СБОРНИК ТЕХНОЛОГИЙ

2020

ОГЛАВЛЕНИЕ

80 ЛЕТ СО ДНЯ РОЖДЕНИЯ В.П.ИВАННИКОВА (1940–2016)

АРУТЮН АВЕТИСЯН

академик РАН,
директор ИСП РАН

В этом году исполнилось 80 лет со дня рождения академика РАН Виктора Петровича Иванникова — основателя и первого директора ИСП РАН, ныне носящего его имя. Его вклад в развитие Института трудно описать в двух словах. Это и научная школа, и модель работы — знаменитый физтеховский треугольник: наука, образование, инновации. Но самым главным было его отношение к работе и к людям.

Виктор Петрович поступил в Физтех вскоре после его основания. Студенты этого вуза уже с третьего курса начинали вовлекаться в проекты академических институтов и конструкторских бюро, создавать технологии без отрыва от образования. Виктор Петрович учился у Л.Н. Королёва на кафедре С.А. Лебедева — основоположника вычислительной техники в СССР, общался с первыми советскими программистами. Сразу после окончания Физтеха, работая в Институте точной механики и вычислительной техники (ИТМиВТ) Академии наук, начал создавать операционную систему для БЭСМ-6. Потом участвовал в создании многомашинного информационно-вычислительного комплекса реального времени АС-6, который использовался в центрах управления полётами космических аппаратов, руководил созданием его операционной системы. В 80-е годы работал в Институте проблем кибернетики Академии наук СССР, где руководил созданием и внедрением систем автоматизации проектирования и программного обеспечения суперЭВМ.

Виктор Петрович понимал, что системное программирование — инновационная область, которая требует постоянного развития и передовых исследований вместе с внедрением. А для этого нужны соответствующие структуры — например, академические институты. В 1994 году ему удалось создать такой институт, которому Виктор Петрович передал лучшее из своих достижений и опыта — свою научную школу, которая сформировалась на основе школы Лебедева в атмосфере открытости и свободного творчества в ИТМиВТ.

Первые годы работы ИСП РАН были непростыми из-за общей экономической ситуации. Потом появились первые совместные проекты с зарубежными партнёрами, увеличилось число студентов. Как говорил Виктор Петрович, в Институте образовались два встречных потока: «качественный поток денег» на основе контрактов на передовые исследования и разработки и поток талантливой молодежи. В начале 2000-х годов начался период стабилизации. Появились новые партнёры — Intel, HP, Dell. Стали развиваться новые научные направления: анализ программ на уязвимости, анализ бинарного кода, интеллектуальный анализ текстов. В 2009 году появился долгосрочный партнёр — Samsung, с которым мы организовали совместную лабораторию. Началось сотрудничество с российскими компаниями. Значительно вырос коллектив.

Управление Институтом требовало много времени и внимания, но Виктор Петрович всё успевал. Занимался научным руководством, продолжал преподавание. Он считал, что преподавать должны именно специалисты с практическим опытом, и нёс этот опыт студентам МГУ и МФТИ. Всегда говорил, что главная ценность — люди. Сейчас его ученики уже сами читают учебные курсы и руководят студентами и аспирантами в Институте.

Время показало, что всё, о чём говорил нам Виктор Петрович, — работает. В ИСП РАН удалось создать экосистему, которая обеспечивает постоянную генерацию кадров и инноваций в области системного программирования. Мы развиваемся, растём, создаём новые технологии. За последние пять лет наш коллектив вырос в два раза, а объём финансирования — в три. И в основе каждого нашего достижения — то самое отношение к работе, которому научил нас Виктор Петрович. Как к делу, которым живёшь.

ИСП РАН 2020: **СОБЫТИЯ**

- Расширение 10-летнего сотрудничества с Samsung Electronics. В 2020 г. финансирование совместной лаборатории увеличилось на 20% по сравнению с 2019 г. Началась разработка технологий искусственного интеллекта в программной инженерии и в анализе программ.
- Расширение сотрудничества с Huawei. Начало работ по анализу лучших практик в тестировании и верификации операционных систем и их применению к различным линейкам HarmonyOS, которую разрабатывает компания.
- Начало полномасштабного внедрения технологий жизненного цикла безопасного ПО в российских компаниях: ООО «Код Безопасности», ОАО «РусБИТех», АО «Лаборатория Касперского», Postgres Professional, МВП «СВЕМЕЛ» и др.
- Победа в конкурсе Минобрнауки РФ (совместно с Математическим институтом им. В.А. Стеклова РАН). Совместный проект вошёл в десятку лучших, набрав 95,5 баллов из 100. Работа двух институтов будет посвящена созданию новых методов обеспечения безопасности федеративных центров обработки данных, в том числе объединенных квантовыми коммуникационными линиями.
- Победа в конкурсном отборе лучших проектов по организации научных центров мирового уровня (совместно с Сеченовским университетом и другими организациями). Главная задача ИСП РАН — создание цифровой платформы для сбора, хранения, разметки и анализа больших медицинских данных. Ее основой станут существующие технологии Института, в частности — облачная среда Asperitas.
- Запуск стипендиальной программы в МГУ им. М.В. Ломоносова, МФТИ и ВШЭ. Цель — вовлечение успешных студентов и аспирантов в передовые проектные исследования и разработки в сотрудничестве с индустрией. Размер ежемесячных выплат составляет до 200 тысяч рублей.
- Рост числа студентов и сотрудников почти на 30% по сравнению с 2019 г.
- Открытие четвёртой внешней лаборатории системного программирования под руководством ИСП РАН — в Орловском государственном университете.
- Проведение трёх из пяти конференций в онлайне. Из-за пандемии коронавирусной инфекции три массовых мероприятия были впервые полностью или частично проведены в онлайн-режиме: Весенняя конференция молодых учёных по программной инженерии (SYRCoSE), Научно-практическая конференция OS DAY, Открытая конференция ИСП РАН.

ИСП РАН 2020: РАЗВИТИЕ ТЕХНОЛОГИЙ

В 2020 г. был получен ряд важных научных результатов, которые стали основой новых и развитием существующих технологий, а также получили признание в научном сообществе. В частности, комплекс для анализа свойств бинарного кода ТРАЛ был включён в число лучшихрезультатов работы Академии наук, а президент РАН А.М. Сергеев назвал его важным вкладом в обеспечение кибербезопасности.

В числе главных достижений этого года:

- Инструмент статического анализа Svace стал поддерживать анализ программ на языках Kotlin и Go. Анализ больших проектов ускорился на 35%. Улучшено качество анализа чувствительных потоков данных, доработана поддержка компиляторов C/C++, в особенности Microsoft Visual C++. Поддерживается анализ программ для архитектур RISC-V и PowerPC64, а сам анализатор теперь может работать на серверах с архитектурой ARM64.
- Фреймворк для анализа социальных сетей Talisman трансформировался в масштабную платформу, которая обеспечивает быструю разработку многопользовательских аналитических систем, объединяющих информацию из внутренних баз данных и открытых источников сети Интернет.
- Технологии динамического анализа программ обновлены и объединены в единый программный комплекс ИСП Crusher. Основные инструменты комплекса: ИСП Fuzzer для проведения фаззинг-тестирования и SyDr, отвечающий за автоматическую генерацию тестов для сложных программных систем.
- Дополнен дистрибутив платформы Asperitas, предназначенной для хранения данных и совершения сложных ресурсоёмких вычислений. В него включён Michman — инструмент оркестрации сервисов уровня PaaS в облачной среде для анализа и хранения больших данных, а также решения задач машинного обучения и распределённых задач на кластере. Кроме того, добавлен инструмент Clouni, который позволяет транслировать шаблоны TOSCA Simple Profile YAML в сценарии развёртывания уровня IaaS на базе Ansible.
- Разработана универсальная открытая система Dedoc, предназначенная для извлечения структуры документов и приведения их к единому формату.

- Разработан инструмент Casr, позволяющий автоматически формировать отчёты об аварийных завершениях, возникающих во время эксплуатации и тестирования ПО.
- Обновлена среда анализа бинарного кода ТРАЛ. Добавлена поддержка выявления утечек чувствительных данных по памяти, работающая на основе динамического анализа помеченных данных в трассе.
- Обновлён статический анализатор бинарного кода BinSide. В частности, повысилось качество работы детекторов ошибок и проведена интеграция со средой обратной разработки Ghidra.
- Расширена сфера использования Lingvodoc — системы для совместной многопользовательской документации исчезающих языков. Теперь она используется филологами 29 вузов и НИИ из 16 городов.

Все эти и другие технологии развиваются в рамках комплексных предметно-ориентированных платформ, в числе которых:

- Платформа жизненного цикла безопасного ПО: статические анализаторы Svace и BinSide, комплекс динамического анализа ИСП Crusher, система дедуктивной верификации ключевых компонентов AstraVer Toolset, инструмент формирования отчётов об ошибках Casr, ИСП Обфускатор.
- Платформа создания распределённых систем. Включает в себя Asperitas и другие облачные решения ИСП РАН, а также программный комплекс DigiTEF для создания сложных цифровых моделей промышленных устройств.
- Платформа анализа данных: платформа для обработки данных Talisman, система поиска SciNoon, платформа для извлечения семантики из текста Texterra, система извлечения структуры документов Dedoc.

ТЕХНОЛОГИИ

ASPERITAS И ДРУГИЕ ОБЛАЧНЫЕ РЕШЕНИЯ

Asperitas — платформа, предназначенная для хранения данных и совершения сложных ресурсоёмких вычислений по запросу. Дистрибутив включает в себя одноимённую облачную среду (№5921 в Едином реестре российского ПО), а также PaaS- и IaaS- оркестраторы (Michman и Clouni). В число облачных решений ИСП РАН входит также Fanlight (№6066 в Едином реестре российского ПО) — платформа для организации web-лабораторий.

ОБЛАЧНАЯ СРЕДА ASPERITAS

Облачная среда Asperitas на базе Openstack и Ceph создана на основе совместного проекта с компанией Dell. Предназначена для вычислений с большими доступными ресурсами. Базируется на открытых современных технологиях, которые являются основными для построения больших приватных облачных систем. Подход к развёртыванию облачной среды из локальных источников реализован в виде заранее подготовленной виртуальной машины, обладающей всеми необходимыми инструментами для запуска процесса развёртывания.

Asperitas — это:

- Отчуждаемость решений (возможность воссоздания инфраструктуры в изолированной среде с полным контролем над ней за счёт использования открытых стандартов, свободного ПО и научных разработок ИСП РАН).
- Высокий уровень безопасности (среда построена на базе уменьшенной кодовой базы и использует собственные решения по усилению безопасности).
- Управление виртуальными сетями и вычислительными кластерами с использованием систем Keystone, Neutron, Nova (аналог Amazon EC2).
- Блочное хранение данных, а также расширяемое объектное хранилище на основе распределённой файловой системы Ceph.
- Возможность адаптации под решение конкретных классов задач (решение задач механики сплошных сред, анализ больших данных, анализ программ на уязвимости и др.).

