CATALOGUE OF TECHNOLOGIES

IVANNIKOV INSTITUTE
FOR SYSTEM PROGRAMMING
OF THE RAS

ISP RAS

MOSCOW
2021

CATALOGUE
OF TECHNOLOGIES

2021

CONTENTS

2021. UNITED FOR A COMMON GOOD

ARUTYUN AVETISYAN

Academician
of the RAS,
ISP RAS Director.

2021 is a very successful year for our institute. We are advancing actively and coherently in all three of our activities: researching, developing technologies, and educating specialists in close cooperation with industry, academia, and regulatory authorities. This year we have grown a lot in quantity (our staff is more than 550 specialists, and we are working with more than 200 students) as well as in quality.

The long term experience of successful collaboration with Russian and foreign partners has allowed us to boost the mission of creating and moderating multidisciplinary communities. We have created and are actively developing three centers that are concentrating knowledge from different fields and are engaging all of institute departments, allowing to solve global practical problems in the area of medicine, cybersecurity, and trusted artificial intelligence.

Each center is being developed in collaboration with industrial, government, and academia partners:

— World-class Research Center (WCRC) “Digital biodesign and personalized healthcare,” jointly with Sechenov University, Institute of Biomedical Chemistry, Yaroslav-the-Wise Novgorod State University, Institute for Design-Technological Informatics of RAS;

— Technology center for security analysis of the Linux kernel, jointly with FSTEC of Russia and with active participation of leading IT companies such as Kaspersky Lab, RusBITech, Base Alt, Rosa, RED SOFT, Open Mobile Platform, Security Code, MCST, Elvis and others;

— Trusted AI center, jointly with Ministry of Economic Development, academia (MIPT, Skoltech, Medical Scientific Center and Faculty of Mechanics and Mathematics of Moscow State University, Innopolis University, Lobachevsky University, Psychology Institute of RAS, Joint Supercomputer Center of RAS) and industry (Kaspersky Lab, EC-leasing, InterProCom, Technoprom).

The institute multidisciplinary laboratories are advancing greatly as well. Two years ago we at ISP RAS have opened the linguistic laboratory, which is formed around the Lingvodoc platform that has gathered audio vocabularies and corpora for more than 900 endangered dialects of Ural and Altai languages already. This year we have signed the agreements on further development of the platform with the Government of Bashkortostan, with the Institute for Humanitarian Science of the Government of Republic of Mordovia and with Mordovia State University. We are creating new laboratories as well, e.g. we have opened the intelligent digital foresight and media data lab in collaboration with MGIMO University of the Ministry of Foreign Affairs.

We are also putting much energy into organizing various events. Together with our traditional conferences, in 2021 we have held a new one: “Data Science in Medicine”. Our partners for this event were D.O. Ott Research Institute of Obstetrics and Gynecology, Sechenov University, Moscow State University Clinic, N.I. Pirogov National Medical and Surgical Center, Yaroslav-the-Wise Novgorod State University, St. Luca Hospital, V.A. Nasonova Research Institute of Rheumatology, and Vorohobov City Clinical Hospital No. 67. Jointly we have created an annual forum where medical experts discuss digital healthcare tasks and communicate their requests to IT developers. This year we have also participated in the “Strategic leadership in the digital era and the AI technology” congress that was co-located with the ARMY-2021 forum.

A very special honor for us is the decision of the COMPSAC conference committee to hold its 2023 edition in Moscow and to name ISP RAS as its organizer. COMPSAC is a large international forum and one of the oldest flagship IEEE Computer Society events. COMPSAC is held annually during last 45 years, and making Russia its venue in the year of 75th anniversary of the Russian IT is a landmark event.

All these achievements allowed us significantly expanding our education activities. ISP RAS performs academic advising to modernize the bachelor program of software engineering on the Faculty of Computer Sciences in Higher School of Economics. 150 freshmen have started their studies using the new program. Similar changes happen in Moscow Institute for Physics and Technology: starting next year, students will be able to choose our new bachelor program named “System programming and applied mathematics.” In Moscow State University, the number of students on the chair for system programming has increased on 50%. We have started working with MGIMO University, MEPhI, and Lobachevsky University to create educational routes from a bachelor to a postgraduate. The Moscow Aviation Institute master program is making progress as well.

We are growing and successfully working on new large scale projects targeted on achieving technological independence and competitiveness of our country. Only when united with government agencies, industry and academia we are able to stand up to global challenges.

2021. WORLD-CLASS RESEARCH CENTER (WCRC) "DIGITAL BIODESIGN AND PERSONALIZED HEALTHCARE"

JOINTLY WITH SECHENOV UNIVERSITY, INSTITUTE OF BIOMEDICAL CHEMISTRY, YAROSLAV-THE-WISE NOVGOROD STATE UNIVERSITY, AND INSTITUTE FOR DESIGN-TECHNOLOGICAL INFORMATICS OF RAS

As a work package for creating the center, ISP RAS develops a cloud digital ecosystem based on existing ISP RAS technologies for big data processing (Asperitas, Talisman, Michman, Fanlight and others).

Most important results of 2021 include the following:

— A web laboratory for the ECG analysis is created. We have received more than one million ECGs for further analysis as a result of the partnership with Telemedicine Information Systems LLC, EC-leasing, and Technion (Israel Institute of Technology). We have trained a neural network model for classifying pathologies or cardiac registries that is successfully deployed in test mode in the "Integrated cardiologist" information system of the Republic of Tatarstan. We have also developed a cross-platform system for creating reference ECG sets via multiuser annotation (http://ecg.ispras.ru).

— A web laboratory for the histological image analysis is created. In a cooperation with our partners (RUDN University, City Clinical Hospital No. 31, V.I. Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology, Yaroslav-the-Wise Novgorod State University, and A.P. Avtsyn Research Institute of Human Morphology) we have gathered the EndoNuke dataset of histological images for cell nuclei detection (https://endonuke.ispras.ru/), which was labeled by doctors and interns via the CVAT open source tool for image annotation. Using the Fanlight technology, we have deployed within the ISP RAS cloud infrastructure the QuPath tool for labeling histological images and the system for blood

vessel segmentation developed jointly with D.O. Ott Research Institute of Obstetrics and Gynecology.

- We have developed a number of modules for the Michman orchestrator within the Asperitas cloud distribution: a module for creating virtualized HPC clusters on demand with an option of choosing an OS, and of including Slurm or OpenMPI, and a module for deploying Kubernetes clusters on demand with an option of choosing Kubernetes version and its specific settings.
- We have developed the new version of the Fanlight web laboratory platform that supports Kubernetes as a service scheduler.
- The Asperitas cloud distribution system components have been fully localized.

2021. TECHNOLOGY CENTER FOR SECURITY ANALYSIS OF THE LINUX KERNEL

JOINTLY WITH FSTEC OF RUSSIA AND WITH ACTIVE PARTICIPATION OF LEADING IT COMPANIES SUCH AS KASPERSKY LAB, RUSBITECH, BASE ALT, ROSA, RED SOFT, OPEN MOBILE PLATFORM, SECURITY CODE, MCST, ELVIS AND OTHERS

In 2021 ISP RAS has won the FSTEC of Russia tender for establishing a technology center for Linux kernel security analysis. The main goal of the center is to improve the security level of national Linux-based systems. The center operation plan includes a number of interconnected work packages such as the following:

- Applying best practices for secure software development (static source code analysis, fuzzing etc.);
- Developing patches for mitigating Linux kernel vulnerabilities;
- Adding Linux kernel vulnerabilities' descriptions to the vulnerability database of FSTEC of Russia.

Currently we are developing requirements and technical documentation for the center. In 2022, the center infrastructure is planned to be deployed in test mode, and in 2023 it will be fully operational.

The center has already attracted the community attention. There was an active discussion of the center work plans during the annual OS DAY conference, which is organized jointly by nine companies and institutions (ISP RAS, DZ Systems, Kaspersky Lab, Open Mobile Platform, RED SOFT, Cryptosoft, RusBITech, Base Alt, Zhukovsky Central Institute of Aerodynamics). Based on the discussion results, FSTEC of Russia invited the companies developing Linux-based certified products to join the center activities providing their best practices and experience in return for the possibility to use the center research results when certifying company products.

2021. TRUSTED AI CENTER

JOINTLY WITH MINISTRY OF ECONOMIC DEVELOPMENT, ACADEMIA (MIPT, SKOLTECH, MEDICAL SCIENTIFIC CENTER AND FACULTY OF MECHANICS AND MATHEMATICS OF MOSCOW STATE UNIVERSITY, INNOPOLIS UNIVERSITY, LOBACHEVSKY UNIVERSITY, PSYCHOLOGY INSTITUTE OF RAS, JOINT SUPERCOMPUTER CENTER OF RAS) AND INDUSTRY (KASPERSKY LAB, EC-LEASING, INTERPROCOM, TECHNOPROM).

ISP RAS has won the tender for supporting research centers for artificial intelligence (AI), and the victory was one of our main events for 2021. The tender was organized as a part of the “Aritifial Intelligence” federal project supervised by the Ministry of Economic Development of Russia. ISP RAS also took part in developing the Russian Code of Ethics for AI.

The main goal of the center is creating practical methods and platforms for developing and verifying the AI technologies with the specified level of trust. The center operation addresses the wide application areas in various business and social industries including image, audio, text and other data processing systems (e.g., face detection, intrusion detection, autonomous driving).

The key technological challenges of the center include:

- Analyzing, detecting and counteracting threats specific for AI technologies (such as adversarial attacks on models, attacks via adding backdoors and malicious code to models, model and data theft).
- Improving interpretability for machine learning models.
- Creating development tools and a cloud platform for AI-based trusted system development.

The key organizational challenges are as follows:

- Educating high quality specialists via developing new courses and involving students and junior engineers in all activities of the center, including the development of practical AI-based trusted systems.
- Creating and moderating both a distributed community of leading Russian scientists and a consumer ecosystem, basing on the principle of fair access to the center results.

Long-term ISP RAS partners are also interested in developing trusted AI technologies. One such example is Kaspersky Lab. There is a rich history of ISP RAS collaboration with industrial partners on the topic as well. For example, during the last few years jointly with Samsung Electronics we are actively developing a number of AI methods including AI techniques for software engineering and the methods for model interpretability in deep learning. We have created visual tools for explaining neural network decisions when analyzing images, audio, text, and table data. Also we are researching the applicability of modern AI techniques and neural network architectures for detecting technical debt. Finally, we are working on computer vision methods for testing mobile applications (developing tools for finding and localizing visual defects and detecting UI elements).

2021. ADVANCING EDUCATIONAL PROGRAMS

— ISP RAS performs academic advising to modernize the software engineering bachelor program of the Faculty of Computer Sciences in Higher School of Economics. Starting from September 1st, 150 freshmen are studying on this new program that includes classical courses on algorithms and data structures and on computer architecture. Further courses on operating systems, compiler technologies, and software engineering will lay the foundation in the area of program analysis, and the courses on databases, natural language processing, and advanced machine learning methods will serve as a base for the data analysis expertise.

— ISP RAS has extended its cooperation with Moscow Institute for Physics and Technology. Our chair for system programming now can accept students from Phystech School of Applied Mathematics and Informatics as well as from Phystech School of Radio Engineering and Computer Technology. Also we are starting the new “System programming and applied mathematics” bachelor program that students of Department of Innovation and High Technology can choose starting next year. In 2021, ISP RAS and Huawei in collaboration with MIPT have launched a two-year free educational program devoted to system programming that can be taken by students of any MIPT departments and years of education. The program includes the courses on algorithms and data structures, modern C++, compiler construction, and developing an operating system kernel.

— In Moscow State University, the number of students on our chair for system programming has increased on 50%. The current applicants are more advanced and talented so they successfully pass the entry testing and are now studying under the supervision of the chair staff. We are also constantly updating our chair courses.

— We have started active work with MGIMO University of the Ministry of Foreign Affairs. ISP RAS and MGIMO signed an agreement with a roadmap including launching a new master program and education courses as well as creating a digital platform for intellectual big data analysis in the area of foreign affairs. Students and junior researchers will use the platform to get involved in the projects working on the Ministry practical tasks. ISP RAS has opened the intelligent digital foresight and media data lab.

TECHNOLOGIES

ASPERITAS AND CLOUD SOLUTIONS FAMILY

Asperitas is a platform for data storage and performing complex resource-intensive calculations on demand. It includes a cloud environment also called Asperitas (listed as No. 5921 in the Unified Register of Russian Programs) as well as Michman, a PaaS orchestrator, and Clouni, an IaaS orchestrator. Fanlight, a web laboratories platform, is also a part of ISP RAS cloud solutions family (listed in the Register as No. 6066).

ASPERITAS CLOUD ENVIRONMENT

Asperitas is based on Openstack and Ceph and was created in a joint project with Dell. It is designed for computations using large amounts of available resources. Asperitas is based on modern open source technologies that are ubiquitous for building large private clouds. The distribution delivery uses an external HDD drive containing a TUI installer for a deployment node and all the necessary tools for launching the deployment process.

ASPERITAS PROVIDES

- An onsite installation option (the provided infrastructure can be installed and fully controlled in an isolated environment due to the usage of open standards and software as well as ISP RAS research).
- High security (the environment is built on top of a smaller code base and uses know-how solutions that increase security).
- Virtual networks and computational clusters management using Keystone, Neutron and Nova (similar to Amazon EC2).
- Block storage and scalable object storage is based on the Ceph distributed file system.
- Adaptation to specific problem classes (e.g. continuum mechanics, big data analysis, program analysis for defect detection etc.).

MICHMAN, A UNIVERSAL ORCHESTRATOR

(https://github.com/ispras/michman, https://michman.ispras.ru). Michman is a PaaS services orchestration tool for a cloud environment performing big data analysis, machine learning, running distributed programs on clusters and storing large amounts of data. It supports automatic cluster deployment with fully set up machines. Users can create clusters in isolated projects and monitor the status of deployed services and clusters in real time.

Michman is able to deploy virtual clusters with PaaS services on demand, including:

- A big data analysis cluster with arbitrary number of nodes having Apache Spark, Apache Hadoop, Apache Ignite and Jupyter Notebook fully set up and ready to work.
- A database for storing large data such as PostgreSQL, Apache Cassandra, CouchDB, ClickHouse or Redis. Some of the databases can be deployed in a distributed mode.
- NextCloud storage and file exchange system.
- Slurm, a cluster management and job scheduling system.
- Kubernetes, a container orchestration system.

