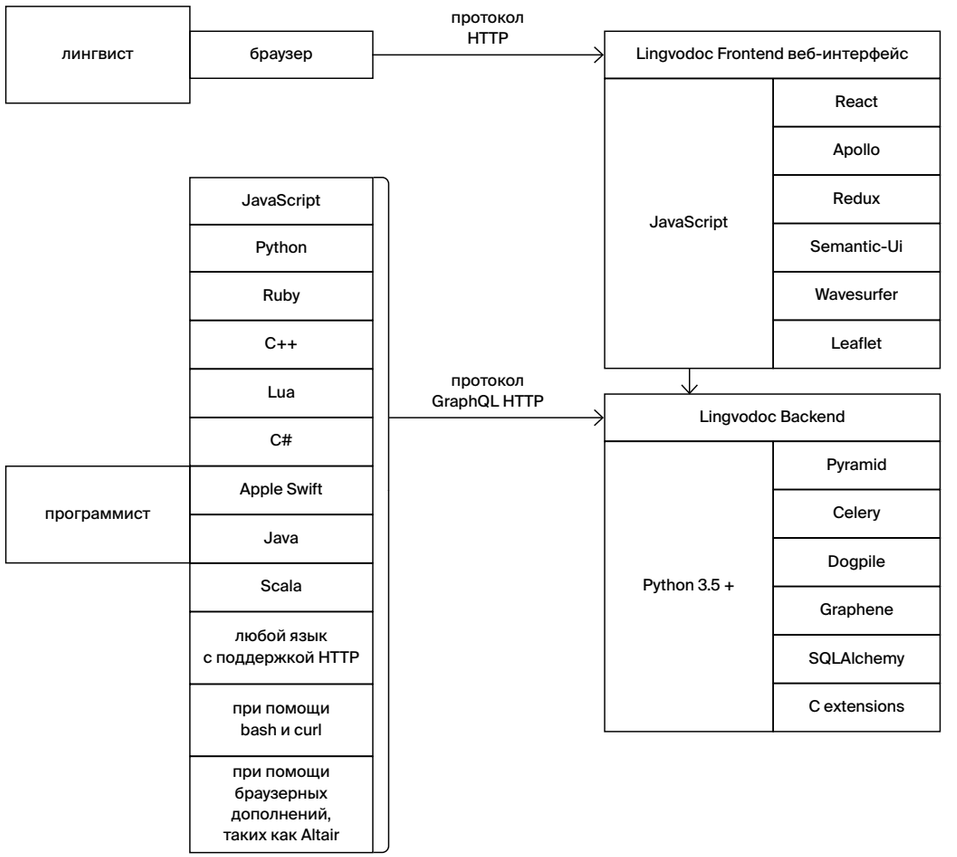
В первую очередь Lingvodoc разработан для лингвистов, ведущих научную работу в сфере документации языков народов России. Однако возможна доработка технологии под другие цели.

ОПЫТ ВНЕДРЕНИЯ

Основной функционал Lingvodoc используется филологами из 29 вузов и НИИ из 16 городов. В числе научно-образовательных организаций: Томский государственный университет, Институт филологии СО РАН, Институт истории, языка и литературы Уфимского научного центра РАН, Удмуртский федеральный исследовательский центр УрО РАН, Северо-Восточный федеральный университет, Югорский государственный университет,

Институт языка, литературы и истории КарНЦ РАН, Мурманский арктический государственный университет. Специалисты, использующие платформу, готовы к проведению мастер-классов для своих коллег. В 2023 г. четыре потока учёных из нескольких городов России прошли курсы дополнительного образования по направлению «Использование возможностей платформы Lingvodoc в работе лингвистов» (в том числе в Башкирском государственном университете и в РУДН).

СХЕМА РАБОТЫ



DEDOC: СИСТЕМА ИЗВЛЕЧЕНИЯ СОДЕРЖИМОГО И СТРУКТУРЫ ТЕКСТОВЫХ ДОКУМЕНТОВ

Dedoc — универсальная открытая библиотека для приведения документов к единому выходному формату. Автоматически извлекает содержимое, логическую структуру, таблицы, форматирование и метаинформацию. Содержимое документов представляется в виде дерева, кодирующего заголовки и списки различного уровня вложенности. Dedoc может встраиваться как отдельный компонент в системы анализа структуры и содержимого документов.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Dedoc реализован на языке Python. Работает со слабоструктурированными форматами данных (DOC*, ODT, XLS/XLSX, CSV, TXT, JSON) и с неструктурированными форматами изображений (PNG, JPG и др.), архивами (ZIP, RAR и др.), PDF, HTML. Извлечение структуры документа проводится в полностью автоматическом режиме вне зависимости от типа входных данных, с извлечением метаинформации и разного вида форматирования текста.