УНИВЕРСАЛЬНЫЙ ОРКЕСТРАТОР MICHMAN

Michman — инструмент оркестрации сервисов уровня PaaS в облачной среде для анализа больших данных, задач машинного обучения, решения распределённых задач на кластере и хранения больших объёмов данных. Поддерживает автоматическое развёртывание кластеров с настроенными системами. Позволяет пользователям создавать кластеры в изолированных проектах и отслеживать актуальную информации о статусе

развёрнутых сервисов и кластеров. Инструмент активно развивается, в будущих версиях планируется добавить поддержку систем Kubernetes и Slurm Workload Manager, а также возможность мониторинга развёрнутых приложений.

Michman предоставляет возможность по запросу развёртывать в облаке виртуальные кластеры с набором сервисов уровня PaaS:

- Кластер для анализа больших данных с полностью настроенными системами Apache Spark, Apache Hadoop, Apache Ignite и Jupyter Notebook, а также с произвольным количеством вычислительных узлов;
- СУБД для хранения больших объемов данных: PostgreSQL, Apache Cassandra, CouchDB, ClickHouse, Redis. Для ряда СУБД поддерживается развёртывание в распределённом режиме.
- Система хранения и обмена файлами NextCloud.

Michman — это:

- Сервис с системой пользователей, изолированных групп и REST API.
- Хранение информации о развёрнутых кластерах, сервисах, их актуальных статусах и точках доступа.
- Хранение шаблонов кластеров, готовых к развертыванию.
- Развёртывание сложных распределённых систем со всеми доступными комбинациями сервисов по запросу.
- Контроль зависимостей между различными сервисами и их версиями.
- Локальное развёртывание без использования сети Интернет.
- Хранение подробной информации о доступных сервисах, их версиях и конфигурируемых параметрах.
- Удобное добавление поддержки новых сервисов при помощи REST API.
- Интеграция с облачными системами виртуализации уровня IaaS.

Вместе с Asperitas и Michman в дистрибутив входит инструмент Clouni (https://github.com/ispras/clouni), который позволяет транслировать шаблоны TOSCA Simple Profile YAML в сценарии развёртывания уровня IaaS на базе Ansible.

ПЛАТФОРМА FANLIGHT

Платформа для организации web-лабораторий Fanlight создана в результате участия ИСП РАН в программе «Университетский кластер» и в международном проекте Open Cirrus (учреждён HP, Intel и Yahoo!). Предназначена для развёртывания SaaS-инфраструктур для вычислительных web-лабораторий средствами Docker Compose. Построена на контейнерных технологиях и предоставляет виртуальные рабочие места в модели DaaS (Desktop as a Service). Доступна для пользователей на сайте fanlight.ispras.ru. Поддерживает только приложения, разработанные для ОС на базе ядра Linux.

Fanlight — это:

- 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. В частности, есть опыт внедрения:
  - в области МСС: 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) или в собственном облачном ЦОД.

ОПЫТ ВНЕДРЕНИЯ

Вычислительный кластер на базе Asperitas используется для анализа информационных потоков платформы Talismanи для работы других 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.

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).

СИСТЕМА ВЕРИФИКАЦИИ ASTRAVER TOOLSET

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).

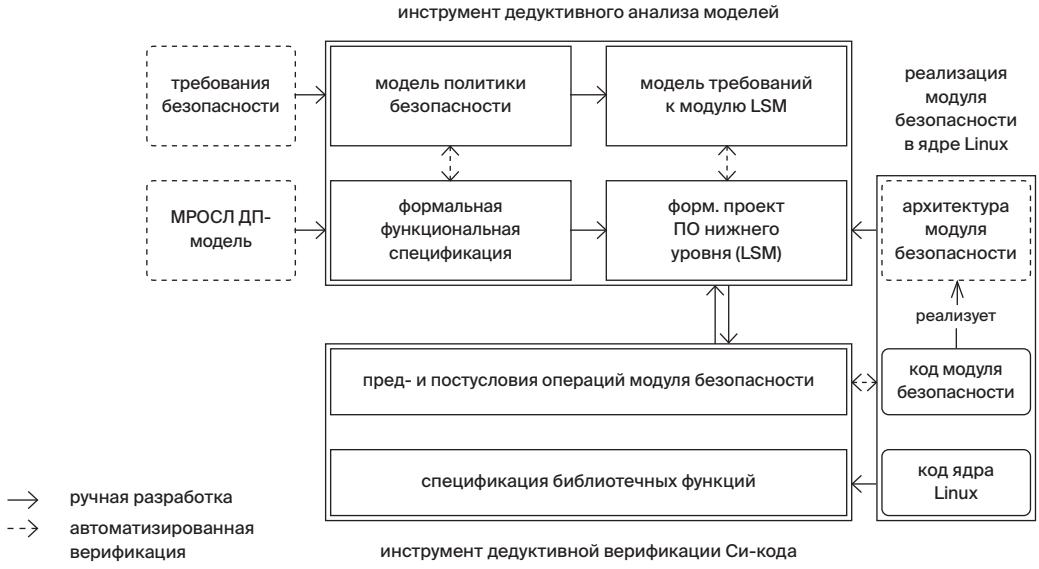
ДЛЯ КОГО ПРЕДНАЗНАЧЕН ASTRAVER TOOLSET?

— 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 ее новых возможностей в ОС Astra Linux Special Edition продолжает верифицироваться с помощью AstraVer Toolset.

СХЕМА РАБОТЫ



BINSIDE: СТАТИЧЕСКИЙ АНАЛИЗАТОР БИНАРНОГО КОДА

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

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

BinSide — платформа для анализа бинарного кода, разрабатываемая на основе фреймворка BinNavi, который переводит ассемблерный код в представление REIL. Данное представление позволяет анализировать код независимо от процессорной архитектуры и операционной системы. Интегрирован с дизассемблерами IDA PRO и Ghidra, которые используются для обратной разработки.

Возможности ядра BinSide:

- Лёгкая расширяемость:
  - детекторы ошибок реализованы в виде подключаемых модулей;
  - используется представление REIL из 17 инструкций без побочных эффектов (каждая ассемблерная инструкция транслируется в набор из REIL-инструкций).
  - возможность разметки для функций источников распространения помеченных данных.
- Поддерживает анализ бинарных файлов и библиотек для архитектур x86-64, ARM и MIPS.
- Поиск дефектов типа: CWE-121 (Stack-based Buffer Overflow), CWE-122 (Heap-based Buffer Overflow), CWE-134 (Use of Externally-Controlled Format String), CWE-415 (Double Free), CWE-416 (Use After Free).
- Ядро анализа позволяет выполнять:
  - анализ значений и указателей;
  - отслеживание помеченных данных, моделей статической и динамической памяти, а также анализ с учётом потока данных и управления;
- поиск ошибок на всех путях (в том числе не покрытых тестированием или динамическим анализом).
- Возможность конвертировать результаты в формат Svace (при наличии отладочной информации) для отображения в веб-интерфейсе в целях навигации по исходному коду;

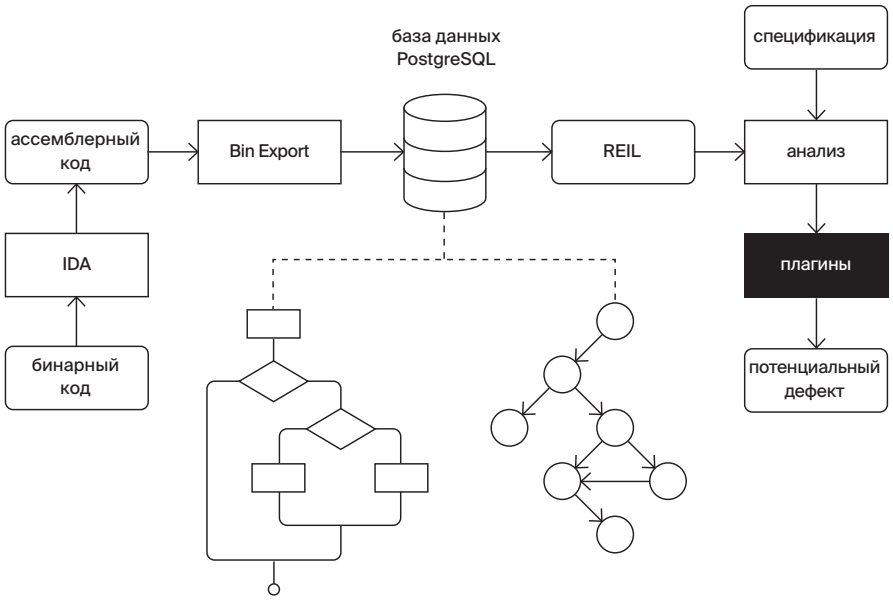
Компоненты BinSide:

- Компонент восстановления семантики функций;
- Компоненты на базе технологии поиска клонов кода:
- Компонент libraryIdentifier:
  - обнаружение использования устаревших версий библиотек;
  - определение нарушений авторских прав.
- Компонент для анализа изменений между версиями программ;
- Компонент переноса разметки имён функций с одного бинарнаго файла на другой.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН BINSIDE?

- Компании, которые нуждаются в тщательной проверке стороннего ПО при отсутствии доступа к исходному коду;
- Разработчики, которым требуется повысить качество работы инструментов динамического анализа за счёт дополнительных данных, полученных с помощью статического анализа.

СХЕМА РАБОТЫ



CASR: ИНСТРУМЕНТ ФОРМИРОВАНИЯ ОТЧЁТОВ ОБ ОШИБКАХ

Casr — это инструмент, позволяющий автоматически формировать отчёты об аварийных завершениях, возникающих во время эксплуатации и тестирования ПО, на основе анализа coredump файлов в ОС Linux. В отчётах содержатся сведения о степени критичности аварийного завершения, а также дополнительная информация, которая помогает установить его причины.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Casr решает те же задачи, что и система с открытым исходным кодом Apport, однако в отличие от неё проводит оценку критичности аварийного завершения, а также предоставляет список открытых файлов и сетевых соединений на момент завершения.

Casr — это:

- Обнаружение критичных аварийных завершений, которые могут привести к перехвату потока управления.
- Классификация аварийных завершений по 23 классам 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.
- Поиск открытых файлов и сетевых соединений, которые могут быть причиной аварийного завершения.
- 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.);
- Отчёты для трудновоспроизводимых ошибок (недетерминированные ошибки, отсутствие возможности настроить правильное окружение и др.);
- Интеграция с системами мониторинга (например, Zabbix), которая позволяет системным администраторам оперативно получать информацию о критичных аварийных завершениях.

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 поставляется в ряд российских компаний и организаций в дополнение к комплексу ИСП Crusher, куда он будет включён в ближайшие 1-2 года.

СХЕМА РАБОТЫ

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, 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 протяженными 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.synchroltd.com/), 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).