MICHMAN PROVIDES

- A service for users and isolated groups management with REST API.
- Storing the data regarding deployed clusters and services, their status, and access points.
- Storing cluster templates ready for deployment.
- Deploying complex distributed systems on demand with arbitrary service combinations.
- Managing service dependencies including per-version dependencies.
- Local deployment without Internet access.
- Storing detailed data regarding available services, their versions and parameters.
- Easy service addition using REST API.
- Integration with IaaS cloud virtualization systems.

Alongside with Michman Asperitas distribution includes Clouni (https://github.com/ispras/clouni), a tool for translating TOSCA Simple Profile YAML patterns to IaaS deployment scenarios based on Ansible.

FANLIGHT

(https://fanlight.ispras.ru). Fanlight is a web laboratories platform that resulted out of ISP RAS participation in the University Cluster program and in the Open Cirrus international project (founded by Hewlett-Packard, Intel and Yahoo!). It is designed for deploying a SaaS infrastructure for a computational web laboratory. Fanlight is built on container technologies and provides virtual workspaces in the Desktop as a Service model (DaaS). It is available in two versions. Initially the platform was based on Docker Compose, and later an implementation based on Kubernetes was added. Fanlight is available on https://fanlight.ispras.ru and supports applications developed for a Linux kernel based OS.

FANLIGHT PROVIDES

- High performance cloud computing through the use of containers:
  - Working comfortably with heavy CAD-CAE engineering applications that require 3D graphics hardware acceleration support for complex visualization.
  - Support for running MPI, OpenMP, CUDA applications via HPC clusters, multi-core processors and NVIDIA graphics accelerators.

- PaaS-provided computing capability expansion via adding hardware resources (HPC/BigData clusters, storage systems, servers with GPUs).
- Application area specific customization via integrating specialized computing application packages. The following systems were successfully deployed:
  - in a CFD area: OpenFOAM, SALOME, Paraview etc.;
  - in a Gas&Oil industry: tNavigator, Eclipse, Roxar, Tempest etc.
- Allows any thin client (including mobile devices) without additional client software.
- Can be deployed on a server, a computing farm, in a cloud (IaaS), in a Kubernetes cluster, or in a local data processing center. The Kubernetes based version additionally allows using various CRI engines for container execution.

CLOUD SOLUTIONS DEPLOYMENT STORIES

The computing cluster based on Asperitas analyzes information flows in the Talisman social media analysis framework and supports other ISP RAS technologies (e.g. analyzing Android OS using Svace). The following projects were also implemented: a joint project with Huawei (large graphs analysis using big data processing) and the Tizen OS lifecycle support infrastructure that allows organizing joint development of OS components and automating regular build and testing of OS images. In addition, a number of projects is performed jointly with the Ministry of Science of Russian Federation.

The Fanlight platform was used in a number of joint projects for web laboratory deployment, including Russian Federal Nuclear Center of the All-Russian Scientific Research Institute of Experimental Physics, OOO RRS-Baltika, Keldysh Institute of Applied Mathematics (developing a technology for increasing and using efficiently the hydrocarbon raw materials resource potential of the Union State), ISP RAS Laboratory of Continuum Mechanics (https://unicfd.ru).

ASTRAVER: A VERIFICATION TOOLSET

AstraVer Toolset is a deductive verification system for key software components. It allows developing and verifying security policy models as well as proving the correctness of software modules written in the C programming language. AstraVer is essential for ensuring the required trust levels from ADV_SPM and ADV_FSP assurance families as defined in the ISO/IEC 15408 standard.

FEATURES AND ADVANTAGES

AstraVer Toolset is a set of tools designed for industrial use. It is based on many years of scientific research and combines two verification approaches: at the model level and at the code level. Parts of the AstraVer Toolset are similar to Microsoft VCC and Frama-C WP, but unlike those AstraVer is specifically designed to support the key security components' verification in the Linux kernel. AstraVer Toolset is free and open source, available at http://linuxtesting.org/astraver.

ASTRAVER PROVIDES:

- An integrated approach to verification, supporting the formalization of high-level requirements and analyzing the C source code behavior.
- Modeling and formalizing functional requirements, proving internal consistency and unreachability of insecure states.
- Testing whether functional requirements are satisfied in an implementation, using their formal models to check the correctness of the observable behavior and to evaluate the quality of testing and generated test cases.
- Verification of critical components written in C (requirements' formalization, correctness proof on all possible input values).
- Support for real industrial C code (GCC compiler extensions, arithmetic operations with bitwise precision, address arithmetic including the container_of intrinsic, function pointers, casting).
- Adhering to the protection profile requirements (ISO/IEC 15408), such as
  - formal security policy modeling;
  - formal verification of internal consistency of a security policy model;
  - formal proof that the target system cannot reach an insecure state;
  - a formal or a semi-formal functional specification development;
  - a formal/semi-formal proof of correspondence between the security policy model and the functional specification;
  - a formal/semi-formal proof of correspondence between different representations of target software, like functional specification, design and source code.

— Ability to adjust the toolset for a specific customer to perform the C source code components verification.

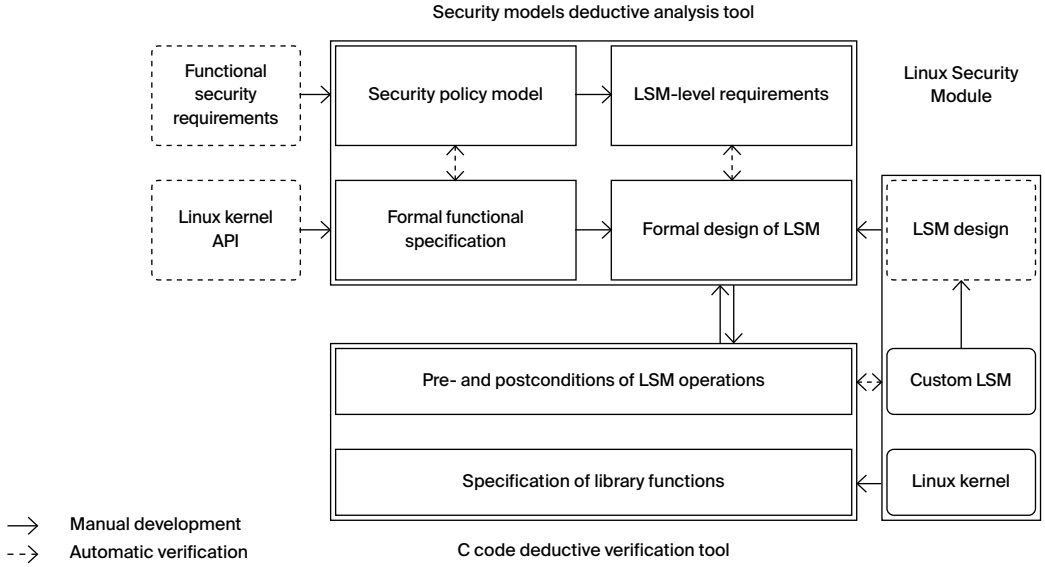
WHO IS ASTRAVER TARGET AUDIENCE?

— Companies developing critical systems, including software in aviation, railway, medical and nuclear power industries.
— Companies that need to certify their software as guided by the ISO/IEC 15408 standard.
— Certification laboratories for information protection software.

ATSTRAVER DEPLOYMENT STORIES

AstraVer Toolset was used in the development of access control mechanisms for Astra Linux Special Edition (RPA RusBITech JSC). As a result, this Astra Linux edition has passed the certification for compliance with the FSTEC information security requirements, which are defined for operating systems of the 2A protection profile. Both the security policy model and the access control mechanisms source code were successfully verified using AstraVer Toolset. The verification work for the new security model features is constantly ongoing.

ASTRAVER WORKFLOW



BINSIDE: A BINARY CODE STATIC ANALYSIS TOOL

BinSide is a static program analysis tool for finding defects in binary code. It is useful when checking programs without source code, such as closed source 3rd party libraries, as well as assisting with required static information to dynamic analysis tools.

FEATURES AND ADVANTAGES

BinSide is a binary code analysis platform based on the BinNavi framework. It translates assembler code into a REIL representation. REIL allows analyzing binary code in a target processor and OS independent way. BinSide provides various analysis types such as defect detection, dynamic analysis optimization, reverse engineering support (via integration with the IDA Pro and Ghidra disassemblers).

BINSIDE CORE PROVIDES:

- Easy extension:
  - individual error detectors are written as plugins;
  - the REIL representation of 17 instructions without side effects is used (each assembly instruction is translated into a set of REIL instructions);
  - it is possible to specify functions' semantics to improve analysis quality.
- Supports analyzing executables and libraries for x86-64, ARM and MIPS architectures.
- Detecting the following CWE types: 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), CWE-77 (Command Injection).
- An analysis core with the following features:
  - Intraprocedural control and dataflow analysis, including value and pointer analysis, taint analysis, tracking dynamic memory, computed jump recovery. The intraprocedural analysis also calculates the function call effect on the calling context and applies call effects for callees.
  - Interprocedural analysis and finding interprocedural defects. Each defect has a trace connecting input data and a program point where the defect manifests itself. At this point the analyzer detects a memory model or a security model violation.
  - Finding errors on all execution paths (including those not covered by testing or dynamic analysis).

— Displaying defect traces on a source code in the Svace web interface in the presence of debug info in an analyzed executable.

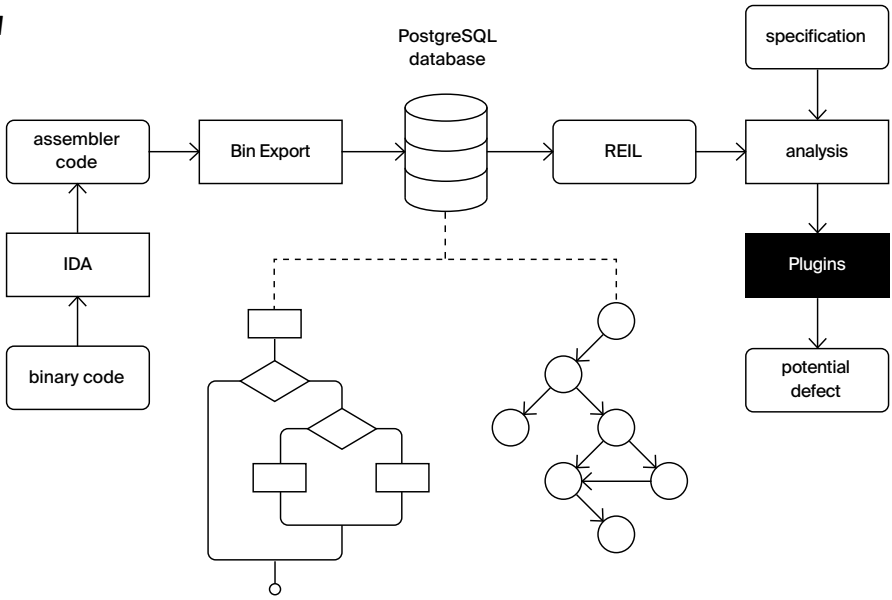
BINSIDE PLUGINS INCLUDE:

— A function call effect recovery plugin.
— A plugin for binary code clone search that serves as a base for the following features:
  — a plugin for detecting changes between different software versions;
  — a plugin for function names markup transfer from one binary file to another.

WHO IS BINSIDE TARGET AUDIENCE?

— Companies that need to check thoroughly the used 3rd party software with no access to its source code.
— Developers who need to increase dynamic analysis quality with the data collected by a static analysis.
— Reverse engineering experts.
— Companies performing software audition or certification.

BINSIDE WORKFLOW



CASR: CRASH ANALYSIS AND SEVERITY REPORTING TOOL

Casr creates automatic reports for crashes happened during program testing or deployment. The tool works by analyzing Linux coredump files. The resulting reports contain the crash severity and additional data that is helpful for pinpointing the error cause.

FEATURES AND ADVANTAGES

Casr solves the same problem as Apport, an open source system, but in contrast with it, Casr estimates a crash severity and provides a list of open files and network connections at a crash time.

CASR PROVIDES:

- Detecting critical program faults that can lead to hijacking control flow.
- Classifying crashes into one of 23 classes based on a program state at a crash time (function return address corruption, null pointer dereference etc.). Fatal errors are further grouped based on severity, such as exploitable, potentially exploitable, or denial of service errors.
- Collecting a list of open files and network connections that could be the reason of a crash.
- An extended crash report containing the fatal error severity and other data (OS and package versions, executed command line, call stack, open files and network connections, register state etc.).
- Reports for hard to reproduce crashes such as non-deterministic errors, cases when the original execution environment is hard or impossible to reconstruct etc.
- Integration with monitoring tools (such as Zabbix) that allows system administrators and devops to receive on the fly notifications about critical program crashes.

WHO IS CASR TARGET AUDIENCE?

- Companies that need to receive the data regarding user-deployed programs' crashes highly reliable and secure software.
- Companies that need to certify the developed software.
- Certification laboratories.

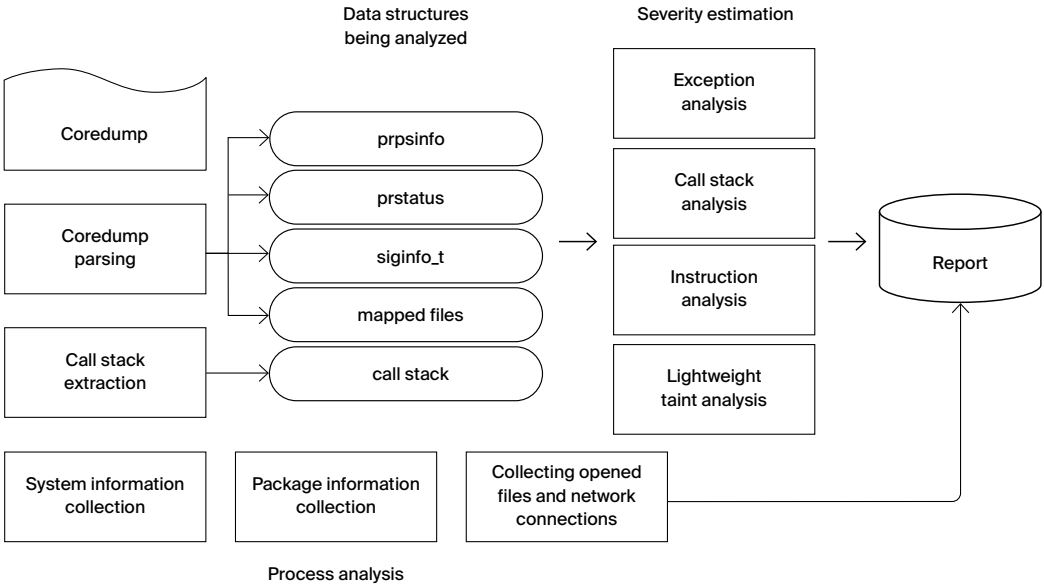
CASR DEPLOYMENT STORIES

Casr is deployed in a number of Russian companies and vendors as an add-on tool to ISP Crusher.