Dedoc — это:

- Python-библиотека с открытым исходным кодом (https://github.com/ispras/dedoc).
- Расширяемость за счёт гибкого добавления поддержки новых форматов документов и простоты изменения выходного формата данных.
- Поддержка извлечения структуры вложенных документов различных форматов.
- Извлечение разного вида форматирования текста (отступы, шрифты, жирность, размер шрифта и др.).
- Работа с документами различной предметной области (технические задания, нормативно-правовые акты, научные отчёты и статьи) и возможность добавления обработки документов новой предметной области.
- Работа с PDF-документами, содержащими текстовый слой:
  - поддержка автоматического определения корректности текстового слоя.
  - Извлечение содержимого и форматирования из PDF-документов с текстовым слоем с помощью разработанного интерпретатора виртуальной стековой машины вывода на печать графики согласно спецификации формата.
- Извлечение табличной информации из DOC*, PDF-документов, HTML, форматов изображений, CSV:
  - распознавание физической структуры и текста ячеек сложных многостраничных таблиц с границами на изображениях с помощью методов контурного анализа.
- Работа со сканированными черно-белыми документами (формата PDF без текстового слоя и с форматами изображений):
  - работа с активно развивающимся движком оптического распознавания символов OCR Tesseract компании Google в совокупности с использованием методов предварительной обработки изображений;
  - использование современных методов машинного обучения для определения ориентации документов, определения одно/многоколоночных документов, полужирного текста и извлечения иерархической структуры на основе классификации строк извлечённых признаков из изображений документов;
  - возможность включения бинаризации для обработки документов с подложкой.

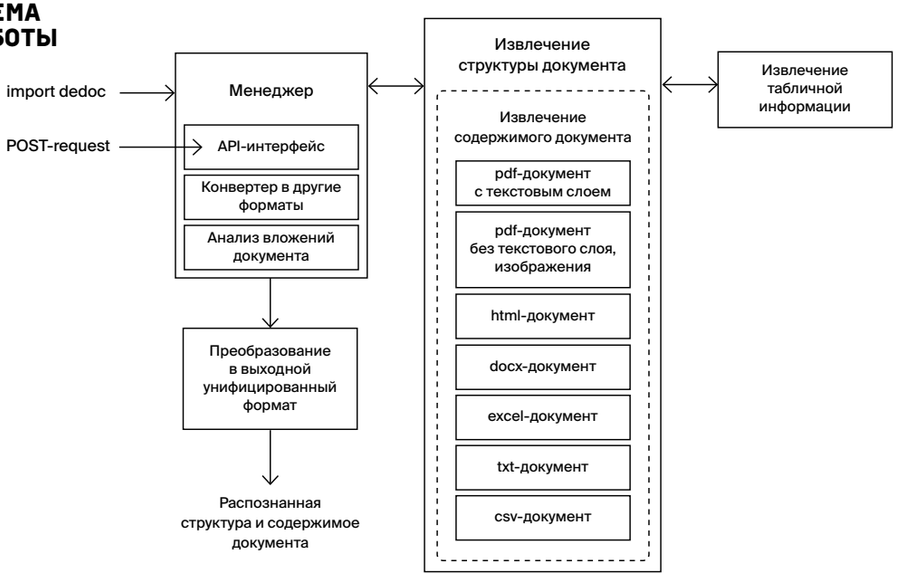
ДЛЯ КОГО ПРЕДНАЗНАЧЕНА СИСТЕМА DEDOC?

- Разработчики прикладных систем анализа содержимого электронных документов и документооборота.
- Разработчики интеллектуального анализа текста документов.
- Разработчики систем автоматической обработки текстов.

ПОДДЕРЖИВАЕМЫЕ ЯЗЫКИ

Русский и английский.