ИСП CRUSHER: КОМПЛЕКС ДИНАМИЧЕСКОГО АНАЛИЗА ПРОГРАММ

ИСП Crusher — программный комплекс, комбинирующий несколько методов динамического анализа. Состоит из двух инструментов: ИСП Fuzzer для проведения фаззинг-тестирования и SyDr, отвечающий за автоматическую генерацию тестов для сложных программных систем. В ближайшие 1-2 года в комплекс планируется включить ещё две технологии ИСП РАН: BinSide и Casr. ИСП Crusher позволяет построить процесс разработки всоответствии с ГОСТ Р 56939-2016 и «Методикой выявления уязвимостей и недекларированных возможностей в программном обеспечении» ФСТЭК России.

ИНСТРУМЕНТ ТЕСТИРОВАНИЯ ИСП FUZZER

ИСП Fuzzer — инструмент проведения фаззинг-тестирования, необходимый на всех этапах разработки, тестирования и эксплуатации ПО. Обнаруживает ошибки или закладки 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:

- Осуществление фаззинг-тестирования через различные источники внешних данных (файл, аргументы командной строки, стандартный поток ввода, аргументы переменных окружений, сеть).
- 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 многопоточного анализа как на одной машине, так и на распределённых.
- Поддержка пользовательских плагинов отправки данных по сети (плагины позволяют осуществлять взаимодействие с клиентским или серверным ПО и отправлять мутированные данные).
- Возможность интеграции с рядом необходимых инструментов жизненного цикла разработки безопасного ПО, созданных в ИСП РАН:
  - использование инструментов динамического символьного выполнения SyDr и Anxiety для повышения эффективности фаззинг-тестирования;

  - возможность получать входные данные, на которых проявляются ошибки, размеченные инструментом статического анализа BinSide в автоматическом 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;
- 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).
- Возможность проведения анализа серверного и клиентского ПО, работающего по протоколам с состояниями и без состояний.
- Возможность запуска систем динамического анализа на новых входных данных (Valgrind, DrMemory, QASan).
- Лёгкая расширяемость и добавление новых методов в рамках существующей инфраструктуры; оперативная адаптация под новые задачи.
- Возможность распределения корпуса входных данных между процессами фаззеров для повышения эффективности их работы.
- Оценка критичности найденных аварийных завершений.
- Использование фаззера Radamsa для генерации новых данных.
- Поддержка различных видов инструментации: DynamoRIO, QemuUserMode, статическая инструментация GCC, статическая инструментация LLVM, QemuSystemArm.

ИНСТРУМЕНТ ДИНАМИЧЕСКОГО СИМВОЛЬНОГО ВЫПОЛНЕНИЯ SYDR

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. 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 — это:

- Инвертирование на конкретном пути выполнения всех условных переходов, которые зависят от входных данных. Реализована возможность параллельного инвертирования.
- Интеграция с ИСП Fuzzer для инвертирования, что решает проблемы при прохождении переходов, зависящих от сравнения с константами.
- Решение задач обратной разработки. Помощь аналитику в достижении интересующей его точки в программе. Получение трассы инструкций, которые зависят от входных данных.

- Поддержка различных источников внешних данных программы (файлы, сетевые сокеты, переменные окружения, стандартный поток ввода, аргументы командной строки).
- Предикаты безопасности. Генерация входных данных, приводящих к проявлению дефекта (деление на ноль, разыменование нулевого указателя, выход за границы буфера).
- Символьное выполнение многопоточных программ.
- 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.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН ИСП CRUSHER?

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

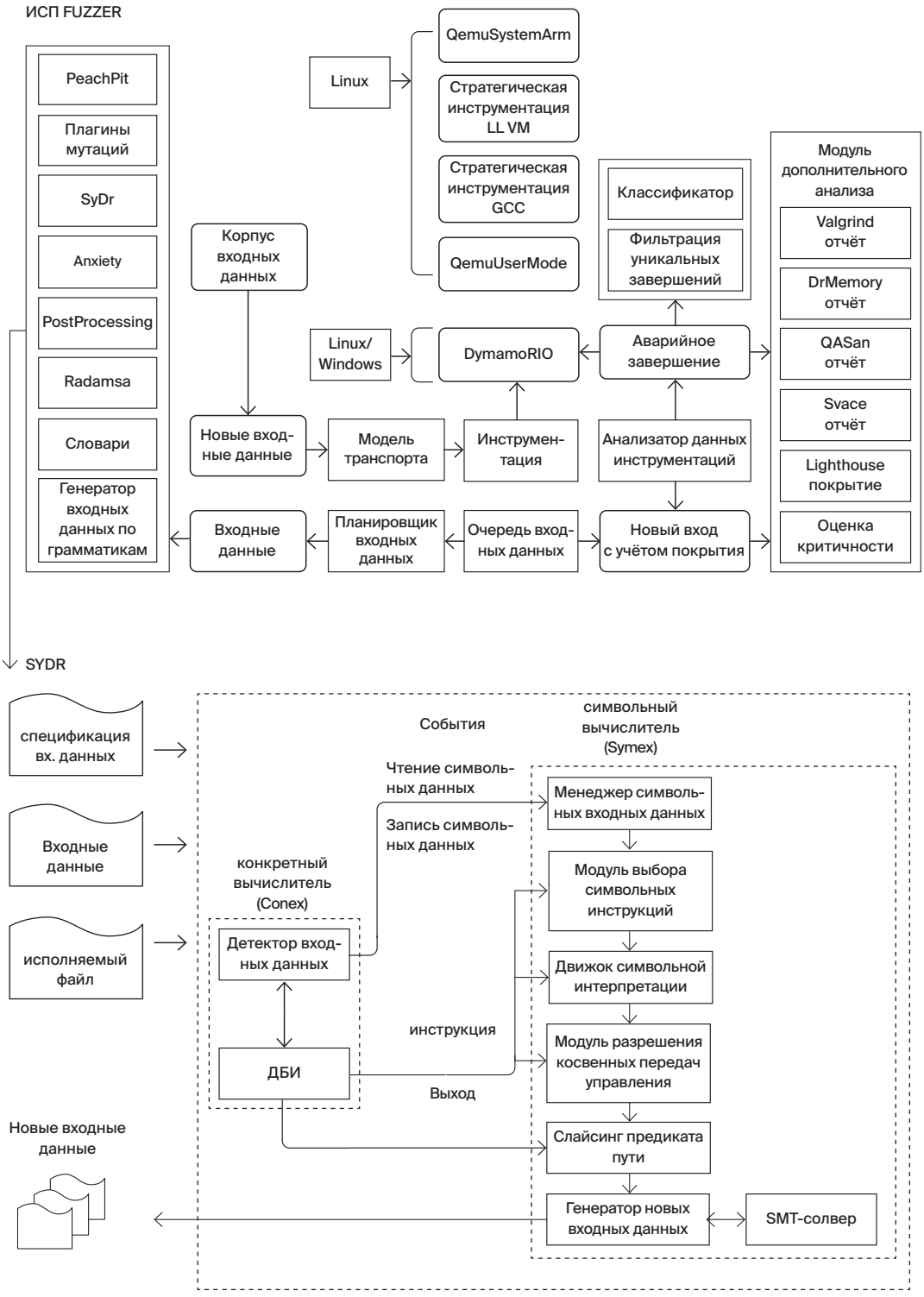
СИСТЕМНЫЕ ТРЕБОВАНИЯ

Поддержка ОС семейства Linux и Windows. 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 20 компаниях и лабораториях, в том числе в ОАО «РусБИТех», Postgres Professional, ООО «Код Безопасности», МВП «СВЕМЕЛ» и др.

СХЕМА РАБОТЫ ИСП CRUSHER



DEDOC: СИСТЕМА ИЗВЛЕЧЕНИЯ СТРУКТУРЫ ДОКУМЕНТОВ

Dedoc — универсальная открытая система для приведения документов к единому формату. Автоматически извлекает логическую структуру, таблицы 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*, csv, txt, json). Позволяет добавлять плагины, в частности, пакет расширений Docreader для работы с изображениями (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:

- Расширяемость за счет 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.).
- Добавление правил корректировки списков, неправильно составленных в документе (с опечатками).
- Извлечение табличной информации из xml формата doc*.

Docreader — это:

- Работа с изображениями сканированных документов различного назначения (технические задания, нормативно-правовые акты, научные отчёты и статьи) и гибкая настройка под документы новой предметной области.
- Работа с pdf-документами — как с текстовым слоем, так и без.
- Распознавание физической структуры и текста ячеек сложных многостраничных таблиц с границами на изображениях с помощью методов контурного анализа; определение ориентации таблиц на изображении.

– 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 и извлечения иерархической структуры на основе классификации строк извлеченных признаков из изображений документов.

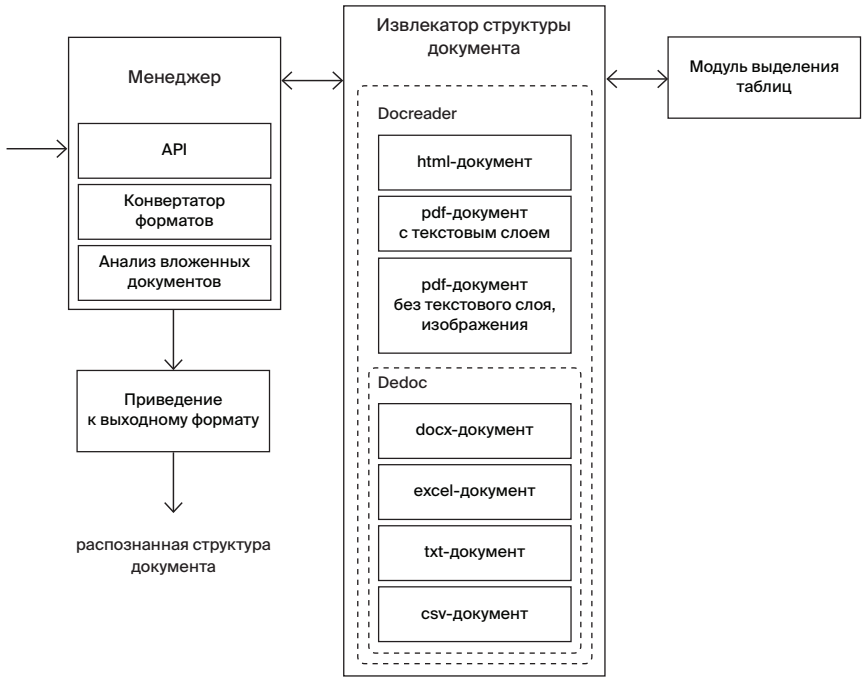
ДЛЯ КОГО ПРЕДНАЗНАЧЕНА СИСТЕМА DEDOC?

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

ПОДДЕРЖИВАЕМЫЕ ФОРМАТЫ

Русский и английский.

СХЕМА РАБОТЫ



ПЛАТФОРМА ДЛЯ СОЗДАНИЯ ЦИФРОВЫХ ДВОЙНИКОВ DIGITEF

DigiTEF — программный комплекс на базе OpenFOAM и утилит других открытых проектов, а также уникальных модулей и библиотек ИСП РАН. Платформа позволяет решать прикладные задачи газовой динамики, аэродинамики, гидродинамики и акустики. Предназначена для создания сложных цифровых моделей промышленных устройств. Включена в Единый реестр российского ПО (№5377).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Платформа решает те же задачи, что и мировые аналоги. 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. Вокруг платформы DigiTEF сформировано сообщество инженеров, исследователей и разработчиков промышленных проектов.