SYSTEM REQUIREMENTS

Linux-based OS for x86, x86-64, or ARMv7. Casr can be deployed as a deb package.

CASR WORKFLOW



CONSTRUCTIVITY 4D: A TECHNOLOGY OF INDEXING, SEARCHING AND ANALYSIS OF LARGE SPATIAL-TEMPORAL DATA

Constructivity 4D is a technology for creating innovative software services that are capable of processing highly dynamic scenes and vast arrays of spatial and temporal data. It performs visual analysis of millions of objects with individual geometry and dynamic behavior. Constructivity is deployed within the Synchro system (Bentley Systems) that is used for 4D modeling of extremely large construction sites.

FEATURES AND ADVANTAGES

Constructivity 4D is a production level technology that puts together original methods of spatio-temporal indexing, search and qualitative and quantitative data analysis. Developed methods account for the specifics of objects' geometric representation, complex organization and the apriori known nature of their dynamic changes.

CONSTRUCTIVITY 4D PROVIDES:

— Support for a well-developed set of operations:
  — Temporal operations implement classical interval algebra introduced by Allen with respect to time stamps of discrete events and their intervals.
  — Metric operations allow determining the individual properties of geometric objects and the characteristics of their mutual arrangement. Diameter, area, volume, center of mass, planar projections, and distances between objects can be calculated for solid geometric objects.
  — Topological operations are intended to classify the relative location of objects and to establish the facts of their coincidence, intersection, coverage, touch, overlap or collision. In contrast with known topological models such as DE-9IM, RCC-8, RCC-3D, these operations allow constructive implementation and are applicable for the analysis of complex objects.
  — Orientational operations generalize known Frank's and Freksa's relative orientation calculi, cardinal direction calculi (CDC), oriented point relation algebra (OPRA) and are

applicable for the analysis of objects with extended boundaries.

- Efficient query execution and typical problems solving, in particular, queries for reconstructing a scene at a given point in time, retrieving objects in a given spatial region, finding nearest neighbors, determining static and dynamic collisions, and conflict-free routing in a global dynamic environment are effectively resolved.
- A spatial-temporal indexing system including binary event trees, spatial decomposition trees, bounding volume trees, object cluster trees, space occupation trees.
- A hybrid computational strategy for determining collisions in scenes that combines methods for precise collision determination, collision localization methods using spatial decomposition, methods of hierarchies of bounding volumes, temporal coherence methods.
- An object-oriented library implemented in C++ that includes extensible set of classes, interfaces and related methods for specifying spatial-temporal data and executing typical queries.
- An original method for navigation in global dynamic environment that is based on extracting spatial, metric and topological information from geometric representation of 3D scenes and its concerted usage on path planning.
- Various options for extending the library so that it can be used both in new software applications development and in legacy applications.

WHO IS CONSTRUCTIVITY 4D TARGET AUDIENCE?

The technology is used for creating application systems in vastly different fields, including but not limited to: computer graphics and animation, geoinformatics, scientific visualization, design and manufacturing automation, robotics, logistics, project management and scheduling.

CONSTRUCTIVITY 4D DEPLOYMENT STORIES

The technology has been successfully deployed within the Synchro software system (https://www.bentley.com/en/products/brands/synchro) that is designed for visual 4D-modeling, planning and management of large-scale industrial projects in the construction and infrastructure areas, as well as others. Synchro is used in more than 300 companies in 36 countries.

DEDOC: A DOCUMENT STRUCTURE RETRIEVAL SYSTEM

Dedoc is an open universal system for converting documents to a unified format. It extracts a document's logical structure, its tables and metadata. The document's contents is represented as a tree storing headings and lists of any level. Dedoc can be integrated in a document contents and structure analysis system as a separate module.

FEATURES AND ADVANTAGES

Dedoc is implemented in Python and works with semi-structured data formats (DOC/DOCX, ODT, XLS/XLSX, CSV, TXT, JSON). Dedoc can be extended via plugins and contains the Docreader plugin package for working with images (PNG, JPG etc.), archives (ZIP, RAR etc.), PDF and HTML formats. Document structure extraction is fully automatic regardless of input data type. Metadata and text formatting is also extracted automatically.

DEDOC PROVIDES:

- Extensibility due to a flexible addition of new document formats and to an easy change of an output data format.
- Support for extracting document structure out of nested documents having different formats.
- Extracting various text formatting features (indentation, font type, size, style etc.).
- Adding rules for correcting mistyped document lists.
- Extracting table data from an XML-style DOC/DOCX format.

DOCREADER PROVIDES:

- Working with scanned document images of various origin (statements of work, legal documents, technical reports, scientific papers) allowing flexible tuning for new domains.
- Working with PDF documents either with or without a text layer.
- Recognizing a physical structure and a cell text for complex multipage tables having explicit borders with the help of contour analysis; detecting table orientation.
- Using Tesseract, an actively developed OCR engine from Google, together with image preprocessing methods.
- Utilizing modern machine learning approaches for detecting a document orientation and extracting its hierarchical structure based on the classification of features extracted from document images.

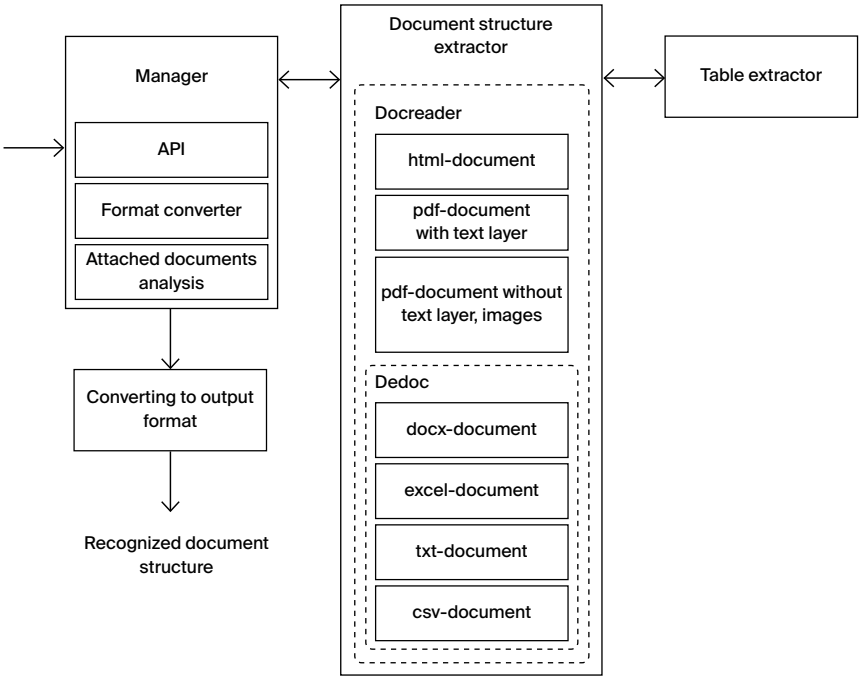
WHO IS DEDOC TARGET AUDIENCE?

- Developers of document contents analysis and management systems.
- Developers of intellectual text analysis algorithms.
- Developers of automatic document processing systems.

SUPPORTED LANGUAGES

Russian and English.

DEDOC WORKFLOW



DIGITEF: A DIGITAL TWIN PLATFORM

DigiTEF is a software platform based on OpenFOAM and other open source tools, as well as unique modules and libraries developed at ISP RAS. DigiTEF solves various application problems of gas dynamics, aerodynamics, hydrodynamics, and acoustics. It is tailored for creating and working with highly sophisticated digital models of industrial devices. DigiTEF is included in the Unified Register of Russian Programs (No. 5377).

FEATURES AND ADVANTAGES

The platform delivers the same level of user experience as its competitors worldwide. DigiTEF core performance and accuracy evaluations compared with ANSYS Fluent and Star CCM+ showed similar (and in some cases lower) computational costs with the same accuracy.

The community of engineers, researchers, and industrial project developers has been formed around the DigiTEF platform.

DIGITEF MEANS:

- open source code (allows controlling and adapting implemented algorithms);
- the development pace of OpenFOAM+;
- automation tools for computation and model integration that allow integrated research of technical objects;
- possibility of developing additional components according to the specific requirements.

DIGITEF CONSISTS OF TWO MAIN BLOCKS:

1. OpenDTE, the platform core based on OpenFOAM. It contains the basic algorithms, procedures, and functions, as well as a set of third-party libraries in C++. It is fully open and can be obtained at https://github.com/unicfdlab. OpenDTEF consists of the following components:

- tools for modeling compressible flows;
- settings setup for advanced cases based on swak4Foam;
- parameterization based on Python. This allows automating calculation cases as well as integrating Salome, ParaView, and CodeAster software systems into DigiTEF.

2. Modules developed at ISP RAS:

- Data analysis for visualizing and retrieving information. It is designed to analyze results and build models of reduced dimension using data processing methods (FFT, POD, DMD, Hilbert transformations).
- Compressible flows simulation based on quasi-gas dynamics (QGD) equations, allowing to use the spatio-temporal averag-

ing procedure to determine the main gas-dynamic quantities (density, velocity, temperature, and others).
- Incompressible flows simulation based on QHD equations. The module is applicable in oceanology, convection, and subsonic flows problems.
- Incompressible and compressible flows simulation based on the Pimple and Kurganov-Tadmore hybrid algorithm.
- Subsonic turbulent flows simulation using the hybrid URANS / LES approach and low dissipative numerical schemes.
- Acoustic analysis. The module implements the Curle and Focs Williams-Hawkings analogies.
- Ice dynamics simulation.

WHO ARE DIGITEF USERS?

DigiTEF is designed for use in the facility of resource-intensive industries. Using digital twin models allows increasing engineering efficiency as well as reducing costs and complexity of the industrial projects implementation.

DEPLOYMENT STORIES

DigiTEF is used in several projects in the fields of wind energy, aerospace, aviation, metallurgy, as well as in the oil and gas industry. DigiTEF open source modules are successfully used in Institut Pprime (France), Korea Atomic Energy Research Institute (Korea), Universität der Bundeswehr München (Germany), Northwestern Polytechnical University (China), Embry-Riddle University (USA), California Institute of Technology (USA), etc.

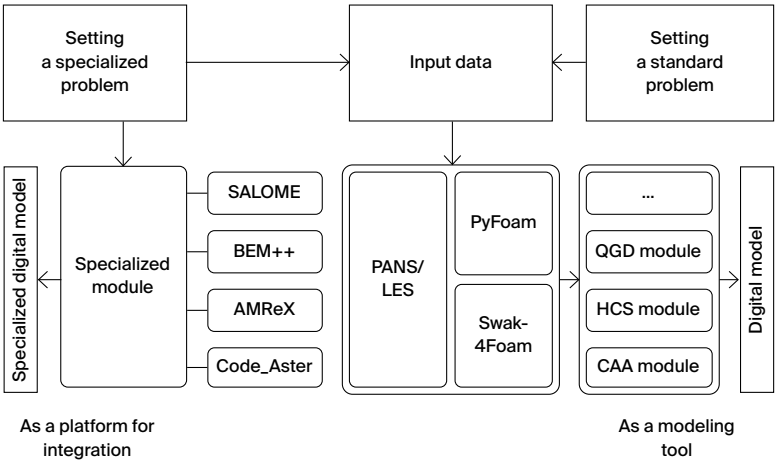
SYSTEM REQUIREMENTS

Linux OS. Other operating systems that support the Oracle VirtualBox virtual machine may also be used (on Microsoft Windows 10 via the Bash shell). Moreover, the performance loss due to virtualization does not exceed 5%.

Required RAM: 16 Gb or higher.

DigiTEF supports parallel computing, which significantly speeds up its work. Also it supports high performance computing systems (supercomputers and clusters) to accelerate calculations. Using up to 1536 computational cores was tested.

WORKFLOW



DOCMARKING: A SYSTEM FOR MARKING TEXT DOCUMENTS

Docmarking is a unique system for embedding digital watermarks into text documents. It allows creating a digital or physical document copy that is almost indistinguishable from the original yet exactly identifies the user or the device that was the intended recipient.

FEATURES AND ADVANTAGES

Docmarking is based on research results in the areas of steganography, digital image processing, and machine learning. The marking system builds on the methods for text detection in images and image-based text classification (the digital watermark is embedded in the text areas that were found in the image).

Docmarking has a number of advantages compared to competing technologies. Watermark extraction does not require the original document. The system supports embedding a watermark in the same scanned document multiple times, and the previous watermark is erased when the new one is being embedded.

DOCMARKING PROVIDES

- Marking algorithms based on machine learning.
- Support for documents of all formats.
- Embedding a watermark when a document is displayed on a screen or printed.
- Standalone setup and work on the client side.
- Real time registering of the embedded watermarks on the server.
- Centralized 24/7 monitoring of the connected client devices.