СХЕМА РАБОТЫ



DOCMARKING: ПРЕДОТВРАЩЕНИЕ АНОНИМНЫХ УТЕЧЕК ДОКУМЕНТОВ

DocMarking — уникальная система внедрения цифровых водяных знаков (меток) в текстовые документы. Позволяет создавать едва отличимые от оригинала цифровые и физические копии документов, однозначно идентифицирующие пользователей и их устройства.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

DocMarking базируется на результатах исследований в областях стеганографии, цифровой обработки изображений и машинного обучения. В основе системы маркирования лежат методы поиска и классификации текста на изображениях, используются статистические особенности изображений документов.

По сравнению с аналогичными технологиями DocMarking обладает рядом преимуществ. При извлечении цифровой метки не требуется оригинал документа. Поддерживается многократное маркирование сканированных документов, при этом ранее встроенная цифровая метка стирается.

DocMarking — это:

- Алгоритмы маркирования на основе технологий машинного обучения.
- Поддержка документов любых форматов.
- Работа во всех приложениях.
- Защита документов при отображении на экране монитора и при печати.
- Извлечение цифровых меток в условиях отсутствия оригинальных немаркированных документов.
- Возможность автономной работы на стороне клиента.
- Централизованный мониторинг всех подключённых к системе устройств в режиме 24/7.

ДЛЯ КОГО ПРЕДНАЗНАЧЕНА СИСТЕМА DOCMARKING?

- Государственные и коммерческие учреждения.
- Компании, заинтересованные в соблюдении сотрудниками протокола работы с конфиденциальными документами.

ПОДДЕРЖИВАЕМЫЕ ОС

Windows (32-bit, 64-bit), Linux (64-bit), в том числе Astra Linux 1.6 SE /1.7 SE.

СХЕМА РАБОТЫ

ECGHUB: КОМПЛЕКС ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ЦИФРОВОЙ ЭКГ

EcgHub — система разметки 12-канальных ЭКГ и нейросетевые модели классификации патологий. Система позволяет предсказывать наличие или отсутствие ряда патологий, а также выполнять и верифицировать синдромальную разметку ЭКГ на основе специализированного опросника, обеспечивая подготовку качественного набора данных для дальнейшего развития нейросетевых моделей.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

EcgHub базируется на результатах исследований в областях цифровой обработки сигналов и алгоритмов машинного обучения. В основе системы классификации патологий лежат глубокие нейронные сети. Верифицированный экспертами подход обеспечивает получение врачами согласованной разметки ЭКГ для обучения и развития предсказательных моделей скрининга, диагностики сердечно-сосудистых заболеваний.

EcgHub — это:

- Глубокие нейронные сети предсказания патологий для цифровой ЭКГ.
- Непрерывное развитие и доработка нейросетевых моделей, в том числе тонкая настройка для сравнительно небольших наборов данных медицинских учреждений.
- Адаптация обученных нейросетевых моделей для классификации патологий одноканальных ЭКГ (кардиокресло, умные часы), а также суточных ЭКГ (холтеры).
- Система согласованной синдромальной разметки, позволяющая получать качественные данные для обучения предсказательных моделей.
- Интеграция нейросетевых моделей в цифровой контур заказчика или удаленный доступ к сервису в контуре ИСП РАН.
- Возможность применения системы разметки в процессе обучения современных специалистов функциональной диагностики.
- Развитие автоматизированной системы скрининга населения.

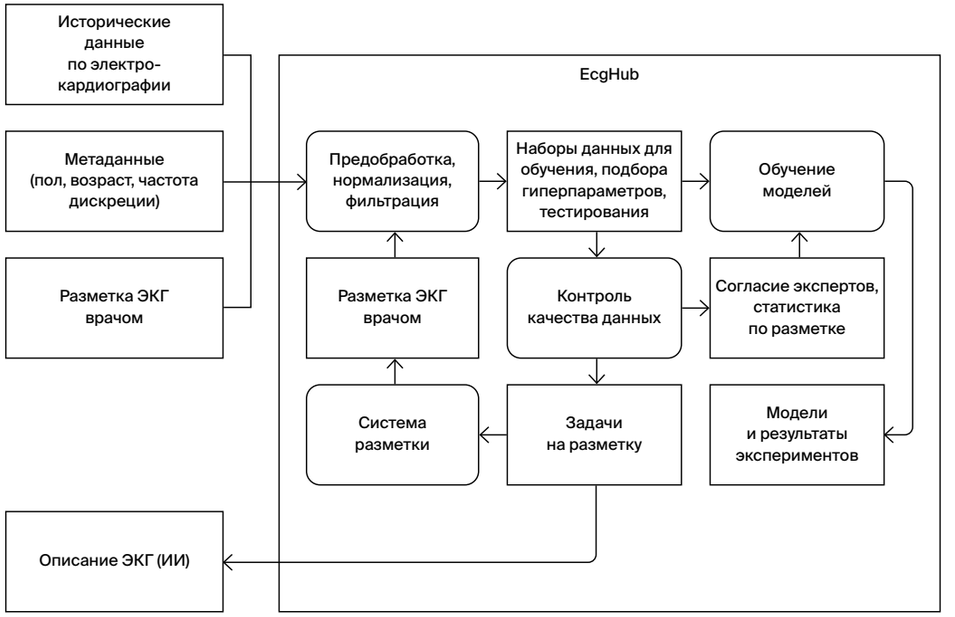
ДЛЯ КОГО ПРЕДНАЗНАЧЕНА СИСТЕМА ECGHUB?