DigiTEF — это:

- открытый исходный код (позволяет контролировать и адаптировать реализованные алгоритмы);
- развитие параллельно с веткой OpenFOAM+;
- наличие средств автоматизации вычислений и интеграции моделей для комплексного исследования технических объектов;
- возможность разработки дополнительных компонентов в соответствии с конкретными требованиями.

КОМПЛЕКС СОСТОИТ ИЗ ДВУХ ОСНОВНЫХ БЛОКОВ:

1 OpenDTEF — ядро программного комплекса на основе OpenFOAM. Содержит основные алгоритмы, процедуры и функции, а также набор сторонних библиотек на языке C++. Находится в открытом доступе (https://github.com/unicfdlab) и состоит из следующих компонентов:

- Компонент инструментов для моделирования сжимаемых течений;
- Компонент расширенных настроек расчетного случая на основе swak4Foam;

– Компонент для параметризации на базе Python. Позволяет проводить автоматизацию расчётных случаев, а также осуществлять интеграцию в DigiTEF программных комплексов Salome, Paraview и CodeAster.

2 Компоненты, разработанные в ИСП РАН:

– компонент анализа данных для визуализации и извлечения информации. Предназначен для анализа результатов и построения моделей пониженной размерности с использованием методов обработки данных (FFT, POD, DMD, Hilbert transformations);
– компонент для расчёта сжимаемых течений на основе квазигазодинамических (КГД) уравнений, позволяющих использовать процедуру пространственно-временного осреднения для определения основных газодинамических величин (плотности, скорости и температуры и др.);
– для расчёта несжимаемых течений на основе КГидД-уравнений. Компонент применим в задачах океанологии, конвекции и дозвуковых течений;
– для расчёта несжимаемых и сжимаемых течений на основе гибридного алгоритма Pimple и Курганова-Тадмора;
– для расчёта дозвуковых турбулентных течений с использованием гибридного URANS/LES подхода и низко диссипативных численных схем;
– для проведения акустического анализа (в компоненте реализованы аналогии Керла и Фокс Уильямса – Хокинга);
– моделирование роста льда.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН DIGITEF?

DigiTEF предназначен для применения на предприятиях ресурсоёмких отраслей промышленности. Использование цифровых моделей позволяет повысить эффективность проектирования, а также снизить стоимость и сложность реализации индустриальных проектов.

ОПЫТ ВНЕДРЕНИЯ

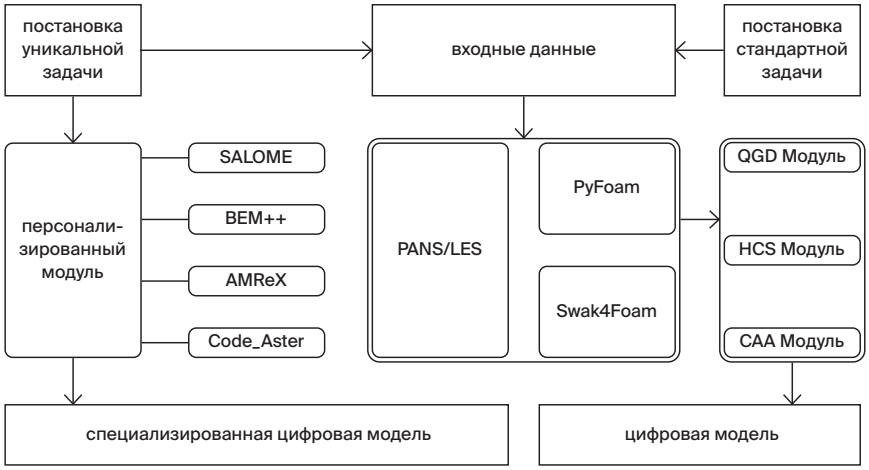
DigiTEF используется в ряде проектов в области ветроэнергетики, космонавтики, авиации, металлургии, а также в нефтегазовой отрасли. Открытые версии модулей DigiTEF успешно применяются в академических, образовательных и промышленных учреждениях мира: Institut Pprime (Франция), Korea Atomic Energy Research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

СИСТЕМНЫЕ ТРЕБОВАНИЯ

ОС Linux. Могут также использоваться другие ОС, которые поддерживают виртуальную машину Oracle VirtualBox (на Microsoft Windows 10 — с применением оболочки Bash). В случае их использования падение производительности не превышает 5%. Требуемая оперативная память — не менее 16 Гб.

DigiTEF поддерживает параллельные вычисления, что существенно ускоряет его работу. Кроме того, поддерживается возможность использования высокопроизводительных систем вычислений (суперкомпьютеров и кластеров) для ускорения расчётов. Проверенное количество вычислительных ядер — до 1536.

СХЕМА РАБОТЫ



Digital Test Facility

KLEVER: ТЕХНОЛОГИЯ ВЕРИФИКАЦИИ МОДЕЛЕЙ КРУПНЫХ ПРОГРАММНЫХ СИСТЕМ

Klever — система верификации моделей, которые генерируются на основе исходного кода крупных программных систем, разработанных на языке программирования Си. Klever позволяет автоматизированным образом проверять различные требования безопасности и надёжности.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Klever базируется на научных исследованиях в области выполнения полностью автоматического доказательства корректности моделей программ, извлекаемых из исходного кода (в том числе без участия пользователя). В основе системы лежат методы для покомпонентной верификации исходного кода программных систем размером в сотни тысяч и миллионы строк кода на языке Си. Klever выложен в открытый доступ (https://forge.ispras.ru/projects/klever).

Klever — это:

- Высокоточный консервативный анализ исходного кода индустриального программного обеспечения (выявление всех возможных ошибок искомых видов и доказательство корректности программ при явно заданных предположениях).
- Проверка расширяемого набора требований к программе (проверка правил безопасного программирования на языке Си и корректности использования программных интерфейсов, специфичных для проверяемых программ).
- Масштабируемость. Модульная верификация программ позволяет применять наиболее точные методы для анализа больших объёмов исходного кода (в частности, методы верификации моделей и символьного выполнения).
- Подробные сведения об обнаруживаемых ошибках. Система верификации не только указывает на места ошибок в исходном коде, но также предоставляет последовательности выражений и значения переменных и аргументов функций для их воспроизведения.
- Удобный многопользовательский веб-интерфейс для проведения верификации, а также для выполнения экспертной оценки результатов верификации;

– Возможность адаптации системы верификации под конкретные нужды заказчиков. Разработка наборов спецификаций для обнаружения нарушений специфичных для программ требований, включая моделирование специфичных окружений целевых программ.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН KLEVER?

– Компании, нацеленные на разработку ПО с высоким уровнем надёжности и безопасности;
– Испытательные лаборатории.

ОПЫТ ВНЕДРЕНИЯ

Klever разработан в рамках работы Центра верификации ОС Linux (http://linuxtesting.org/), организованного на базе ИСП РАН при поддержке The Linux Foundation. В настоящее время система верификации в основном применяется для различных операционных систем.

Для демонстрации возможностей Klever были выполнены работы по верификации драйверов устройств ядра операционной системы Linux. В результате удалось обнаружить более 350 ошибок, подтверждённых разработчиками: ошибки выхода за границу буфера, разыменование нулевого указателя, использование неинициализированной памяти, повторное или некорректное освобождение памяти, состояния гонки и взаимные блокировки, утечки специфичных ресурсов ядра Linux, некорректные вызовы функций в зависимости от контекста, некорректная инициализация структур данных ядра Linux.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

Ubuntu 18.04, 16 гигабайт оперативной памяти, от 100 Гб свободного места на диске.

СХЕМА РАБОТЫ

LINGVODOC: ВИРТУАЛЬНАЯ ЛАБОРАТОРИЯ ДЛЯ ДОКУМЕНТАЦИИ ИСЧЕЗАЮЩИХ ЯЗЫКОВ

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;
- 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.

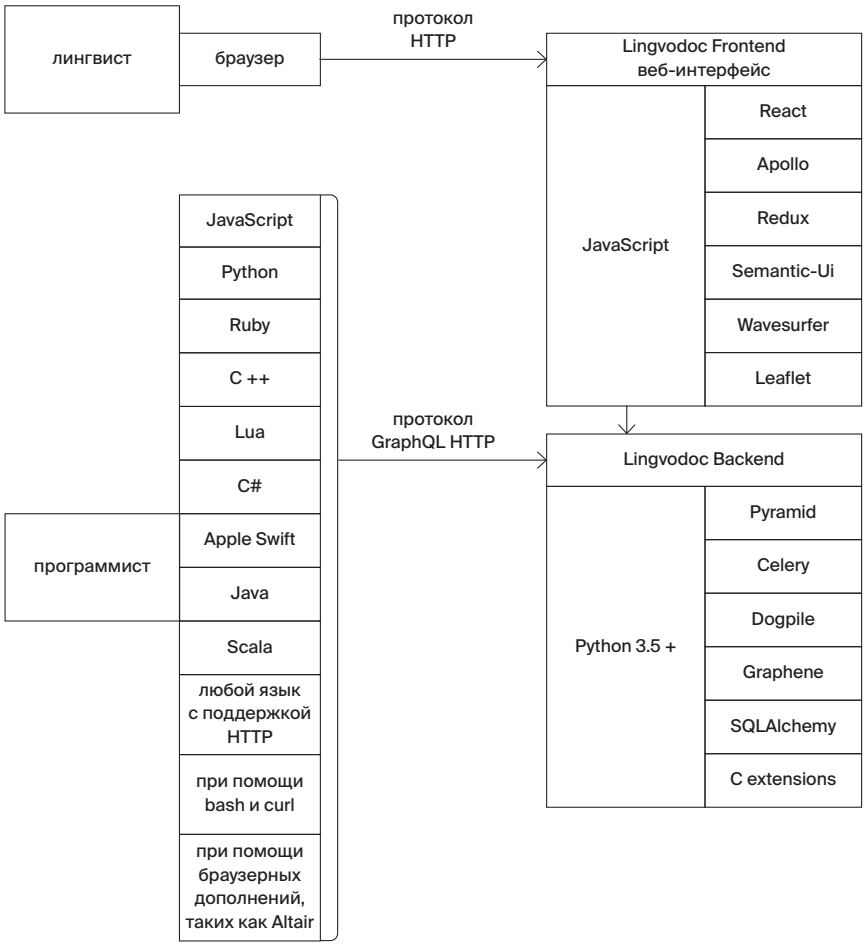
ДЛЯ КОГО ПРЕДНАЗНАЧЕН LINGVODOC?

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.

СХЕМА РАБОТЫ



MASIW: ПОДДЕРЖКА 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 реального времени).
- 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

ГЕНЕРАТОР ТЕСТОВЫХ ПРОГРАММ MICROTESK

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 (GenesysPro и RAVEN), однако отличается от них повышенной производительностью и удобством использования, а также распространением по лицензии открытого исходного кода.