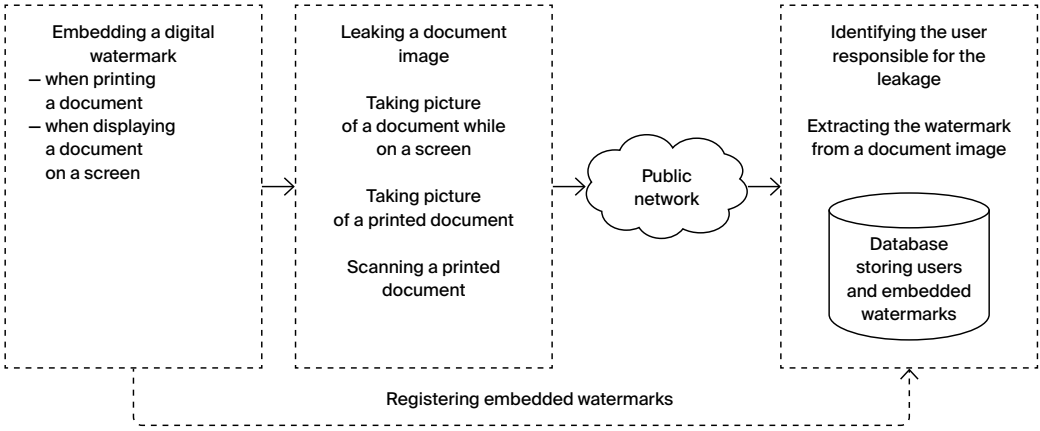
WHO IS DOCMARKING TARGET AUDIENCE?

- Government entities and public offices.
- Companies that would like to enforce their guides for handling classified documents.

SUPPORTED OPERATING SYSTEMS

Windows (32-bit, 64-bit), Linux (64-bit), including Astra Linux 1.5/1.6 SE.

DOCMARKING WORKFLOW



ISP CRUSHER: A DYNAMIC ANALYSIS TOOLSET

ISP Crusher is a toolset that combines various dynamic analysis approaches. It includes ISP Fuzzer, a fuzzing tool, and Sydr, an automatic test generation tool for complex programs. In the near future Crusher will also include the BinSide analyzer, another ISP RAS technology. Crusher allows organizing a development process that is fully compliant with GOST R 56939-2016 and other regulatory requirements of FSTEC of Russia.

ISP FUZZER, A TESTING TOOL

ISP Fuzzer is a fuzzing tool that is essential for any software development phase, be it coding, testing, or deployment. The fuzzer finds program errors either with or without access to the program source code. It solves the same problems as its global competitors (Synopsys Codenomicon, beSTORM, Peach Fuzzer), but it is more convenient for Russian companies in the import substitution context.

ISP FUZZER PROVIDES:

- Fuzzing various input data sources (files, command line arguments, standard input stream, environment variable arguments, network sockets, raw memory writes).
- Adding custom mutational transformations (for new input data generation and increased fuzzing efficiency).
- Input data pre- and post-processing plugins for performing data transformations before feeding the data to the software being analyzed.
- Multicore and distributed parallel analysis support.
- Custom network plugins support allowing to interact with a client or a server and to send mutated data.
- Integration with other ISP RAS security development lifecycle tools, such as:
  - using the Sydr dynamic symbolic execution analyzer for improving fuzzing efficiency;
  - automated input data generation for triggering errors found by the BinSide static analysis tool;
  - using the Svace analyzer GUI to show a function call trace that resulted in a crash;
  - using the ANTLR parser generator to create an input data corpora.
- Integration with the IDA PRO disassembler (exporting coverage data to the Lighthouse plugin for displaying covered basic blocks and showing the percentage of the coverage achieved).
- Analysis of client/server software working via stateless or stateful protocols.
- Support for describing scenarios for GUI software fuzzing.

- Broad support for embedded software fuzzing with partial emulation and symbolic execution.
- Adding user specified handlers to be called automatically for the newly generated input data.
- Running dynamic analyzers on generated input (Valgrind, DrMemory, QASan).
- Easy extension via adding new analysis algorithms to the existing infrastructure as well as fast adaptation to new fuzzing problems.
- Distributing input data between the fuzzer processes for more effective work.
- Estimating severity for the crashes found.
- Using the Radamsa fuzzer for new input data generation.

Automated fuzzing wrappers generation.
Supporting various instrumentation tools such as DynamoRIO, QemuUserMode, GCC or LLVM static instrumentation, QemuSystemArm.

SYDR, A DYNAMIC SYMBOLIC EXECUTION TOOL

Sydr is an automatic test generation tool for complex programs that finds errors and increases code coverage during testing. Sydr constructs the program's mathematical model that allows a fuzzer to explore new execution paths that are hard to follow via genetic mutation approaches. The tool advances further the dynamic symbolic execution approach used earlier in Avalanche and Anxiety analyzers developed in ISP RAS. In contrast with similar open source tools, Sydr ensures the correctness of generated input data by checking whether the newly found execution path has the target conditional branches inverted.

SYDR PROVIDES:

- Given an execution path, inverting all conditional branches that depend on input data. Inverting in parallel is also supported.
- ISP Fuzzer integration to reach the code behind the branches depending on comparisons with constants.
- Automating reverse engineering such as assisting an expert in reaching a given program point or collecting a trace of instructions that depend on input data.
- Supporting various input data sources (files, network sockets, environment variables, standard input stream, command line arguments).
- Safety predicates that are used to find errors and to generate input data for the errors found (for division by zero, null pointer dereference, buffer overflow).
- Symbolic execution of multithreaded programs.
- Inverting indirect branches (in switch statements). Table jumps and computed jumps are detected.
- Formula slicing. Sydr removes excessive formulae (not influencing the conditional branch being inverted) from the path predicate. This feature solves the problem of undertainting and speeds up SMT solver work for generated queries.

WHO IS ISP CRUSHER TARGET AUDIENCE?

- Companies developing highly reliable and secure software.
- Companies auditing or certifying software.

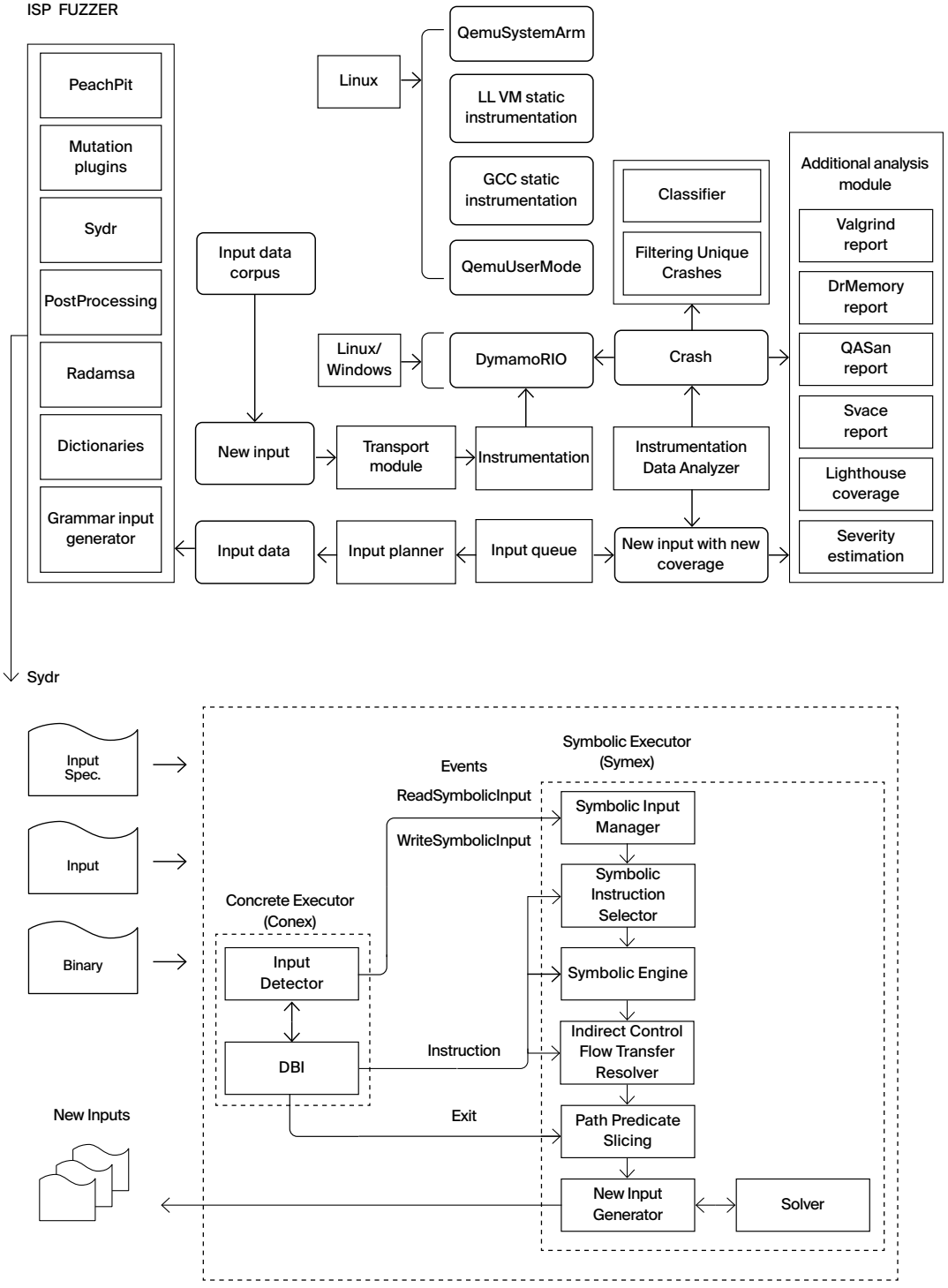
SUPPORTED SYSTEMS AND PLATFORMS

Linux and Windows OS family support. Crusher can also fuzz embedded devices (controllers, IoT devices) as well as Windows services and COM objects.

ISP CRUSHER DEPLOYMENT STORIES

ISP Crusher is used in more than 40 companies and certification labs, including RusBITech, Postgres Professional, Security Code, Swemel and others.

ISP CRUSHER WORKFLOW



ISP OBFUSCATOR

Obfuscator is a set of technologies to prevent mass exploitation of vulnerabilities resulting from errors or backdoors. In case a hacker has attacked one of the devices that has a certain software installed, the rest will remain protected by changes to the code that the tool made.

FEATURES AND ADVANTAGES

Obfuscator protects a software system from mass exploitation of vulnerabilities using various code diversification methods and allows compiling a full OS distribution.

ISP OBFUSCATOR PROVIDES:

- Fine-tuning the balance of obfuscation level and performance (when protecting against reverse engineering). The minimum speed degradation is 1.2 times, the maximum is 8 times.
- Full automation (no need to change the program source code or to spend efforts for build system integration).
- Based on the GCC compiler, which allows correctly building the full OS source code.
- The original control flow integrity technique (CFI), which successfully counteracts most of code reuse attacks (ROP, JOP, ret-to-libc, etc.). The implemented CFI support within the GCC compiler shows the average slowdown of about 2% on the SPEC CPU2006 test suite, which is noticeably lower than that of the traditional methods.
- Two diversification approaches:
  - Dynamic code diversification at program startup. It is used when the customer needs the same binary code deployed on all devices (for example, because of the certification procedure). This method allows shuffling up to 98% of code with a slight increase in size and a performance degradation of about 1.5%. The obfuscator provides the following advantages over the similar products:
    - shuffling with function granularity (as opposed to ASLR and Pagerando technologies that randomize only large blocks of code);
    - shuffling functions throughout the full OS code except the kernel, and avoiding conflicts with the antivirus software (compared to the Selfrando technology developed for the Tor Browser).
- Static code diversification. During each separate compilation, depending on the specified key, the unique executable file is created. This approach has the following advantages:
  - the binary code size does not increase (which is especially important for the Internet of things use case);
  - performance degradation is close to zero;
  - an extended set of diversifying transformations can be applied and more flexibly customized, as the required operations are performed within the compiler during build time, as opposed to working at link time;
  - performs also Control Flow Integrity (CFI).
- Conflict-free combination with other software protection tools (including the ASLR system mechanism).

WHO IS OBFUSCATOR TARGET AUDIENCE?

— Developers of specialized operating systems;
— Application software developers.

OBFUSCATOR DEPLOYMENT STORIES

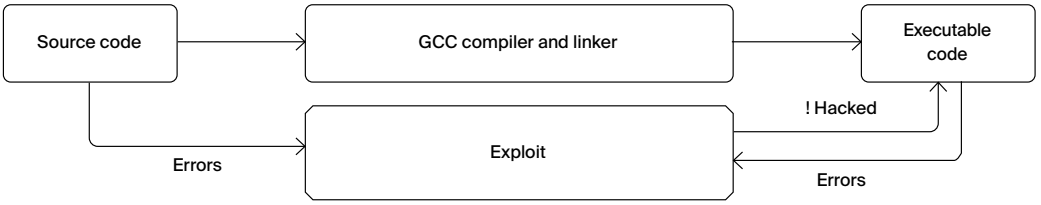
ISP Obfuscator is deployed in the Zirkon OS, which is used by the Ministry of Foreign Affairs and the Border Guard Service of the Federal Security Service of Russia.

SYSTEM REQUIREMENTS

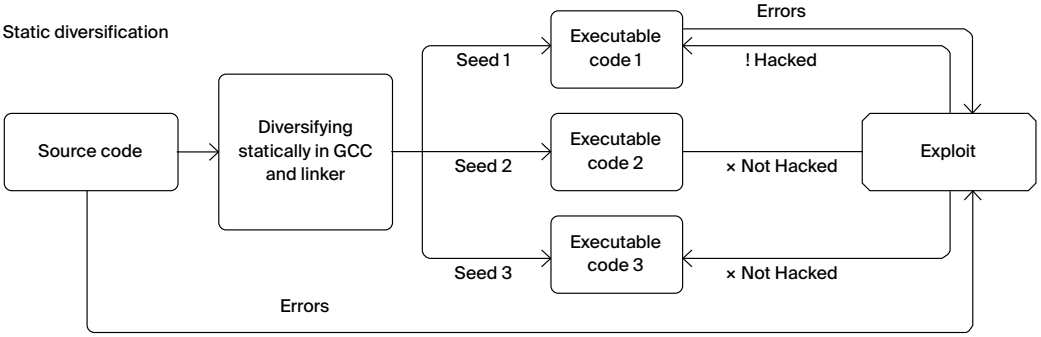
Obfuscator is a universal product that can be adapted to many system requirements. The production version is currently running on a Linux-based OS (version 2.6 and higher) with the Intel x86/x86-64 architecture support.