- Медицинские учреждения (результат работы нейросетевых моделей может использоваться в качестве второго мнения);
- Образовательные учреждения (верифицированные наборы данных позволяют оценивать качество работы студентов или начинающих врачей соответствующих специализаций);
- Разработчики устройств и приложений (автономная ЭКГ-диагностика).

ОПЫТ ВНЕДРЕНИЯ

Нейросетевая модель классификации 12-канальных ЭКГ обучена на данных Республики Татарстан, интегрирована в режиме опытной эксплуатации в систему «Единый Кардиолог» и апробирована на данных ЭКГ из разных регионов (Республика Татарстан, Москва, Великий Новгород).

СХЕМА РАБОТЫ



3

ПРОЧИЕ ТЕХНОЛОГИИ

CONSTRUCTIVITY 4D: ТЕХНОЛОГИЯ ИНДЕКСИРОВАНИЯ, ПОИСКА И АНАЛИЗА БОЛЬШИХ ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ДАННЫХ

Constructivity 4D — технология для создания перспективных программных систем и сервисов, оперирующих динамическими сценами и большими массивами пространственно-временных данных. Способна проводить визуальный анализ миллионов объектов с различным геометрическим представлением и индивидуальным динамическим поведением. Внедрена в систему Synchro (Bentley Systems), предназначенную для 4D-моделирования крупных строительных объектов.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Constructivity 4D — технология для промышленного использования, объединяющая оригинальные методы пространственно-временного индексирования, поиска, а также качественного и количественного анализа данных с учётом особенностей их геометрического представления, сложной организации и предопределённого характера динамики.

Constructivity 4D — это:

- Использование развитых наборов операций для эффективного исполнения запросов:
  - темпоральные операции (реализуют классическую интервальную алгебру Аллена применительно к временным штампам дискретных событий и их интервалам);
  - метрические операции (позволяют определять индивидуальные свойства геометрических объектов и характеристики их взаимного расположения: диаметр, площадь, объем, центр масс, планарные проекции и др.);

- топологические операции (предназначены для классификации взаимного расположения объектов и установления фактов их совпадения, пересечения, покрытия, касания, перекрытия или коллизии). Допускают конструктивную имплементацию и применимы для анализа сложных объектов (в отличие от известных топологических моделей DE-9IM, RCC-8 и RCC-3D);
- ориентационные операции (обобщают известные системы исчисления направлений Франка, Фрексы, CDC, OPRA и применимы для анализа объектов с протяжёнными границами).

— Эффективное исполнение запросов и решение типовых задач (реконструкция сцены на заданный момент времени, выборка объектов в заданной пространственной области, поиск ближайших соседей, определение статических и динамических столкновений, бесконфликтная маршрутизация в глобальном динамическом окружении).
— Система пространственно-временного индексирования (бинарные деревья событий, октарные деревья пространственной декомпозиции, деревья ограничивающих объёмов, объектных кластеров, занятости пространства).
— Комбинированная вычислительная стратегия для определения столкновений в сценах. Объединяет методы точного определения столкновений, методы локализации на основе пространственной декомпозиции, иерархии ограничивающих объемов и методы темпоральной когерентности.
— Объектно-ориентированная реализация на языке C++ (расширяемый набор классов, интерфейсов и связанных с ними методов для задания пространственно-временных данных и исполнения типовых запросов к ним).
— Оригинальный метод маршрутизации в глобальном динамическом окружении. Основан на извлечении пространственной, метрической и топологической информации, а также на её согласованном использовании при планировании путей.
— Различные возможности расширения библиотеки, которая может использоваться при разработке новых приложений, а также для оптимизации и расширения функций уже существующих.