Выложен в открытом доступе на сайте ИСП РАН: https://forge.ispras.ru/projects/microtesk. Описание технологии доступно 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 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 счет повторного использования).
- Широкий набор поддерживаемых архитектур микропроцессоров:
    - поддержка особенностей 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.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

ОС Windows или ОС на базе ядра GNU/Linux, Java 8.

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

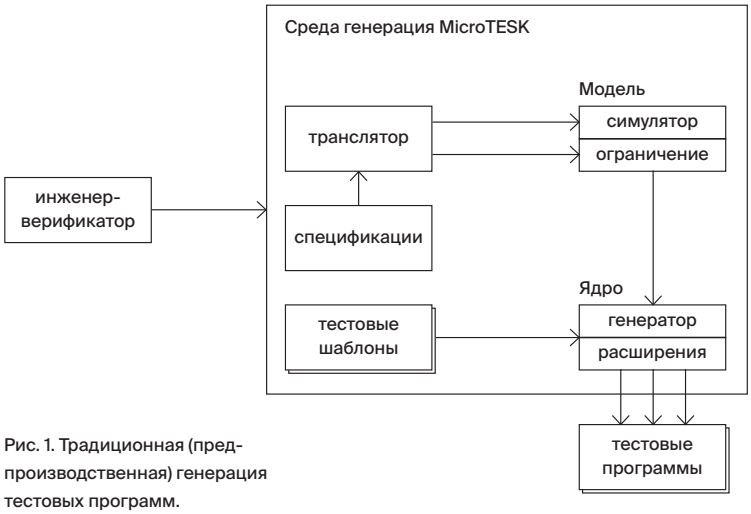


Рис. 1. Традиционная (предпроизводственная) генерация тестовых программ.

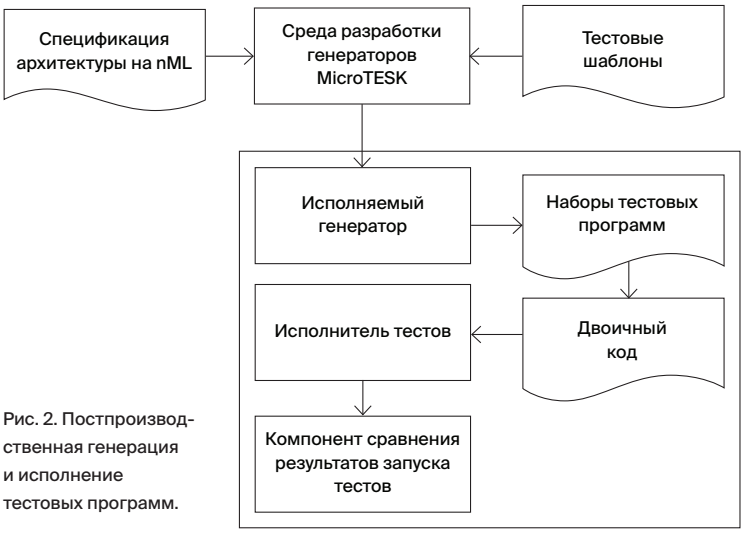


Рис. 2. Постпроизводственная генерация и исполнение тестовых программ.

ИСП ОБФУСКАТОР

ИСП Обфускатор — 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

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

Обфускатор — 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.). На базе компилятора GCC реализован прототип CFI, который показал 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 Обфускатора 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 ОБФУСКАТОР?

— Разработчики специализированных дистрибутивов операционных систем;
— Разработчики прикладного ПО.

ОПЫТ ВНЕДРЕНИЯ

ИСП Обфускатор внедрен в ОС «Циркон», которую используют МИД и Пограничная служба ФСБ России.

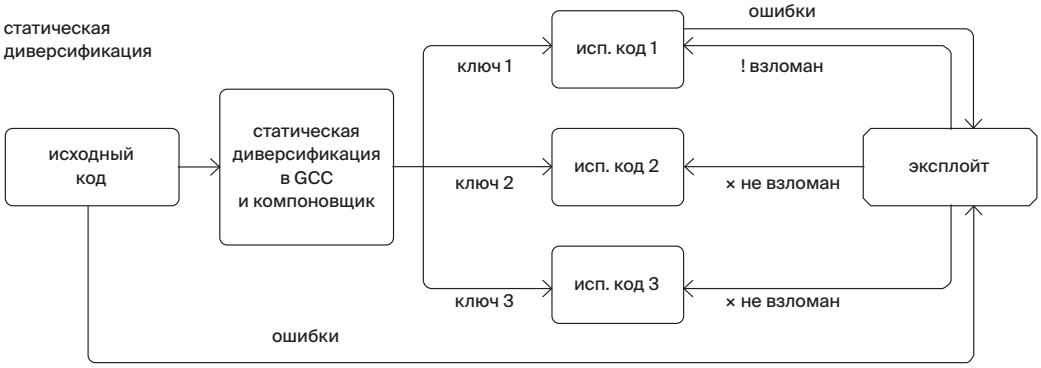
СИСТЕМНЫЕ ТРЕБОВАНИЯ

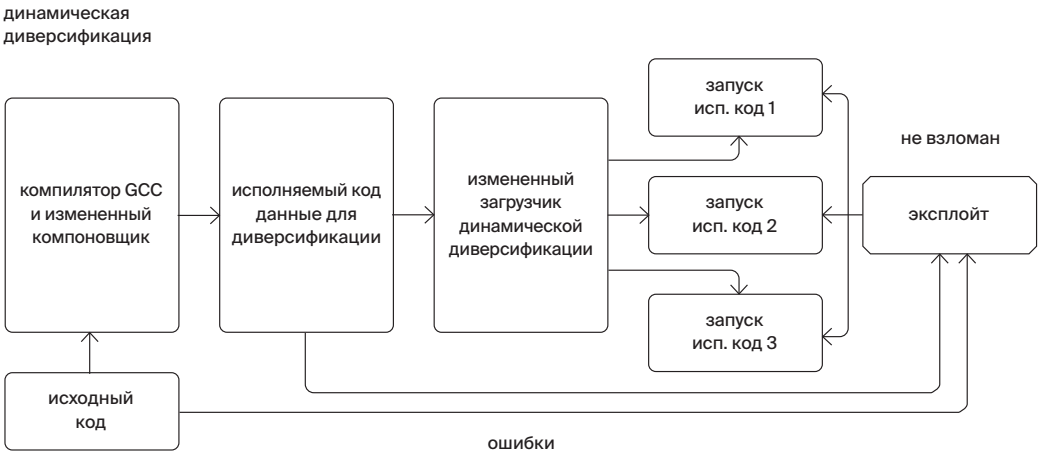
Обфускатор — универсальный продукт, который можно адаптировать под любые системные требования. В настоящее время основная версия работает в ОС на базе ядра Linux (начиная с версии 2.6) с поддержкой архитектуры Intel x86/x86-64.

СХЕМА РАБОТЫ

обычная сборка

статическая диверсификация




динамическая
диверсификация
компилятор GCC
и измененный
компоновщик
исполняемый код
данные для
диверсификации
измененный
загрузчик
динамической
диверсификации
запуск
исп. код 1
запуск
исп. код 2
запуск
исп. код 3
не взломан
эксплойт
исходный
код
ошибки

СИСТЕМА АНАЛИЗА СЕТЕВОГО ТРАФИКА PROTOSPHERE

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) 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+).
- Быстрое расширение списка поддерживаемых протоколов:
  - API доступ к результатам разбора;
  - локализация ошибок разбора;
  - возможность отладки разрабатываемого модуля на потоке, позволяющая существенно ускорить поддержку новых протоколов.
- Поддержка двух режимов работы: на потоке и в отложенном режиме.

- Продвинутый графический интерфейс, позволяющий выбирать наиболее удобный вариант представления результатов проводимого анализа.
- Ускоренная кастомизация благодаря универсальности внутреннего представления:
  - поддержка новых протоколов;
  - извлечение новых типов данных;
  - настройка формата выдачи результатов анализа;
- Адаптация под сетевой канал и доступные вычислительные ресурсы: гибкая система конфигурирования позволяет находить баланс между детализацией/точностью анализа и потребляемыми ресурсами.

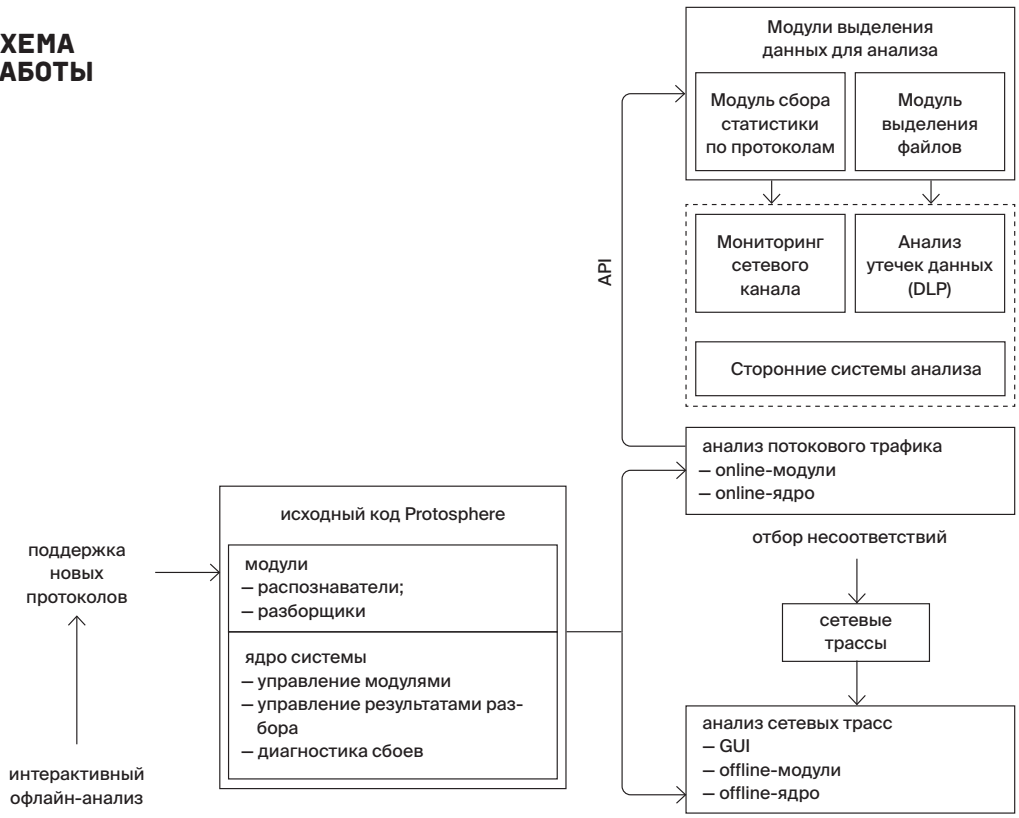
ДЛЯ КОГО ПРЕДНАЗНАЧЕНА PROTOSPHERE?