OBFUSCATOR WORKFLOW

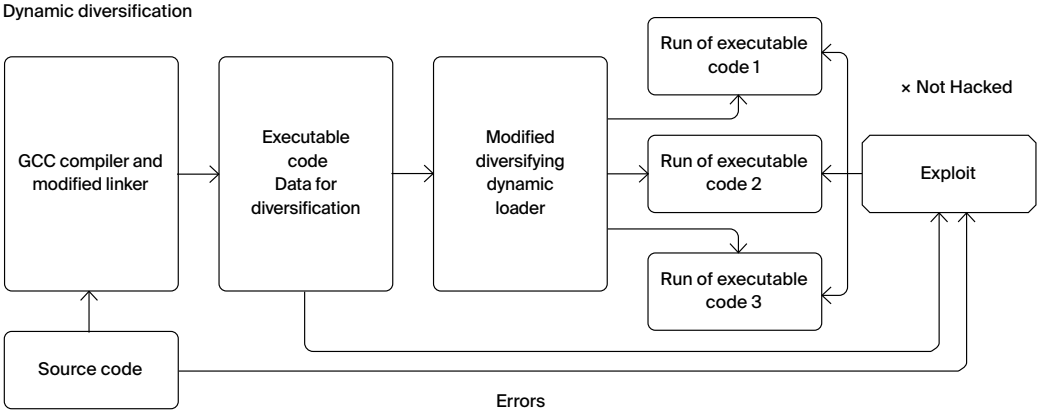
Standard compilation



Static diversification



Dynamic diversification



ISP RAS SOFTWARE ANALYSIS PLATFORM BASED ON QEMU

ISP RAS Foundation Platform for creating program analysis systems is built on top of open source QEMU emulator. This framework is essential for organizing cross platform development. It supports reverse debugging and introspection features, as well as full system emulation mode for debugging low-level software.

FEATURES AND ADVANTAGES

QEMU supports emulation of more than 10 instruction set architectures (i386 and x86-64, ARM and Thumb, MIPS, PowerPC, etc.). It implements guest debugging via GDB Remote Serial Protocol and is compatible with IDA Pro, GDB, and various IDEs. QEMU supports full system emulation mode that allows debugging low-level software such as a bootloader and an OS kernel. The QEMU source code is regularly checked by static code analysis tools, including Coverity and Svace. Thus performing malware analysis with QEMU is more secure.

QEMU with reverse debugging and introspection support is available on the ISPRAS GitHub page: https://github.com/ispras/swat. The developed QEMU automatization tools are available at https://github.com/ispras/qdt, https://github.com/ispras/i3s.

ISP RAS QEMU FOUNDATION PLATFORM PROVIDES:

- A record and replay mechanism for a virtual machine:
  - The same VM execution is replayed every time, deterministically. All external events are recorded and replayed by the emulator. It makes finding bugs in multi-threaded applications (race conditions, deadlocks) easier;
  - GDB-compatible reverse debugging is implemented based on the record and replay mechanism. The debugging is performed by restoring previous VM snapshots and searching for the previous breakpoint stop or the previous instruction;
  - The minimum required information is recorded. This allows recording longer for debugging rarely occurring errors;
  - Low performance overhead caused by recording. This enables analysis of software that requires interacting with an uncontrolled external environment in real time.
- VM introspection solution (getting high-level information regarding guest OS work) without any guest OS kernel modifications or installing monitors:
  - Getting the list of executed system calls, accesses to named functions in shared libraries, the list of running processes, the list of open files and loaded modules;

- Supports all Linux-based virtual machine images as well as embedded software images for various devices;
- WinDbg server support in QEMU that allows showing guest software information in terms of Windows kernel abstractions. There is no need to enable the OS debugging mode in the guest OS.

- Speeding up QEMU development:
  - Faster development of dynamic analysis tools that can analyze binary code for specific hardware;
  - Automated support for new processor architectures using a machine instruction decoder generator and a C-like language for describing machine instructions semantics;
  - An automatic tool for preliminary virtual machine testing. The tool only requires GNU Binutils and a C compiler;
  - A tool for automating QEMU virtual devices development;
  - VM generation tool in the form of QEMU module source code. The tool can create VMs from both existing devices and new devices out of Python description. The tool provides GUI for sketching the virtual machine;
  - A Python API for an automated debugging via GDB Remote Serial Protocol. It is used to debug QEMU, the guest OS, or both at the same time.
- Convenience and user experience:
  - Easy QEMU extension due to open source code and own ISPRAS toolkit for speeding up development;
  - Binary code analysis without any guest OS modifications;
  - VM introspection mechanism that can be extended using plugins;
  - A convenient API for developing own introspection plugins;
  - Can be easily adapted for specific use cases;
  - Support for latest QEMU versions that have support for newest peripherals and CPUs.

WHO IS ISP RAS FOUNDATION PLATFORM TARGET AUDIENCE?

- Bootloader, driver, OS and other system software developers.
- DevOps teams for reproducing of software bugs, cross-platform development, and scalable cloud testing.
- Programmers analyzing potential malware.
- Software certification engineers.

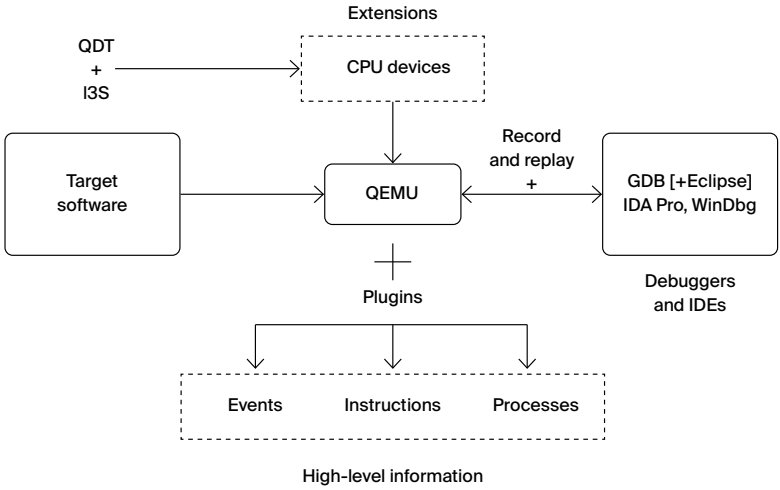
SUPPORTED GUEST PLATFORMS

Emulation of the following ISAs: i386, x86-64, ARM, MIPS, PowerPC, and others.
Guest systems supported by the introspection mechanism: Windows XP (x86), Windows 10 (x86-64), Linux 2.x-5.x (x86, x86-64, ARM, AArch64).

ISP RAS QEMU DEPLOYMENT STORIES

The QEMU community has accepted ISP RAS patches for the record and replay mechanism and added them to the open source QEMU version 3.1.

WORKFLOW


Extensions
QDT
+
I3S
CPU devices
Target
software
QEMU
Record
and replay
+
GDB [+Eclipse]
IDA Pro, WinDbg
Debuggers
and IDEs
Plugins
Events
Instructions
Processes
High-level information

KLEVER: A VERIFICATION FRAMEWORK FOR MODELS OF INDUSTRIAL SOFTWARE

Klever is a framework for verifying models that are automatically extracted from large software systems' source code written in the C programming language. Klever allows specifying various security and safety requirements and verifying them automatically with the preconfigured precision level.

FEATURES AND ADVANTAGES

Klever is a result of advanced research and development in the field of automated extraction and verification of program models. The framework base includes per-component verification, environment modeling, and requirements specification methods. This allows applying formal methods to the industrial software of hundreds of thousands or millions of lines of the C source code. Klever is an open-source project (https://forge.ispras.ru/projects/klever).

KLEVER PROVIDES:

- Thorough sound analysis of industrial software (allows detecting all possible errors of specified types and proving program correctness under explicitly stated assumptions).
- Scalability. Modular program verification allows applying the most rigorous program analysis methods to the large code base. The methods are model checking and symbolic execution.
- Adapting software verification framework to customer needs. Developing specifications for modeling target programs' environments and for detecting violations of program specific requirements. This specific customization is performed in addition to checking regular safe programming rules for the C language.
- Comprehensive representation of found faults. When an error is detected, the verification system provides the detailed error trace that includes concrete variable values and called functions' arguments.
- A convenient multi-user web-interface for setting and running verification and for expert analysis of verification results.

WHO IS KLEVER TARGET AUDIENCE?

— Companies developing safety-critical and security-critical software.
— Certification laboratories.

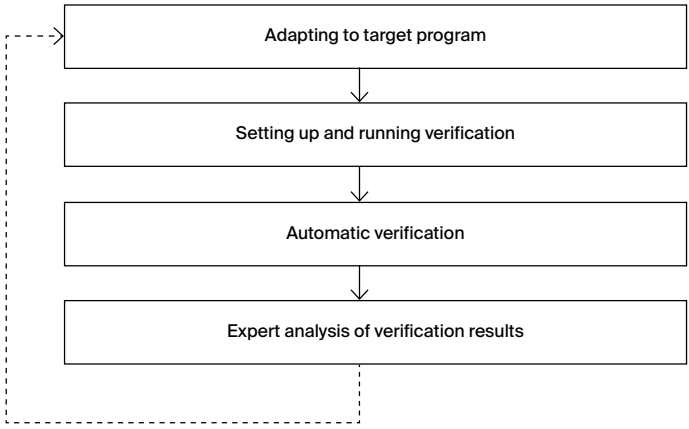
KLEVER DEPLOYMENT STORIES

The Klever verification system is mostly used for thorough checking of various operating system kernels and drivers. To showcase Klever features, it was used for verification of Linux kernel device drivers. As a result more than 400 errors of the following types have been found: buffer overruns, null pointer dereferences, uninitialized memory usages, double or incorrect memory deallocations, memory leaks, race conditions and deadlocks, incorrect function calls (depending on a certain context), incorrect initialization of Linux kernel data structures etc. Linux kernel developers have acknowledged these errors.

SYSTEM REQUIREMENTS

Ubuntu 18.04/20.04, at least 4 x86-64 CPU cores, 16 GB of memory, 100 GB of disk space.

WORKFLOW



LINGVODOC: A VIRTUAL LABORATORY FOR DOCUMENTING ENDANGERED LANGUAGES

Lingvodoc is a system intended for collaborative multi-user documentation of endangered languages, for creating multi-layered dictionaries and performing scientific work with the received sound and text data. It is a joint project with the Institute of Linguistics of the Russian Academy of Sciences and Tomsk State University. Lingvodoc is under active development since 2012 and can be found on lingvodoc.ispras.ru.

FEATURES AND ADVANTAGES

Lingvodoc is an open source cross-platform system based on an innovative research (https://github.com/ispras/lingvodoc, https://github.com/ispras/lingvodoc-react).

LINGVODOC PROVIDES:

- Collaborative work on dictionaries (as opposed to the similar Starling project that doesn't support this feature).
- Saving full history of user actions.
- Working with audio-textual corpuses and dictionaries simultaneously based on the integration with the ELAN system developed by Max Planck Institute of Psycholinguistics (Netherlands).
- Creating and editing unidirectional and bidirectional connections between lexical entries within dictionaries as well as external connections between dictionaries.
- Recording, playing and storing sounds with markup (in WAV, MP3 and FLAC formats), as well as constructing vowel formants followed with data visualization.
- Advanced search that supports multiple parameters (as opposed to the similar TypeCraft project).
- Ability to search data on a map with automatic construction of isoglosses.
- Conflict-free bilateral delayed synchronization.
- High automation level (compared to the similar Kielipankki project): ability to carry out automatic etymological and phonetic analysis.
- Creating dictionaries of any structure, such as typical two-layer dictionaries with lexical entry layer and paradigms layer or multi-layer dictionaries. Importing dictionary structures is also supported.

- Algorithms mimicking scientists' work for phonetic and etymological analysis.
- Support for storing text corpora in the Word format and dictionaries in the Excel format.
- Built-in morphological analysis for the languages of Russia in the Aperitum format.
- A convenient interface for disambiguating homonyms after completing morphological analysis.
- Either using the ISP RAS cloud infrastructure resources or locally deployed resources with data isolation.
- Desktop and web-based versions.
- Open registration (confirmation required).
- Fast development for extending the system features as well as easy adaptation to another scientific field.

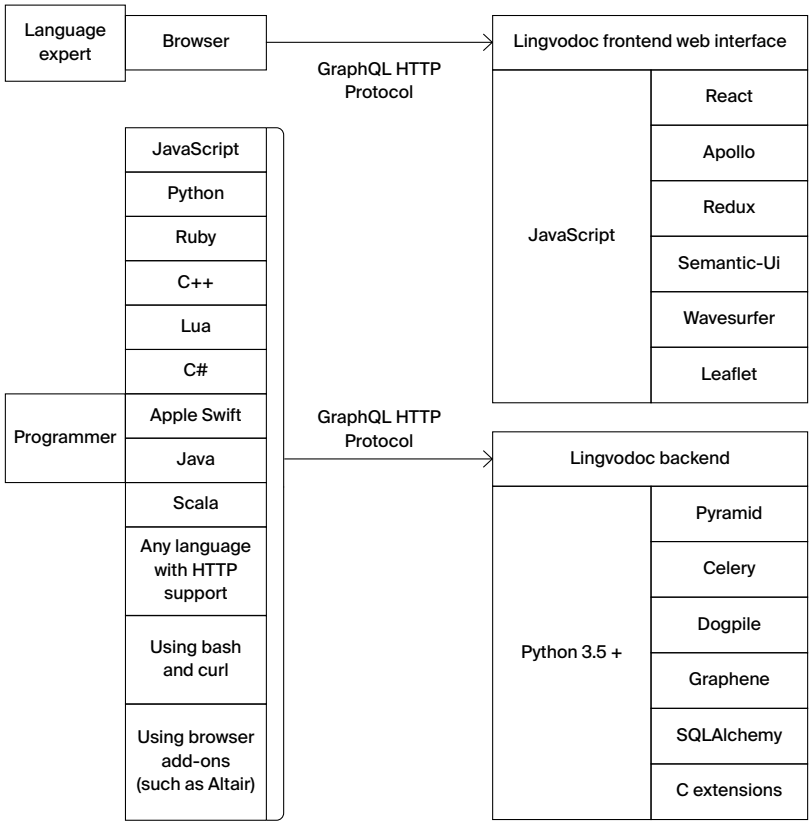
WHO IS LINGVODOC TARGET AUDIENCE?