ДЛЯ КОГО ПРЕДНАЗНАЧЕНА CONSTRUCTIVITY 4D?

Технология используется для создания приложений в самых разных предметных областях, в числе которых: компьютерная графика и анимация, геоинформатика, научная визуализация, автоматизация проектирования и производства, робототехника, логистика, планирование и управление проектами.

ОПЫТ ВНЕДРЕНИЯ

Технология успешно используется в составе программной системы Synchro (https://www.bentley.com/ru/products/brands/synchro), предназначенной для визуального 4D-моделирования, планирования и управления масштабными индустриальными проектами в сфере строительства зданий, инфраструктурных объектов и др. Применяется более чем 300 компаниями в 36 странах (в том числе в России).

VALIDBIM: СЕРВИС ВЕРИФИКАЦИИ ИНФОРМАЦИОННЫХ МОДЕЛЕЙ В АРХИТЕКТУРЕ И СТРОИТЕЛЬСТВЕ

VALIDBIM — сервис верификации информационных моделей в архитектуре и строительстве, представленных в формате IFC SPF и обеспечивающих функциональную совместимость программных приложений на уровне BIM Level3. Данный уровень технологической зрелости в модели Бью-Ричардса предполагает интероперабельность ТИМ-приложений и возможность их интеграции в составе перспективных мультидисциплинарных программных комплексов для проектной деятельности в архитектуре, инженерии, строительстве и эксплуатации зданий и сооружений.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

VALIDBIM — сервис, способный качественно улучшить ситуацию в области перспективного ПО, которое удовлетворяет требованиям технической, синтаксической и семантической интероперабельности и соответствует новому уровню технологической зрелости ТИМ.

VALIDBIM — это:

— Верификация информационных моделей в архитектуре и строительстве на соответствие международным и национальным стандартам IFC (Industry Foundation Classes — ISO 16739; ГОСТ Р 10.0.02: 2019) и SPF (STEP Physical File — ISO 10303-21; ГОСТ Р ИСО 10303-21: 2002, 2022).
— Проверка синтаксиса и ссылочной целостности файловых данных модели.
— Полная и математически строгая проверка семантики данных модели на основе формальной схемы, специфицированной на языке объектно-ориентированного моделирования EXPRESS.

Проверка обеспечивает контроль:
- типов объектов (ENTITY),
- количества и типов атрибутов объекта, в том числе атрибутов с перечислимой типизацией (SELECT),
- обязательных и опциональных атрибутов (OPTIONAL),
- ограничений для длин символьных и бинарных строк (STRING, BINARY),
- размеров коллекций (BAG, SET, LIST, ARRAY),
- обязательных и опциональных элементов коллекций (OPTIONAL ARRAY),
- уникальности элементов коллекций, являющихся множествами (SET, LIST OF UNIQUE),
- размеров и состава коллекций, являющихся обратными атрибутами (INVERSE).

Кроме того, проверка обеспечивает выполнимость:
- правил для областей значений простых типов (TYPE WHERE),
- правил согласованности атрибутов объектов (ENTITY WHERE),
- правил уникальности атрибутов объектов (ENTITY UNIQUE),
- глобальных правил согласованности коллекций объектов (RULE).

- Верификация программного обеспечения, для которого декларируется техническая, синтаксическая и семантическая интероперабельность на уровне BIM Level3.
- Поддержка актуальных версий стандарта IFC, включая IFC 2x3, IFC 4 и IFC 4.3.
- Журнализация выявленных ошибок и отправка на почту зарегистрированного пользователя.
- Оперативная обработка пользовательских заданий по верификации.

ДЛЯ КОГО ПРЕДНАЗНАЧЕНА VALIDBIM?

- Для разработчиков программного обеспечения ТИМ, нацеленных на создание перспективных интероперабельных продуктов и нуждающихся в надежных средствах их верификации.
- Для пользователей ТИМ, желающих удостовериться в качестве и полноте информационных моделей и возможности работы с ними с использованием продуктов от разных производителей.

ОПЫТ ВНЕДРЕНИЯ

Сервис реализован и развернут в составе Национальной платформы ТИМ на сайте проекта bim.ispras.ru. В настоящее время сервис активно используется российскими разработчиками и пользователями программного обеспечения ТИМ.