- 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).
- 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.
Платформы: ОС Windows, ОС на базе ядра Linux.

СХЕМА РАБОТЫ



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 Intel 64, ARM и Thumb, MIPS, PowerPC и др.). Реализует отладку по удаленному GDB-протоколу 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, что упрощает отладку ошибок, связанных с параллельной работой приложения (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, требующее взаимодействия с удалённой системой в режиме реального 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 C-подобного языка описания семантики инструкций;
  - Система автоматического первичного тестирования виртуальной машины. Для работы системы требуются только утилиты GNU Binutils и компилятор языка C;
  - 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 машины;
  - 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.

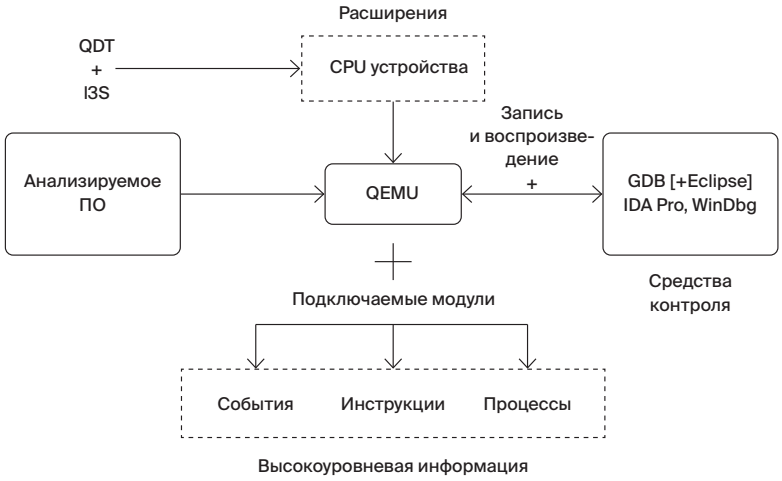
ПОДДЕРЖИВАЕМЫЕ ГОСТЕВЫЕ СРЕДЫ

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-4.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



RETRASCOPE: ИНСТРУМЕНТ СТАТИЧЕСКОГО АНАЛИЗА HDL-ОПИСАНИЙ

Retrascope — инструмент функциональной верификации модулей цифровой аппаратуры. Retrascope предоставляет автоматизированные средства анализа кода, извлечения формальных моделей и генерации функциональных тестов. В качестве входных данных инструмент принимает описания модулей цифровой аппаратуры на синтезируемых подмножествах языков Verilog и VHDL, а также спецификации поведения.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Retrascope — открытый инструмент функциональной верификации модулей цифровой аппаратуры. Инструмент реализует ряд методов извлечения и анализа формальных моделей, а также генерации функциональных тестов. Модульная архитектура Retrascope позволяет разрабатывать гибридные техники верификации HDL-описаний за счёт комбинирования различных средств анализа формальных моделей. Retrascope доступен на сайте ИСП РАН: https://forge.ispras.ru/projects/retrascope.

Retrascope — это:

- Извлечение формальных моделей из исходного кода:
  - граф потока управления;
  - решающая диаграмма охраняемых действий;
  - высокоуровневая решающая диаграмма;
  - расширенный конечный автомат.
- Генерация функциональных тестов:
  - случайные тесты;
  - выявление недостижимого кода;
  - выявление типовых ошибок;
  - проверка пользовательских свойств.
- Проверка формальных моделей (model checking) на соответствие спецификациям:
  - PSL;
  - SystemVerilog Assertions.
- Графический интерфейс на основе Eclipse IDE (также доступен интерфейс командной строки):
  - запуск инструмента с параметрами;
  - визуализация извлеченных моделей (Zest, GraphML).

- Открытый исходный код (лицензия Apache License Version 2.0);
- Расширяемость на уровне исходного кода:
  - добавление новых моделей;
  - расширение набора средств анализа.
- Открытые интерфейсы взаимодействия позволяют использовать различные средства для достижения целей анализа и верификации без изменения кода инструмента:
  - SMT-решатели — язык SMT-LIB v2;
  - Средства проверки моделей — язык SMV.
  - Функциональные тесты — языки VHDL и Verilog, формат VCD.

ДЛЯ КОГО ПРЕДНАЗНАЧЕН RETRASCOPE?

- Компании, занимающиеся проектированием цифровой аппаратуры;
- Коллективы, проводящие исследования в области функциональной верификации цифровой аппаратуры.

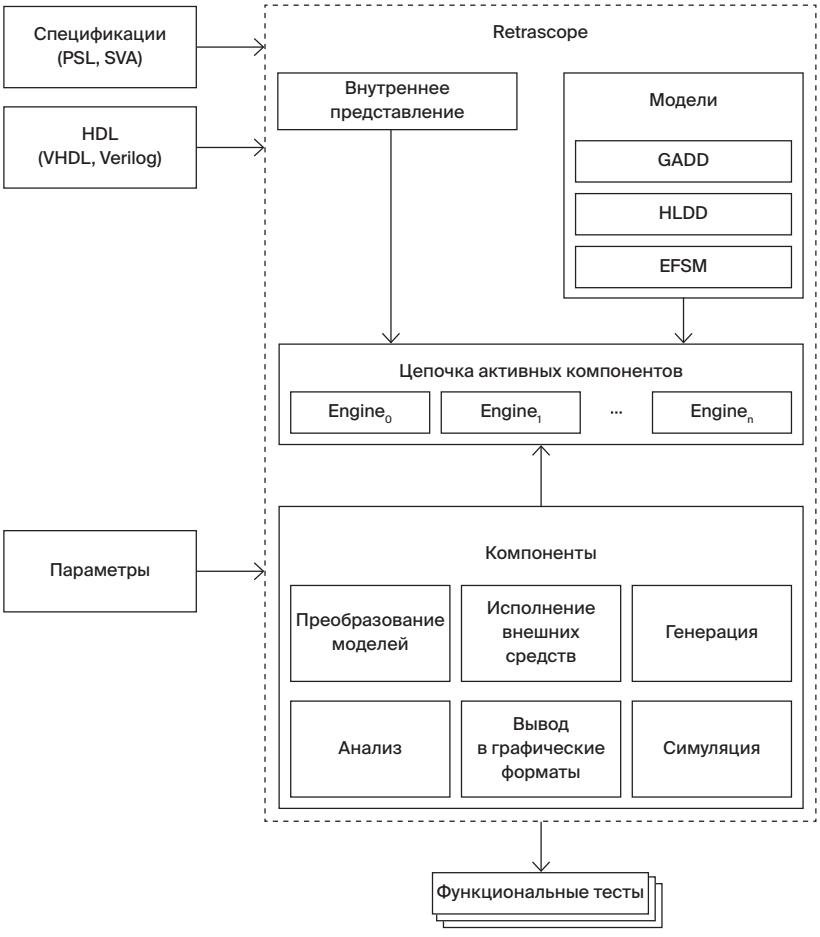
ОПЫТ ВНЕДРЕНИЯ

Инструмент находится на стадии исследовательского прототипа, ведётся разработка.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

Программное обеспечение: ОС Windows или ОС на базе ядра GNU/Linux, Java 8.

СХЕМА РАБОТЫ



СИСТЕМА ИССЛЕДОВАТЕЛЬСКОГО ПОИСКА SCINOON

SciNoon — система совместного исследовательского поиска научных статей. Позволяет группе исследователей быстро погружаться в новую предметную область и находить ответы на свои вопросы, а затем отслеживать новые публикации по изучаемой тематике.

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

SciNoon — инновационная система, созданная с целью оптимизации длительной командной работы с научными публикациями. Статьи в SciNoon можно добавлять как из широко известных поисковых систем и электронных библиотек (Google Scholar, arxiv.org, Semantic Scholar, PubMed), так и с помощью загрузки PDF-файлов. Уникальная особенность — графические карты исследований, на которые все члены группы могут добавлять найденные ими статьи.

SciNoon — это:

- Общее рабочее место для совместной обработки публикаций.
- Возможность масштабирования карты исследований для управления степенью подробности отображаемой информации о статьях.
- Нормализация и дедупликация метаданных загружаемых статей благодаря внутренней базе данных. Построение связей между статьями и авторами.
- Классификация контекстов цитирований в один из пяти классов (с точки зрения цели цитирования):
  - Background (цитируемая статья содержит общую информацию, относящуюся к области исследования в рассматриваемой статье);
  - Use (рассматриваемая статья использует методы, данные и т.д. из цитируемой статьи);
  - Compare (рассматриваемая статья указывает на различия/сходства с цитируемой статьёй);
  - Extend (рассматриваемая статья продолжает развитие методов из цитируемой статьи);
  - Weak (рассматриваемая статья критикует цитируемую статью, указывает на ошибки авторов).
- Возможность находить релевантные исследованию статьи без использования поиска по ключевым словам (благодаря встроенной рекомендательной системе);
- Настройка собственного перечня вопросов, на которые надо искать ответы в статьях. В зависимости от полученных ответов можно по-разному отображать статью на карте исследований;
- Объединение близких статей в кластеры;

- Возможность получать уведомления о действиях каждого исследователя в команде, а также оперативно обмениваться мнениями и помогать друг другу;
- Анализ всех собранных ответов на вопросы с помощью встроенной табличной формы представления, а также экспорт в формате CSV (если требуется более сложная обработка);
- Отслеживание новых статей по рассмотренной тематике после завершения исследования и быстрая актуализация первоначально полученных результатов.

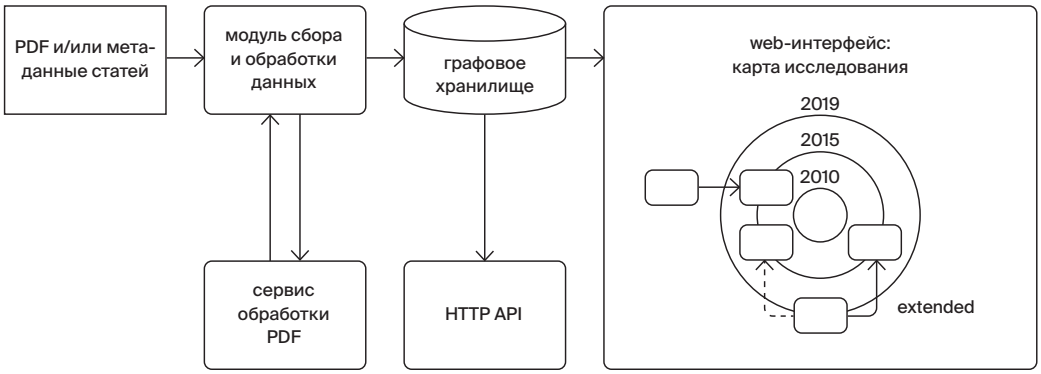
ДЛЯ КОГО ПРЕДНАЗНАЧЕН SCINOON?

- Сотрудники R&D отделов корпораций, которым нужно быстро найти решение возникшей научной задачи;
- Сотрудники научно-исследовательских институтов, нуждающиеся в инструменте для командной работы;
- Преподаватели и студенты ВУЗов, занимающиеся исследовательским поиском для подготовки научных работ.