Lingvodoc is designed primarily for linguists performing a research in the area of documenting the endangered languages in Russia. However, it is possible to adapt the technology for other purposes.

LINGVODOC DEPLOYMENT STORIES

Lingvodoc is currently used by philologists in 29 universities and scientific centers of 16 cities, including Tomsk State University, Institute of Philology (Siberian Branch of RAS), Institute of History, Language and Literature (Ufa Scientific Center of RAS), Udmurt Federal Research Center UB RAS, North-Eastern Federal University, Ugra State University, Institute of Linguistics, Literature and History (Karelian Research Centre of RAS), Murmansk Arctic State University.

LINGVODOC WORKFLOW



MASIW: SUPPORT FOR DESIGNING HIGHLY RELIABLE SOFTWARE SYSTEMS

MASIW is a toolset for developing highly reliable hardware and software systems for avionics, medicine, and other safety critical areas. It is designed for engineers creating airborne hardware/software systems that are developed using the integrated modular avionics (IMA) approach. MASIW can be easily adapted for other application areas.

FEATURES AND ADVANTAGES

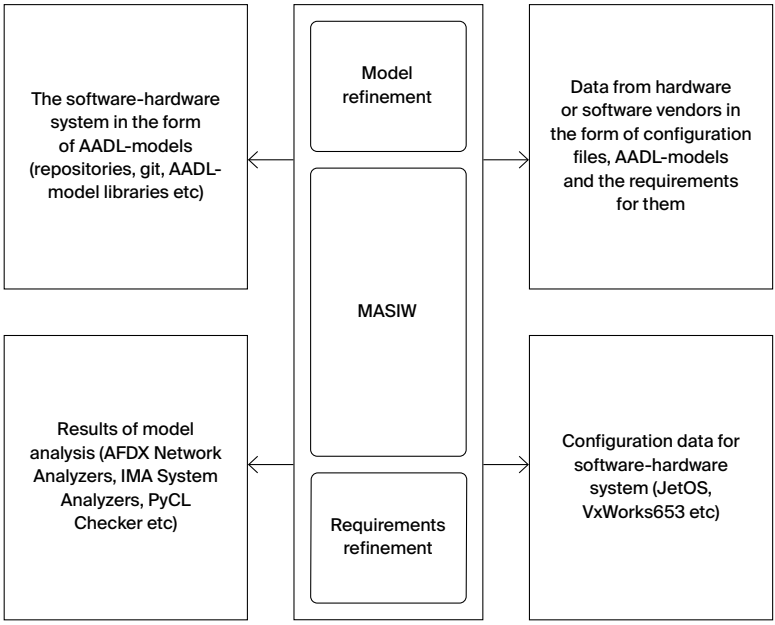
MASIW is the technology for optimizing the development and verification process of complex hardware/software systems. It allows performing a preliminary quality assessment of the product before making the first prototype, as well as performing the fault tolerance analysis. This reduces the risk of errors and defects. MASIW is being developed jointly with GosNIIAS. Despite the presence of the OSATE tool at the start of development, MASIW currently is more functional in the areas of verification, static, and dynamic analysis.

MASIW PROVIDES:

- Creation, editing and management of models based on the AADL modeling language:
  - creation and editing of models using the text and diagram editors;
  - support for team development with the ability to track and modify individual elements of a model;
  - support for the third-party AADL models reuse.
- Model analysis:
  - hardware+software system structure analysis: hardware resources sufficiency, interfaces consistency, etc.;
  - verification of the developed system for compliance with the requirements;
  - transmission characteristics analysis for the AFDX networks: message latencies, port queue depth, etc.;
  - generation and analysis of fault trees (FTA) to determine probabilities of high-level fault events;
  - architecture-model based analysis of failures and their consequences, including generation of special descriptive tables;
  - simulation of hardware+software system model with user reports generation including software-in-the-loop execution of on-board partitions with RTOS co-emulated with QEMU and with a universal AADL model simulator.
- Model synthesis:
  - distribution of software applications by computational modules taking into account hardware resource

limitations and additional restrictions regarding reliability and security;
  - processor schedule generation (in particular, for ARINC-653 compatible real-time operating systems).
- Configuration data generation:
  - development of specialized configuration data tools based on the provided software interface (API);
  - configuration data generation for the VxWorks653 RTOS and for the AFDX network equipment.
- The ability to extend the toolset by creating own modules.

MASIW WORKFLOW



MICROTESK: A TEST PROGRAM GENERATOR

MicroTESK is a reconfigurable and extendable framework for generating test programs for functional verification of microprocessors. MicroTESK allows automatically constructing test program generators based on formal specifications of microprocessor architectures. MicroTESK supports a wide range of architectures including RISC, CISC, VLIW, and DSP. MicroTESK supports online test program generation.

FEATURES AND ADVANTAGES

MicroTESK is a set of technologies for industrial use that includes the basic modeling framework (building microprocessor models based on formal specifications) and the generation framework (building test programs based on test templates). MicroTESK delivers value similar to its global competitors (e.g., Genesys Pro and RAVEN) but outperforms them via increased usability and performance. Also, it is distributed under the open-source Apache 2.0 license. MicroTESK is available at the ISP RAS website: https://forge.ispras.ru/projects/microtesk. The technology is also presented at http://www.microtesk.org.

MICROTESK PROVIDES:

- Using formal specification as a source of knowledge about the microprocessor under verification:
  - architecture specification in the nML language (registers, memory, addressing modes, instruction logic, text/binary instruction representation);
  - additional memory subsystem specifications in the mmuSL language (memory buffer properties (TLB, L1, and L2), address translation logic, read/write operations logic);
  - an option to make a transition to formal verification and to automatic toolchain generation for the microprocessor under development (disassembler, emulator, etc.).
- Test programs generation based on object-oriented test templates:
  - test templates in the Ruby language (so that the templates are human readable and easy to support);
  - allows using different generation techniques for instruction sequences and test data simultaneously (random generation, combinatorial generation, constrained-based generation, etc.);
  - generation framework scalability (can develop complex test templates at low cost due to reuse).
- Wide range of supported microprocessor architectures:
  - supporting architecture specific features for various architectures (RISC, CISC, VLIW, DSP) at the generator development framework level;

  - MicroTESK-based test program generators have been developed for RISC-V, ARM, MIPS, and PowerPC architectures;
  - multicore architectures are supported.
- Quick framework adaptation for the new microprocessor architecture with minimal costs and automatic information extraction for test situations (due to formal specifications).
- Convenient language for developing test templates that allows describing complex verification scenarios quickly.
- Support for online test program generation for performing post-silicon verification of the target microprocessor. The online generation is performed by an executable generator included into MicroTESK. The generator constructs test sequences using formal specifications, and then modifies the sequences by making functionally equivalent substitutions. It also allows repeated execution of the test sequences on the target microprocessor.

SYSTEM REQUIREMENTS

Windows or GNU/Linux-based OS, Java 8.

MICROTESK DEPLOYMENT STORIES

MicroTESK is developed since 2007. It was used in various Russian and international projects on developing modern industrial microprocessors, including production projects on verifying ARMv8, MIPS64, and RISC-V microprocessors.

WORKFLOW

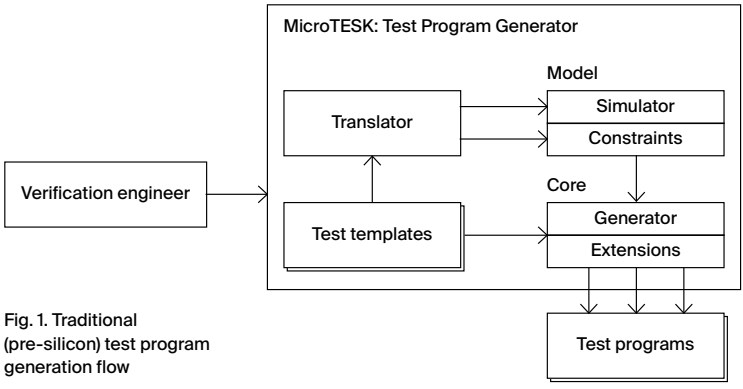


Fig. 1. Traditional (pre-silicon) test program generation flow

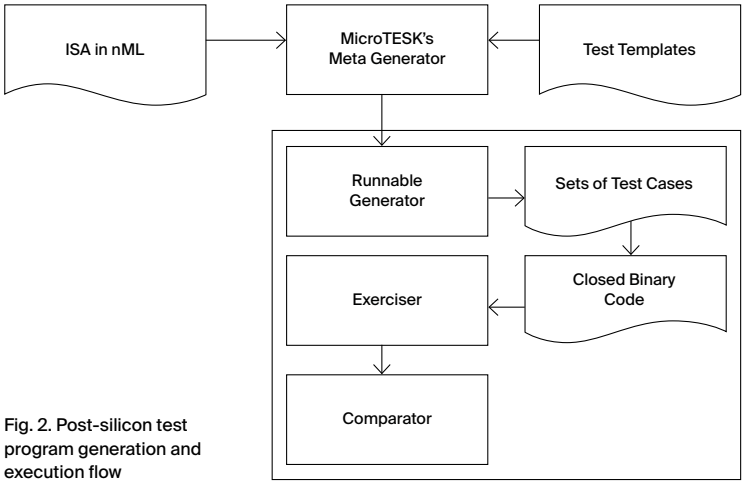


Fig. 2. Post-silicon test program generation and execution flow

PROTOSPHERE: A NETWORK TRAFFIC ANALYZER

Protosphere is a system of deep packet inspection (DPI). It can serve as a part of intrusion and information leak protection systems. Protosphere detects inconsistencies between a protocol specification and the actual traffic. It allows to add support quickly for new protocols (either open or closed) due to the flexibility of its internal representation.

FEATURES AND ADVANTAGES

Protosphere is an innovative system based on the innovative research in the area of network traffic analysis. It combines the key features of similar tools (e.g. Wireshark, Microsoft Message Analyzer) with a universal data representation model that enables rapid expansion of analysis capabilities.

PROTOSHPERE PROVIDES:

- Advanced system core:
  - universal data representation model used when parsing network traffic;
  - processing of corrupted, reordered or duplicated packets, as well as handling of packet loss and processing of asymmetric traffic;
  - compressed/encrypted data analysis;
  - support for tunnels of arbitrary configuration;
  - support for network flows causality.
- Support for all stages of network trace analysis (each stage has a visualization component that are synchronized between stages):
  - network connections localization in the network interaction graph and the network flow tree;
  - detailed view of the selected connections in the timeline diagram;
  - interactive visualization of the parsed network packets in the stream tree;
  - detection of discrepancies between a protocol implementation and the actual traffic in the diagnostic log;
  - arbitrary OSI-layer data extraction and analysis (L7+).
- Easy support for new protocols:
  - access to parsing results via API;
  - localize parsing errors;
  - debugging the module being developed on real-time traffic and network traces.
- Support for both online and offline analysis modes.
- Advanced GUI provides choice of the most convenient way to present the analysis results.
- Universal data representation model to accelerate customization:
  - support for new protocols;
  - extract data in a desired format;

- configuring the analysis results format.
- Adjustment to network bandwidth and available computational resources to find a balance between accuracy of the analysis and the resources consumed.

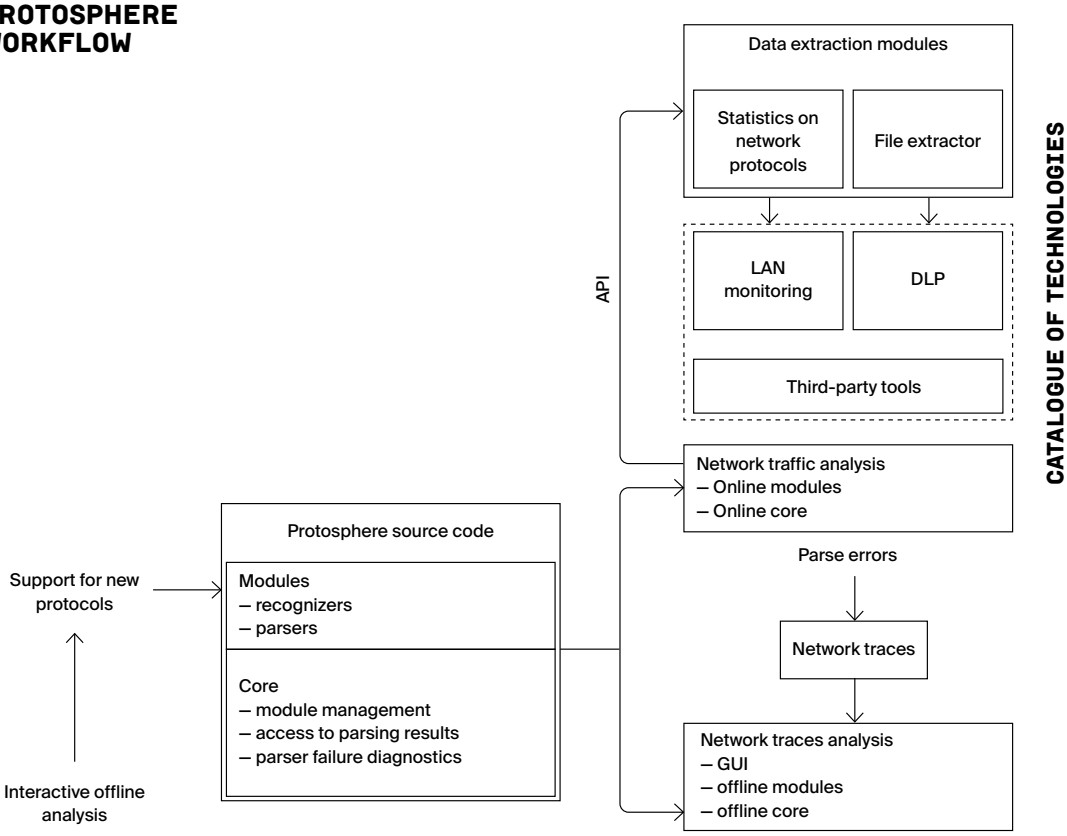
WHO IS PROTOSPHERE TARGET AUDIENCE?