DIGITEF: ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ ПРОВЕДЕНИЯ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

DigiTEF — программный комплекс, предназначенный для разработки средств цифрового моделирования, для проведения компьютерного моделирования и инженерного анализа прикладных и научно-технических задач промышленности. Позволяет решать задачи газовой динамики, аэродинамики, гидродинамики, акустики, а также проводить сопряжённые расчеты. Включён в Единый реестр российского ПО (№5377).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

DigiTEF разрабатывается на базе программ с открытым исходным кодом, а также уникальных модулей и библиотек ИСП РАН. Использование собственных разработок позволяет получать для отдельного класса задач решения, точность и достоверность которых превышает уровень мировых аналогов. Сравнительные исследования производительности и точности ядра DigiTEF с ANSYS Fluent и Star CCM+ показали сопоставимые (а в некоторых случаях и более низкие) вычислительные затраты при одинаковой точности.

DigiTEF — это:

- открытый исходный код (позволяет повысить сохранность данных, контролировать и адаптировать реализованные алгоритмы под конкретные задачи);
- понятный графический интерфейс, который при необходимости может быть адаптирован под конкретное предприятие и решаемые задачи;
- отсутствие ограничений на количество пользователей, ячеек расчётной сетки и используемых ядер, что позволяет снизить экономические затраты на вычисления и дальнейшее использование;
- использование современных моделей и алгоритмов за счет синхронизации технологического уровня с международным сообществом;

- наличие средств автоматизации вычислений и интеграции моделей для комплексного исследования технических объектов;
- возможность разработки дополнительных компонентов в соответствии с конкретными требованиями;
- возможность использования высокопроизводительных систем вычислений (суперкомпьютеров и кластеров) для ускорения вычислений.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН DIGITEF?

DigiTEF предназначен для применения в компаниях и на предприятиях ресурсоёмких отраслей промышленности. Использование DigiTEF позволяет повысить эффективность проектирования и рентабельность производства, снизить стоимость и сложность при реализации индустриальных проектов.

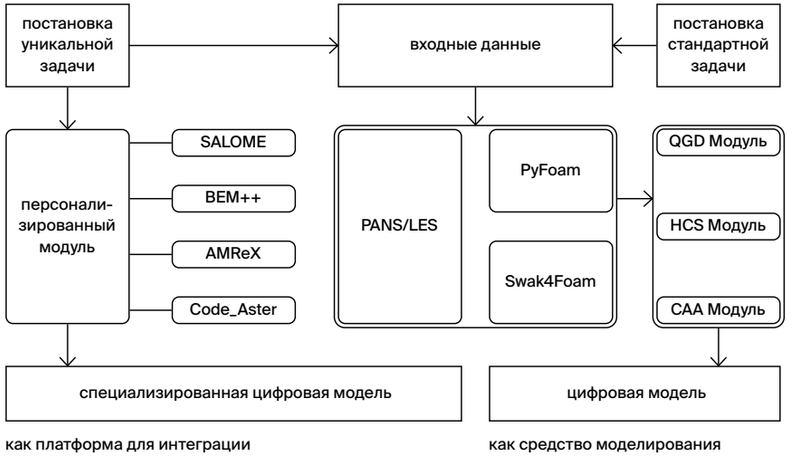
ОПЫТ ВНЕДРЕНИЯ

DigiTEF используется в ряде проектов в области ветроэнергетики, космонавтики, авиации, кораблестроения, металлургии, а также в нефтегазовой отрасли. Открытые версии модулей DigiTEF успешно применяются в академических, образовательных и промышленных учреждениях мира: Institut Pprime (Франция), Korea Atomic Energy Research Institute (Корея), Universität der Bundeswehr München (Германия), Northwestern Polytechnical University (КНР), Ocean University of China (КНР), Embry-Riddle University (США), California Institute of Technology (США) и др.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

Поддержка ОС семейства Linux (в том числе Astra Linux), а также Microsoft Windows 10. Не менее четырёх процессорных ядер x86-64, 16 ГБ оперативной памяти и 100 ГБ свободного места на диске. DigiTEF также поддерживает параллельные вычисления. Проверенное количество вычислительных ядер — до 1536.

СХЕМА РАБОТЫ



Digital Test Facility

Институт системного программирования
им. В.П. Иванникова РАН

109004, Москва, улица А. Солженицына, дом 25.
По всем вопросам: scsec@ispras.ru

ИСП РАН им. В. П. Иванникова 2023