ОПЫТ ВНЕДРЕНИЯ

SciNoon используется в ИСП РАН при проведении исследований и при руководстве студентами.

СХЕМА РАБОТЫ



СТАТИЧЕСКИЙ АНАЛИЗАТОР SVACE

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 и Go — в предварительной версии. Включён в Единый реестр российского ПО (№4047).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Svace — постоянно развивающийся инновационный продукт, основанный на многолетних исследованиях. Объединяет ключевые качества иностранных аналогов (Coverity Scan Static Analysis, Fortify Static Code Analyzer, Klocwork Static Code Analysis) 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);
  - полное покрытие всех путей с учетом связей между функциями для поиска сложных ошибок;
  - 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 Android 6 из 8 миллионов строк занимает 5-6 часов);
  - поддержка не только полного, но и инкрементального анализа системы (подразумевает быструю повторную проверку недавно измененного кода).
- Удобный интерфейс просмотра предупреждений:
  - подробное описание ошибок с навигацией по коду;
  - разделение срабатываний на истинные и ложные;
  - миграция результатов между запусками и сокрытие ложных срабатываний.
- Ускоренная кастомизация (конфигурация существующих детекторов, а также написание индивидуальных, доступных только данному заказчику; создание специфических интерфейсов);
- 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 (при необходимости

сертификации ПО для использования на территории России), в том числе полная совместимость с новой «Методикой выявления уязвимостей и недекларированных возможностей в программном обеспечении» ФСТЭК России (при необходимости сертификации ПО для использования на территории России).

ДЛЯ КОГО ПРЕДНАЗНАЧЕН SVACE?

- Компании, нацеленные на разработку ПО с высокой степенью надёжности и безопасности.
- Компании, которые нуждаются в сертификации разрабатываемого ПО.
- Испытательные 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 30 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: ОС на базе ядра Linux (начиная с версии 3.13), ОС Windows (начиная с Windows 7 SP 1) и WSL (версий 1 и 2).
- Архитектуры: Intel x86/x86-64, ARM/ARM64, MIPS/MIPS64, Power PC/Power PC 64, RISC-V 32/64, Hexagon (частично — AEON, TriCore, HIDSP, OpenRISC для компилятора GCC).

ПОДДЕРЖИВАЕМЫЕ КОМПИЛЯТОРЫ

Для C/C++ (версий до C++17): GCC (GNU Compiler Collection), Clang (LLVM compiler), Microsoft Visual C++ Compiler, RealView/ARM Compilation Tools (ARMCC), Intel C++ Compiler, Wind River Diab Compiler, 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 Family, Samsung CalmSHINE16 Compilation Tools, Texas Instruments TMS320C6* Optimizing Compiler и др.

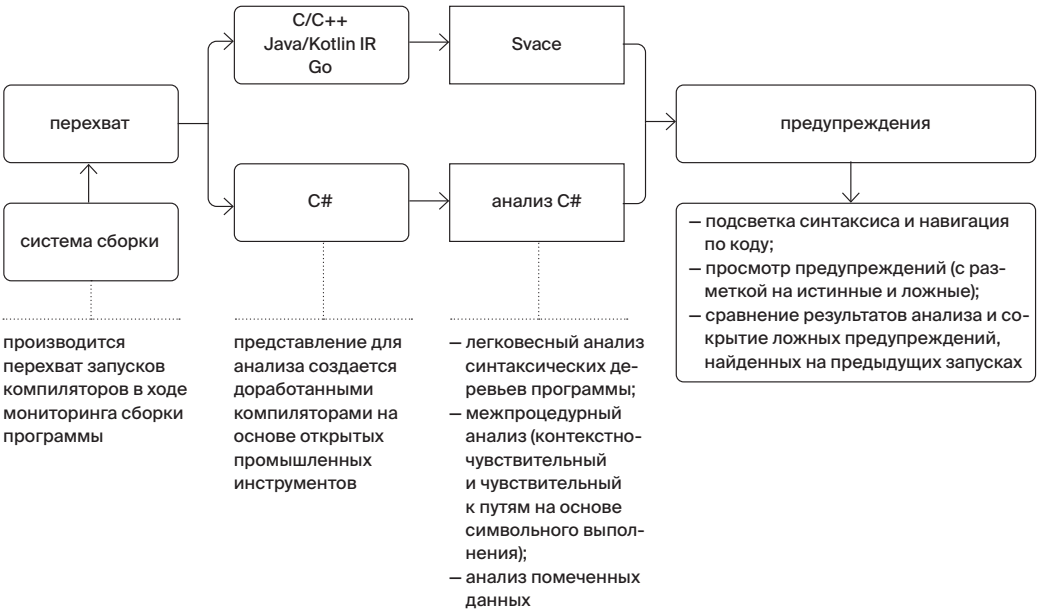
Для C# (версий до C# 7): Roslyn, Mono.

Для Java (версий до Java 11): OpenJDK Javac Compiler, Eclipse ECJ compiler.

Для Kotlin: Kotlin 1.4.

Для Go: Go 1.14.

СХЕМА РАБОТЫ



ПЛАТФОРМА ДЛЯ ОБРАБОТКИ ДАННЫХ TALISMAN

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 аналитических систем, объединяющих информацию из внутренних баз данных и открытых источников сети Интернет (в том числе из социальных сетей).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Talisman объединяет компоненты для работы с большими данными. В качестве базовых сервисов использует такие технологии ИСП РАН, как Dedoc (система извлечения структуры документов) и Texterra (платформа для извлечения семантики из текста). Сопоставим с лучшими мировыми аналогами (Palantir Gotham и IBM Watson Content Analytics). Преимущество — автоматизация рутинных процессов с помощью последних научных достижений (сокращает затраты на аналитиков).

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 даных. Инструменты анализа представляют собой Docker-контейнеры с программным интерфейсом под управлением системы «Talisman.Поток» (№6045 в Едином реестре российского ПО). На выходе данные сохраняются в файлы на жёстких дисках или в СУБД (PostgreSQL, ElasticSearch, Cassandra и др.). В качестве базовых сервисов используются система распознавания текста на изображениях Tesseract и собственные разработки ИСП РАН.
  - 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.

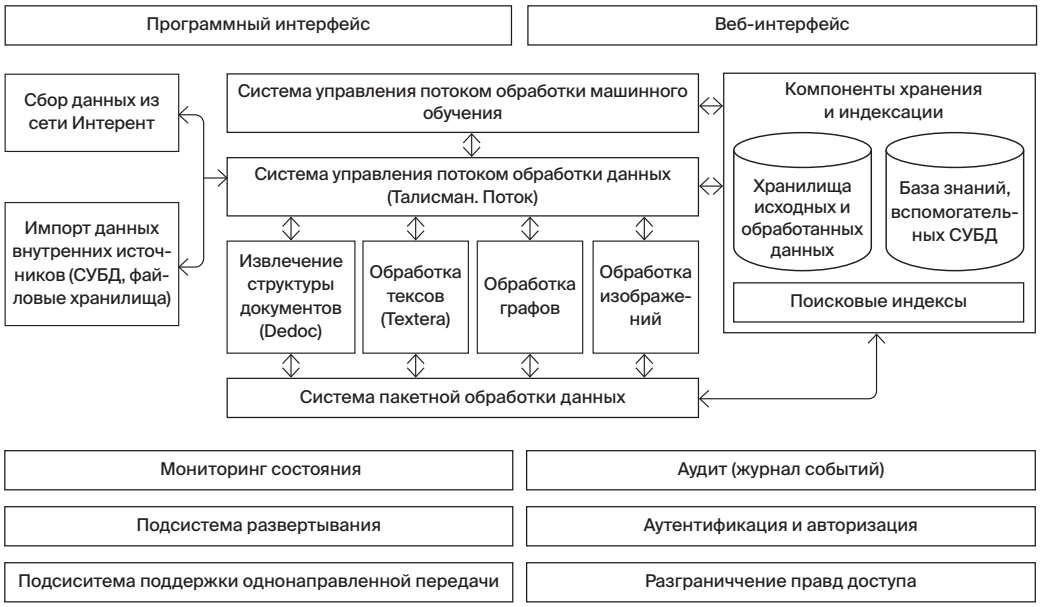
ОБЛАСТИ ПРИМЕНЕНИЯ

- Автоматизация построения базы знаний по интересующей предметной области и обеспечение постоянного мониторинга новой информации об объектах интереса.
- Проведение конкурентной разведки по открытым данным (OSINT).
- 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.
- Оптимизация управления персоналом (эффективный подбор сотрудников, верификация анкетных данных, выявление скрытой деятельности, помощь в разработке систем мотивации).
- Объективная оценка эффективности деятельности, а также тестирование стратегий на целевой аудитории в целях получения обратной связи.
- Выявление и управление точками социального напряжения; обнаружение и своевременное предупреждение эскалации конфликтов.

ИСПОЛЬЗУЕМЫЕ ЯЗЫКИ

В настоящее время Talisman использует языки, распознаваемые анализатором Texterra (русский и английский).

СХЕМА РАБОТЫ



БАЗОВЫЙ СЕМАНТИЧЕСКИЙ АНАЛИЗАТОР TEXTERRA

Texterra — масштабируемая платформа для извлечения семантики из текста. Базовый комплекс технологий для создания многофункциональных прикладных приложений. Анализирует тексты с помощью выделения концептов. Включена в Единый реестр российского ПО (№4048).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

Texterra осуществляет уникальный анализ русскоязычных текстов на основе выделения концептов, а не только слов. Отличается от иностранных аналогов преимущественным вниманием к русскому языку. Базируется на результатах фундаментальных исследований и предоставляет возможность интеграции с поисковой системой Elasticsearch, существенно расширяя ее возможности. Удачное сочетание технологий позволяет платформе конкурировать с проектами уровня IBM Watson Natural Language Understanding.

Texterra — это:

- Высокая скорость обработки текста (морфологический анализ — 69 000 слов в секунду, синтаксический — 39 100 слов/сек, разрешение кореферентности — 10 100 слов/сек, полный разбор текста — приблизительно 13 600 слов/сек);
- Максимальное внимание к русскому языку (в отличие от аналогичных проектов spaCy и UDPipe, а также IBM Watson Natural Language Understanding, который не поддерживает анализ эмоций и концептов в русскоязычных текстах);
- Большой объем знаний (более 7 миллионов понятий);
- Построение базы знаний без привлечения экспертов (автоматическое пополнение с помощью Wikipedia, MediaWiki, Linked Open Data и др.);
- Масштабируемость как по скорости обработки текстов, так и по объему знаний (с помощью Apache Ignite и облачной среды Asperitas (ИСП РАН));
- Высокая точность анализа текста благодаря ряду ключевых особенностей:
  - Многоуровневый поиск по смежным понятиям;
  - Адаптивность к сленгу, хэштегам и ошибкам;
  - Анализ эмоциональной окраски (с разделением отношения к объектам и их атрибутам);

  - Определение взаимосвязей людей и компаний (на основе информации в тексте);
  - Определение неявных упоминаний объектов в дискуссиях.
- Высокая скорость разработки индивидуального решения;
- Два варианта использования:
  - в качестве отчуждаемого продукта на локальном сервере заказчика с доступом как по протоколу HTTP (REST-архитектура), так и по протоколу RMI;
  - онлайн на сайте https://texterra.ispras.ru/;
- Простое и быстрое освоение специфичных предметных областей и возможность интеграции новых языков для анализа (благодаря современному подходу к машинному обучению).