- Companies that are testing network protocol implementations including those in embedded OS and network hardware.
- Developers of network security tools, such as firewalls and IDS/IPS.
- Manufacturers of network hardware that must be certified.
- Companies requiring real-time control and monitoring of network channels.

SUPPORTED PLATFORMS AND ARCHITECTURES

Architecture: Intel x86-64.
Platforms: Windows and Linux-based OSes.

PROTOSPHERE WORKFLOW



RETRASCOPE: STATIC ANALYSIS OF HDL DESCRIPTIONS

Retrascope is a functional verification toolkit for digital hardware designs. Retrascope provides automated engines for code analysis, formal model extraction and functional test generation. The toolkit accepts digital hardware module descriptions as inputs, written on the synthesizable subset of Verilog and VHDL languages, as well as their behavioral specifications.

FEATURES AND ADVANTAGES

Retrascope is an open source toolkit for functional verification of digital hardware modules. The toolkit implements formal model extraction and analysis methods and functional test generation methods. A component-based architecture of Retrascope allows developing hybrid formal model verification techniques by combining various analysis methods. Retrascope is available at https://forge.ispras.ru/projects/retrascope.

RETRASCOPE PROVIDES:

- Formal model extraction from source code:
  - control flow graph;
  - guarded actions decision diagram;
  - high-level decision diagram;
  - extended finite state machine.
- Functional test generation:
  - random tests;
  - dead code detection;
  - typical read-write error detection;
  - user-defined property checking.
- Formal model analysis (model checking) that verifies specification conformance via:
  - PSL;
  - SystemVerilog Assertions.
- Graphical user interface based on Eclipse IDE (command line interface is also available):
  - running the tool with specified parameters;
  - model visualization (Zest, GraphML).
- Open source (Apache License Version 2.0).
- Extensibility at the source code level:
  - new models;
  - new engines for analysis and test generation.

- Open APIs and formats allow using various tools for analysis and verification:
  - SMT solvers via the SMT-LIB v2 language;
  - model checkers via the SMV language;
  - functional tests via Verilog/VHDL languages and the VCD format.

WHO IS RETRASCOPE TARGET AUDIENCE?

- Companies working in the area of digital hardware design.
- Research groups in the field of digital hardware verification.

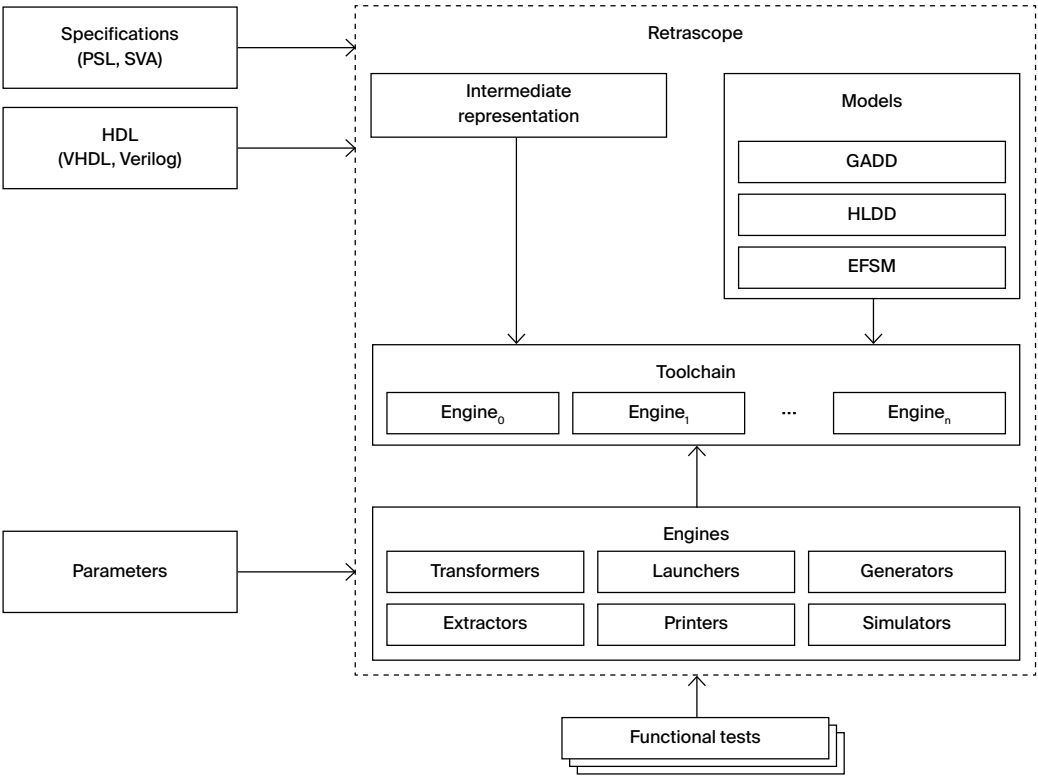
RETRASCOPE DEPLOYMENT STORIES

Retrascope is at the research prototype stage with active development.

SYSTEM REQUIREMENTS

Windows or GNU/Linux-based OS, Java Runtime Environment 8.

RETRASCOPE WORKFLOW



SAFE COMPILER

The safe compiler avoids introducing new vulnerabilities in program's binary code w.r.t. its source code when aggressively optimizing (e.g., when making use of source code constructs exhibiting undefined behavior). The compiler tries to restrict optimizations as little as possible, which allows avoiding the large performance drop compared to the solution with all optimizations turned off.

FEATURES AND ADVANTAGES

The safe compiler is based on the GCC compiler and can act as a drop-in replacement for GCC (for example, when building the complete Linux distribution). The compiler retains the generated code quality and produces the ready-to-use safe build of a program.

THE SAFE COMPILER PROVIDES

- Refined compiler optimizations for conservative treatment of source code places with undefined behavior so that for these places program semantics gets defined safely and naturally.
- Forced initialization for uninitialized automatic variables.
- Issuing warnings when detecting undefined behavior.
- Adding dynamic checks for certain constructs to prevent exhibiting undefined behavior during program execution.
- Diversifying code generation during either compilation or program execution.
- No need to modify either source code or build system configuration, which makes using the compiler as simple as possible.
- Three different safety levels that provide trade-offs between generated code safety and performance. The lowest level is the third; the highest is the first one.

The safe compiler performs the following actions:

On the third level:

- Avoiding integer overflow, accessing objects via pointers of incompatible types, dereferencing null pointers, using compiler builtins instead of standard library implementations for input/output functions and for functions working with memory.
- Detecting division by zero, incorrect bitwise shifts, accesses beyond stack frames, array loads/stores outside of the memory allocated for the array. Detecting automatic variables that are stored in registers during function calls.

On the second level:

- Analyzing arguments of bitwise shifts, redundant memory operations, data alignment when working with vector instructions, address arithmetic when optimizing memory accesses and changing their order.
- Initializing all automatic variables (with zero) that are not initialized explicitly by the user.

— Treating certain compiler warnings as errors and stopping compilation when they are issued.

On the first level:

— Generating unique memory layout for function code either statically during compilation or when performing dynamic linking.
— Adding machine code that aborts the program when detecting undefined behavior during program execution (sanitization) in the following situations:

1 Integer and floating point operations:
   — loading a non-boolean value in a boolean variable;
   — floating point conversion that results in either integer or floating point overflow;
   — performing a bitwise shift with a negative shift value or with a shift value that is equal or greater than the shifted type width;
   — signed integer operation with the result that is non representable in the output type;
   — integer division or module with the divisor equal to zero.
2 Pointer and array operations:
   — loads/stores via incorrectly aligned or null pointer;
   — array loads/stores using the address outside of the memory allocated for the array;
   — passing null pointer as a function parameter marked with the nonnull attribute;
   — address arithmetic resulting in integer overflow;
   — returning null value out of function that is marked with the returns_nonnull attribute;
   — allocating an automatic VLA array with incorrect size (zero or negative).
3 Function operations:
   — a function pointer call via a pointer whose type does not match the function prototype;
   — returning from a non-void function without actually executing the return statement;
   — calling a compiler builtin with incorrect arguments;
   — reaching a program point during program execution that is marked in the source code as unreachable.

WHO IS THE SAFE COMPILER TARGET AUDIENCE?

— Operating system developers.
— Companies developing high-level safe and secure software.

SAFE COMPILER DEPLOYMENT STORIES

The safe compiler is deployed in a number of Russian companies and government institutions as an add-on to the ISP Crusher framework.

SUPPORTED PLATFORMS

Linux-based OS for x86 32/64 and ARMv7.

SCINOON: EXPLORATORY SEARCH SYSTEM FOR SCIENTIFIC GROUPS

SciNoon is a system for collaborative exploration of scientific papers. SciNoon is an essential tool for a group of researchers to dive quickly into the new area of knowledge and to find answers on their questions, following up with tracking new research on the topic of interest with highly customizable alerts.

FEATURES AND ADVANTAGES

SciNoon is an innovative system designed to optimize long-term teamwork with scientific papers. The papers could be added both from search systems and from digital libraries (like Google Scholar, arxiv.org, Semantic Scholar, PubMed) or could be uploaded directly as PDF files. The key feature of SciNoon is graphical research maps that all team members can add papers to.

SCINOON PROVIDES:

- Shared workspace for collaborative processing of found scientific papers.
- Zooming research map to control paper visualization details.
- Deduplication and cleansing of the uploaded metadata made possible by the internal database that allows discovering connections between papers or authors.
- Citation context classification into five classes depending on a citation role.
  - Background: a cited paper contains general information about the research area;
  - Use: a citing paper uses methods, data, and so on from the cited paper;
  - Compare: a citing paper points out differences (or similarities) with the cited paper;
  - Extend: a citing paper continues the research from the cited paper;
  - Weak: a citing paper criticizes the cited one pointing to the authors' mistakes.
- Possibility to find relevant papers without a keyword search using an integrated recommendation system.
- A customizable list of questions whose answers should be looked up in a paper. Based on the retrieved answers the paper is visualized differently on the research map.
- Possibility to group similar papers into clusters.
- Notifications for team member actions, as well as support

for quick opinion exchange within the team and helping each other.
- Collected answers analysis via built-in spreadsheet view and an option to export data to a CSV file for more complex processing.
- Tracking new papers on a given research topic and updating previously gathered results.

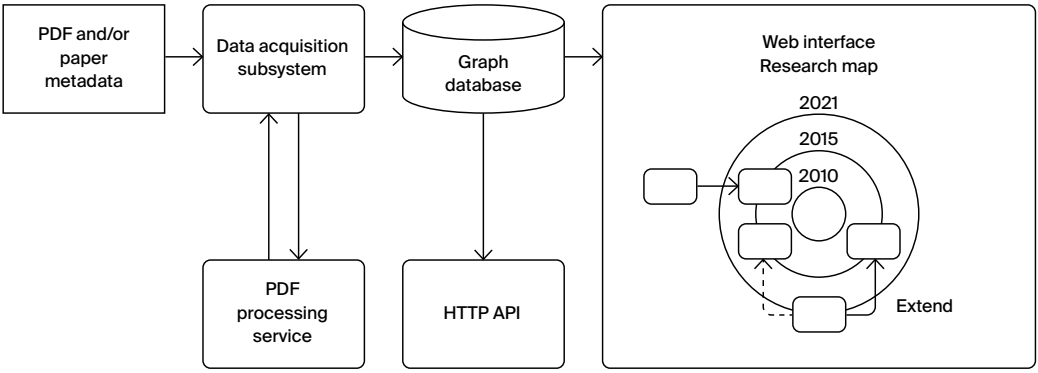
WHO IS SCINOON TARGET AUDIENCE?

- R&D department researchers that need a fast solution to their scientific problem.
- Scientists who need a tool for collaborative teamwork.
- Scientific advisers and their students who are doing exploratory search on research projects.

SCINOON DEPLOYMENT STORIES

SciNoon is used in ISP RAS while doing research and when advising students.

SCINOON WORKFLOW



SVACE STATIC ANALYZER

Svace is an essential tool of the secure software development life cycle, the main static analyzer that is used in Samsung Corp. It detects more than 50 critical error types as well as hundreds of coding issues. Svace supports C, C++, C#, Java, Kotlin, and Go. Svace is included in the Unified Register of Russian Programs (No.4047).

FEATURES AND ADVANTAGES

Svace is an innovative technology based on years of research that constantly evolves for customer's needs. It combines the key qualities of foreign competitors (Coverity Scan Static Analysis, Klocwork Static Code Analysis, Micro Focus Fortify Static Code Analyzer) with the unique open industrial compilers usage to provide the maximal support level for new programming language standards.

SVACE PROVIDES:

- High-quality deep analysis:
  - an accurate representation of the source code (due to integration with any build system);
  - full path coverage taking into account function calling contexts when searching for complex defects;
  - high percentage of true positives (60-90%).
- Scalability and high speed:
  - parallel analysis using all available processor cores;
  - ability to analyze software with the code size of tens of millions of lines (analysis of the Tizen 6.5 mobile OS having 44 million lines of code takes 7-8 hours);
  - supporting incremental system analysis in addition to the full analysis mode (performs a quick re-analysis of the recently changed source files).
- Convenient warnings viewing interface:
  - detailed error description with code navigation;
  - review interface for marking true and false positives;
  - analysis results migration between runs with hiding any issues previously marked as false positives.
- Accelerated customization (configuring existing detectors as well as writing individual ones available exclusively to this customer; creating tailored user interfaces).
- Ultra fast adaptation to new environments and tools (adding new compilers within 1-2 weeks, in complex cases up to 2 months).
- Full compatibility with regulatory documents and requirements of regulators (FSTEC of the Russian Federation).
- Can be used for adhering to the GOST R 56939-2016 requirements and to the requirements of the FSTEC regulation document mandating software vulnerability detection process (when certifying software within Russia).

WHAT IS SVACE TARGET AUDIENCE?

- Companies aimed at software development with a special focus on high reliability and security.
- Companies that need to certify the developed software.
- Certification laboratories.

SVACE DEPLOYMENT STORIES

Svace is the main static analyzer used in Samsung Corp. since 2015. It is used to check the company's own software based on Android OS as well as the Tizen OS source code. Tizen is used in smartphones, infotainment systems and Samsung home appliances. Since 2017, Svace checks all changes submitted for review and inclusion in the Tizen OS.