ДЛЯ КОГО ПРЕДНАЗНАЧЕНА TEXTERRA?

- Разработчики корпоративного ПО (в частности, чат-ботов);
- Разработчики систем семантического поиска для специфических предметных областей (информационная безопасность, медицина, аудит и т.п.);
- Разработчики прикладных систем обработки текста.

ОПЫТ ВНЕДРЕНИЯ

Texterra доработана до промышленного уровня в рамках сотрудничества с HP и Samsung (цель совместных проектов — получение технологий для анализа корпоративной отчетности и поддержки работы смарт-телевидения).
В настоящее время Texterra используется в работе ряда технологий ИСП РАН (в частности, платформы для обработки данных Talisman). Texterra используется также рядом государственных ведомств России.

ПОДДЕРЖИВАЕМЫЕ ЯЗЫКИ

Texterra анализирует тексты на русском и английском языках.

СИСТЕМНЫЕ ТРЕБОВАНИЯ

- Любые платформы, поддерживаемые Java 8;
- Не менее 16 Гб оперативной памяти для каждого из анализируемых языков;
- Рекомендуется применение 64-битной версии ОС.

СХЕМА РАБОТЫ

ТРАЛ: СРЕДА АНАЛИЗА БИНАРНОГО КОДА

ТРАЛ — уникальный промышленный инструмент для анализа свойств бинарного кода. Позволяет работать с кодом различных целевых процессорных архитектур. Не требует наличия отладочной информации и исходных кодов. Применим для анализа всего программного стека от загрузчика до прикладного ПО. Включён в Единый реестр российского ПО (№5323).

ОСОБЕННОСТИ И ПРЕИМУЩЕСТВА

ТРАЛ — комплекс технологий, основанный на многолетнем опыте разработчиков компиляторов и специалистов по информационной безопасности. В отличие от аналогичных научно-исследовательских технологий в области анализа бинарного кода, доработан до промышленного использования.

Ключевые возможности:

- полносистемный анализ трасс выполнения (исследуется весь стек развернутого ПО, включая системное);
- восстановление потоков данных и управления на уровне машинных команд без ограничений;
- локализация в коде отдельных алгоритмов, формальное представление их структуры и семантики;
- выявление утечек чувствительных данных по памяти на основе динамического анализа помеченных данных по полносистемным трассам выполнения;
- автоматизация ручного анализа, полностью автоматическое решение многих прикладных задач.

ТРАЛ — это:

- Модульная архитектура среды (позволяет расширять набор поддерживаемых целевых платформ и развивать функциональное наполнение среды);
- Поддержка автоматизации анализа с помощью сценариев и открытого API (предоставляет возможность интегрировать среду с другими инструментами: IDA Pro и Wireshark).
- Глубокий анализ:
  - для анализа достаточно наличия лишь исполняемого бинарного кода;
  - в основе подхода — динамический анализ по трассам выполнения, при необходимости дополняемый статическим анализом снимков памяти;
  - предварительное автоматическое повышение уровня представления;
  - восстановление статического представления программ, входящих в состав анализируемой системы, в том числе по нескольким запускам;

- точный анализ потоков данных, учитывающий особенности аппаратуры (конвейер команд, прерывания, трансляция виртуальных адресов, DMA);
- интерактивное восстановление блок-схемы алгоритма, основанное на построении срезов информационных потоков;
- подход, реализованный в среде, невосприимчив к большинству известных приёмов противодействия анализу.

- Высокая производительность:
  - параллельный анализ с высокими показателями масштабируемости на многоядерных рабочих станциях;
  - возможность анализа длительных сценариев работы анализируемой системы.
- Развитый графический интерфейс:
  - просмотр трасс выполнения с обширными возможностями поиска и навигации, аналогичными классическому отладчику, но с возможностью мгновенного перемещения по потокам данных как вперёд, так и назад во времени;
  - автоматическая разметка высокоуровневой структуры трассы: процессов и потоков выполнения, обработчиков прерываний, стеков вызовов, динамически загружаемых модулей и символов в них;
  - просмотр значений параметров и возвращаемых значений вызванных функций;
  - разметка трассы с указанием внешних событий (сетевые взаимодействия и пользовательский ввод-вывод) и событий, связанных с работой аппаратуры.
- Техническая поддержка с возможностью доработки под особенности проводимых исследований конкретного ПО.

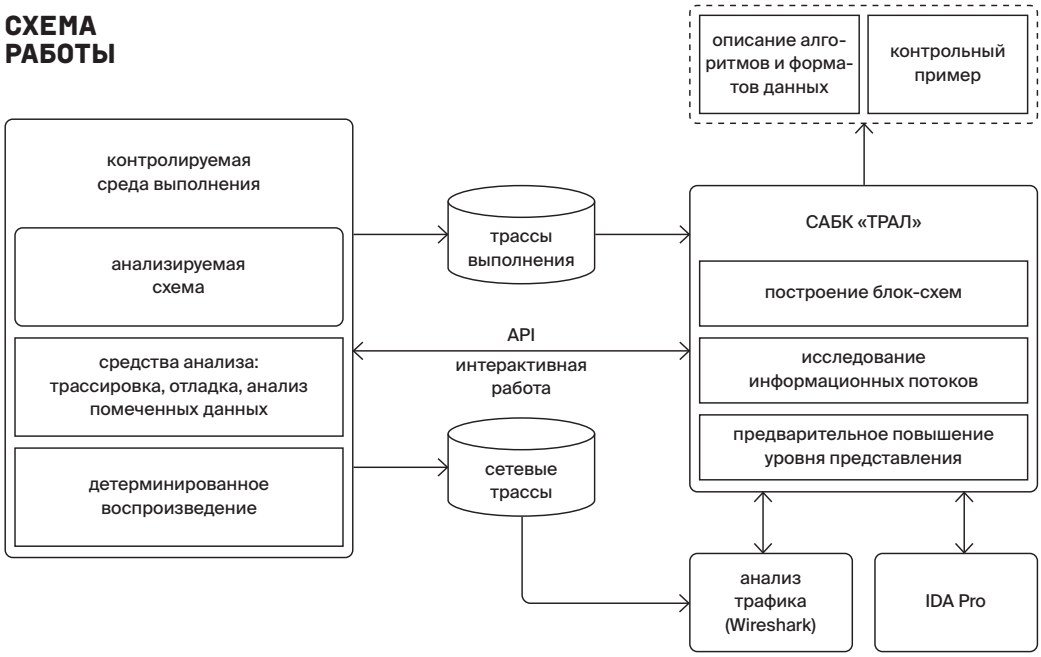
ДЛЯ КОГО ПРЕДНАЗНАЧЕН ТРАЛ?

- Лаборатории, проводящие анализ вредоносного кода;
- Компании-разработчики встраиваемого ПО и компонентов ОС;
- Компании-разработчики безопасного/доверенного ПО.
- Испытательные лаборатории.

ПОДДЕРЖИВАЕМЫЕ ПЛАТФОРМЫ И АРХИТЕКТУРЫ

- Системные требования среды анализа ТРАЛ: ОС Windows или ОС на базе ядра Linux, 64-разрядный процессор архитектуры x86, 16 и более Гбайт ОЗУ.
- Целевые процессорные архитектуры: x86, x86-64, ARMv6, ARMv7.
- Целевые ОС: семейство Windows, семейство Linux, поддерживается возможность работы с неопознанной ОС и с кодом, работающим вне ОС.

СХЕМА РАБОТЫ



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 (по контрактам с российскими и зарубежными компаниями, Минобрнауки РФ, программам РАН, грантам РФФИ и ФПИ и т.п.), нацеленные на создание новых технологий;
- 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 ИТ, 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 (ITRI (Тайвань), Университет Пассау (Германия), Израильский технологический институт Технион, Белградский университет и др.). 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 Института системного программирования РАН» (индексируется в РИНЦ и Scopus). Институт отвечает также за выпуск и редактуру журнала РАН «Программирование» (индексируется в Web of Science и Scopus). Оба издания входят в перечень ВАК.

Кроме того, при Институте функционирует российский Центр верификации ОС Linux, созданный для развития и продвижения открытых стандартов Linux.

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, которые сотрудничают с ИСП РАН более пяти лет. В числе главных зарубежных партнёров – Samsung, Huawei, HP, Intel, Nvidia, Bentley Systems (ранее Synchro Software), Linux Foundation; в числе отечественных — «РусБИТех», ГосНИИАС, «Вымпелком», «Базальт СПО», «МВП Свемел».

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.

В настоящее время в Институте функционируют совместные лаборатории с Samsung Electronics (нацелена на компиляторные технологии, в том числе на обеспечение безопасности в контексте ОС Android и Tizen) и Huawei (первая лаборатория проводит исследования и разработки в области компиляторных технологий и компонентов операционных систем, вторая — в области статического и динамического анализа). Кроме того, на базе облачной платформы Fanlight создана и успешно функционирует лаборатория для решения задач механики сплошных сред, реализующая исследовательские проекты в интересах промышленных предприятий. На базе платформы Lingvodoc работает лингвистическая лаборатория по документации исчезающих языков; исследования ведутся совместно с Институтом языкознания РАН. Томским государственным университетом и другими вузами и НИИ.

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 (например, с Free Software Foundation). 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 Фондом перспективных исследований (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 ИТ-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 ИТ-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 в нескольких направлениях:

- Интеграция ИСП РАН с ведущими вузами. Кафедры системного программирования, на которых работают сотрудники Института, открыты в МГУ им. М.В. Ломоносова, МФТИ и ВШЭ. В первый год обучения в ИСП РАН студенты-третьекурсники слушают лекции специалистов, 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 реальными специалистами по системному программированию.

- Стипендиальная программа. В рамках поддержки образовательных процессов ИСП РАН 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. В настоящее время функционируют и развиваются четыре лаборатории: в Ереване, в Великом Новгороде, а также в Орловском государственном университете и Академии ФСО. Лаборатории привлекают к работе успешных студентов и аспирантов, которые занимаются разработкой перспективных технологий в тесном сотрудничестве с индустрией.

В 2017 г. ИСП РАН совместно с компанией Samsung открыл на базе МФТИ «IoT Академию Samsung», в рамках которой студенты проходят спецкурс, направленный на изучение реальных случаев использования технологий Интернета вещей в различных отраслях, а также создают собственные прототипы IoT-устройств. В 2018 г. была запущена вторая часть этого проекта на Факультете аэромеханики и летательной техники (ФАЛТ) МФТИ в Жуковском.

ДЛЯ ЗАМЕТОК