Within Russia, Svace is deployed in more than 40 companies and certification labs, including RusBITech, Kaspersky, Postgres Professional, Security Code, and Swemel.

SUPPORTED PLATFORMS AND ARCHITECTURES

- Host platforms for the analyzer: Linux kernel based OS (version 3.13 and later), Windows 7 SP1 and later, including WSL 1 and 2, macOS on x86-64 (starting from 10.12 Sierra); x86-64 and ARM64/Linux architectures; x86 architecture for build capture.
- Target architectures of the analyzed code: Intel x86/x86-64, ARM/ARM64, MIPS/MIPS64, Power PC/Power PC 64, RISC-V 32/64, SPARC/SPARC64, Hexagon (AEON, TriCore, HIDSP, OpenRISC targets of the GCC compiler are partially supported).

SUPPORTED COMPILERS

For C/C++ (up to 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 and T900 Family, Samsung CalmSHINE16 Compilation Tools, Texas Instruments TMS320C6* Optimizing Compiler.

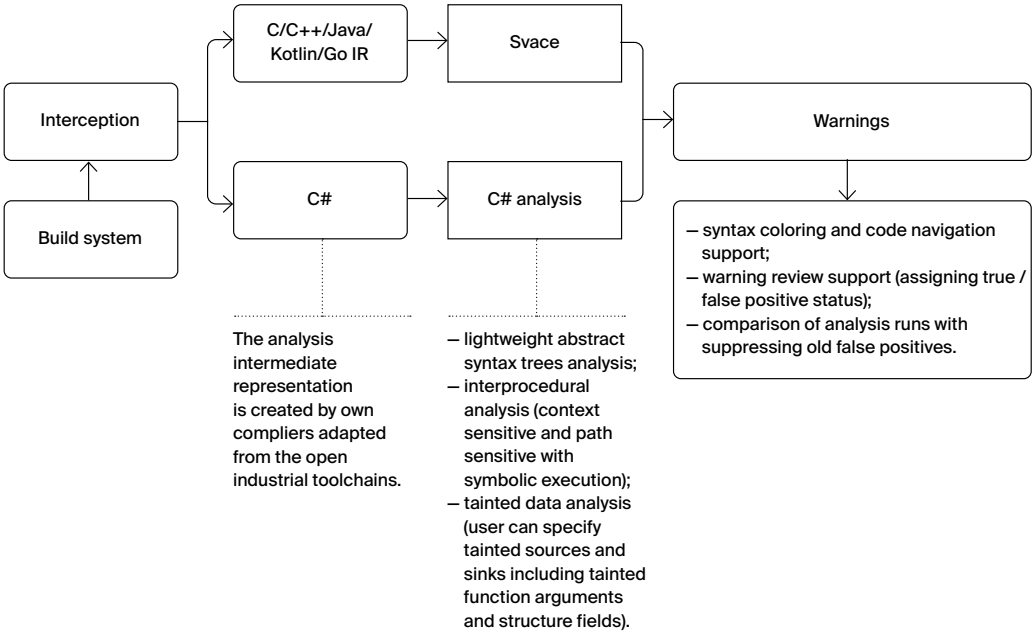
For C# (up to C# 9): Roslyn, Mono.

For Java (up to Java 11): OpenJDK Javac Compiler, Eclipse Java compiler.

For Kotlin: Kotlin 1.5.

For Go: Go 1.15.

SVACE ARCHITECTURE



TALISMAN: A DATA PROCESSING FRAMEWORK

Talisman is a unified set of tools that automate typical data processing tasks, such as data retrieval, integration, analysis, storage and visualization. It makes possible the fast development of specialized multi-user analytical systems that merge and work uniformly with the data from private databases and Internet sources (including social networks).

FEATURES AND ADVANTAGES

Talisman unifies the tools necessary for big data. It builds on two ISP RAS technologies: Dedoc, a system for document structure retrieval, and Texterra, a platform for extracting semantics from text. Talisman is comparable to world's best competitors (Palantir Gotham and IBM Watson Content Analytics). Its advantage is automating routine analysis processes with state of the art research results (reducing resources required for manual analysis).

TALISMAN PROVIDES:

- A rich set of reusable components that have APIs for easy management and integration:
  - Data retrieval components. They include a framework for Internet data collection, namely, from social media (Facebook, VKontakte, Twitter, Instragram, Odnoklassniki, Youtube, LinkedIn etc.), blogs, news, MediaWiki sites, developer portals etc. Also there's a system for importing data from file storages and databases.
  - Automatic data analysis components. Analysis tools are designed as Docker containers that are managed through APIs by the Talisman.Stream system (included in the Unified Register of Russian Programs as No.6045). The output is stored on hard disks or in databases (PostgreSQL, ElasticSearch, Cassandra etc.). The basic services used are the Tesseract OCR system and own ISP RAS tools.
  - Storage and indexing components. These include a number of databases and information search engines that store source data, automatic analysis results, and results of manual user work.
- An easy to use web interface that unifies all components requiring user interaction.
- A flexible modular architecture that allows adding new features to the interesting components without changing others.
- A scalable architecture that allows processing and storing more data just by adding more hardware without any software change.

- Specialized components that monitor system status, manage event log, perform deployment, authentication and authorization, access control, and unidirectional data transfer.
- Tools and methods for training machine learning models as well as for transferring existing algorithms to other knowledge domains.
- A configurable knowledge domain scheme that can be changed by a user when the system is in operation.
- An onsite deployment option using existing customer hardware or the new hardware provided and configured with the framework.
- Integration with private customer systems via provided component APIs.
- Closed license free. Talisman is based on open source and know-how ISP RAS tools.

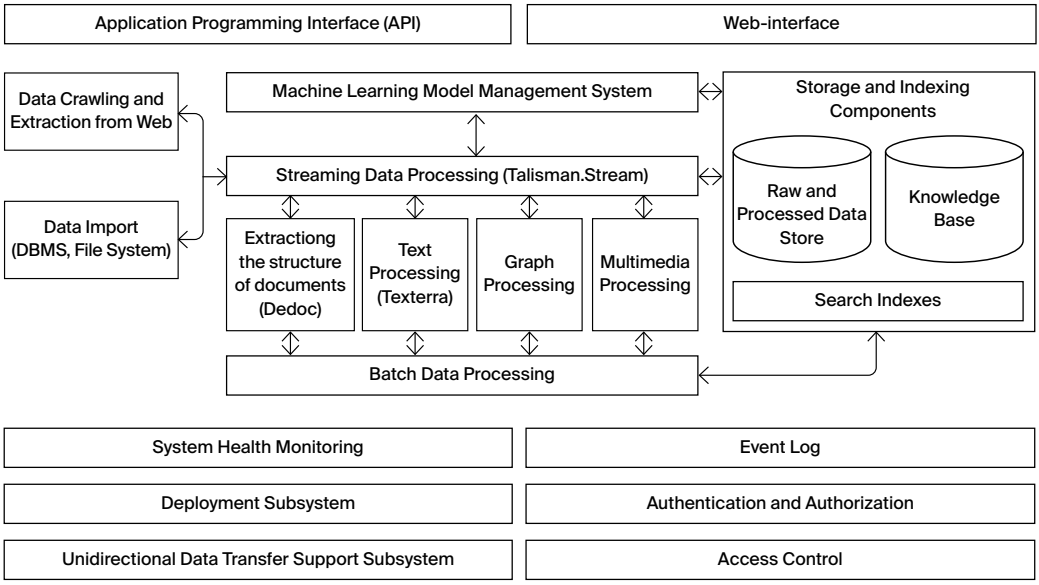
TALISMAN APPLICATION AREAS

- Automated knowledge base construction for a given knowledge domain and non-stop monitoring for new information regarding objects of interest.
- Competitor intelligence based on open sources (OSINT).
- Detecting information campaigns that aim to manipulate target audience as well as detecting the target audience for a campaign.
- Detecting and analyzing means for disseminating information (used resources, people, bots) as well as analyzing community member communication roles (news source, opinion leader, disseminator, moderator, bot, commentator).
- Reputation management for persons and companies, including monitoring relevant news, detecting possible complaints, monitoring leaks and information disclosure.
- Staff management optimization including efficient recruitment, data verification, detecting hidden activity, assisting in developing motivational systems.
- Evaluating activity effectiveness objectively and testing strategies on a target audience to gather feedback.
- Finding and managing social tension to detect and prevent conflict escalation.

SUPPORTED LANGUAGES

Talisman supports languages recognized by the Texterra analyzer, namely, Russian and English.

TALISMAN WORKFLOW



TEXTERRA: A SEMANTIC ANALYZER

Texterra is a scalable platform for extracting semantics from text. It contains the complete fundamental set of technologies for creating multifunctional applications for text analysis. Texterra bases its semantic analysis approach on concept identification. The platform is included in the Unified Register of Russian Programs (No.4048).

FEATURES AND ADVANTAGES

Texterra performs a unique analysis of Russian texts based on the identification of concepts instead of just words. It differs from foreign competitors by paying the most attention to the Russian language. The analyzer builds on fundamental research results and integrates with the Elasticsearch search system greatly expanding its capabilities. The successful combination of technologies allows the platform to compete with the projects similar to IBM Watson Natural Language Understanding.

TEXTERRA PROVIDES:

- High text processing speed (morphological analysis: 69 000 words per second, syntactic analysis: 39 100 words per second, coreference resolution: 10 100 words per second, full text analysis: approximately 13 600 words per second).
- Maximum attention to the Russian language (unlike similar spaCy and UDPipe projects, as well as IBM Watson Natural Language Understanding, which does not support the analysis of emotions and concepts in Russian texts).
- Large knowledge base (more than 7 million concepts).
- Building knowledge base without expert involvement (automatic construction and update using Wikipedia, MediaWiki, Linked Open Data, etc.).
- Scalability both in word processing speed and in knowledge base size (using Apache Ignite and the Asperitas cloud technology developed at ISP RAS).
- High text analysis accuracy due to a number of key features:
  - multi-level search by related concepts;
  - adaptability to slang, hashtags (#) and errors in text;
  - emotion analysis (with separation of attitude towards objects and their attributes);
  - determining relationships between people and companies based on text information;
  - detecting implicit object references in discussions.
- Fast adaptation and tailored solutions development.
- Supporting two use cases:
  - as a deployed software system on a customer's local server providing either HTTP REST-based or RMI protocol access;
  - online at https://texterra.ispras.ru/.

— Simple and fast support for specific domains and ability to integrate new languages backed up by a modern machine learning approach.

WHO IS TEXTERRA TARGET AUDIENCE?

— Corporate software developers (e.g. chat bot developers).
— Developers of semantic search systems for specific domains (such as information security, medicine, auditing, etc.).
— Developers of arbitrary text processing applications.

TEXTERRA DEPLOYMENT STORIES

Texterra has been productized in the joint projects with HP and Samsung (the project goals were to develop a technology for analyzing corporate reports or supporting smart TVs). Currently Texterra backs up several ISP RAS innovative products, e.g. Talisman social media analysis. A number of Russian government agencies also use Texterra.

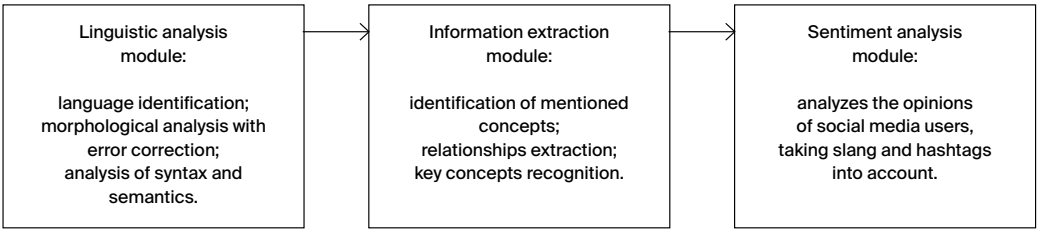
SUPPORTED LANGUAGES

Texterra analyzes Russian and English texts.

SYSTEM REQUIREMENTS

— A system platform supported by Java 8.
— 16 Gb RAM or more for each supported language.
— 64-bit operating system is recommended.

TEXTERRA WORKFLOW



ISP RAS: AN INNOVATION ECOSYSTEM

ISP RAS activities are aimed at deploying fundamental research results in industry. The institute's business model consists of three closely related activities producing a synergistic effect:

- project-oriented fundamental and applied research aimed at creating new technologies (under contracts with Russian and foreign companies, the Ministry of Science and Higher Education of Russia, RAS programs, grants from Russian Science Foundation and from Advanced Research Foundation, etc.);
- deploying new technologies in partner companies and developing innovations based on industry feedback;
- educating students and postgraduates based on developed technologies (while participating in the institute's research and industrial projects).

This model of industrial research plus education is well known and applied in research laboratories of leading universities (Stanford, MIT, Berkeley, Carnegie Mellon) and industrial giants (IBM, Intel), as well as in state research centers (INRIA, Fraunhofer). When implemented effectively, the model solves the problem of the gap between science and industry, and produces highly qualified specialists in system programming.

FUNDAMENTAL RESEARCH

Fundamental research and experimental works are necessary elements of the institute's activities, allowing it to move in line with the latest trends in the IT world, as well as to generate own ideas for projects with business partners. ISP RAS works on a large number of scientific and educational programs and cooperates with leading Russian and foreign universities and scientific centers, including Technion — Israel Institute of Technology; Fraunhofer Society – an association of applied research institutes, Germany; ITRI, an industrial research institute, Taiwan, and many others. This allows providing high quality research results, and ISP RAS reputation in academic and university circles helps presenting its technologies on international markets.

ISP RAS publishes its own journal called "Proceedings of ISP RAS" (indexed in the Russian Science Citation Index (RSCI) on the Web of Science platform and in Google Scholar).

The institute is also responsible for publishing and editing the RAS journal called "Programming" (indexed in the Web of Science and Scopus). Both are included in the journal list of Higher Attestation Commission (the VAK).

DEPLOYMENT

ISP RAS deploys its research results in various industrial and research organizations, which use and promote further the institute's technologies. Most of the contract work is performed with long-term partners who have been collaborating with ISP RAS for more than five years. Such companies include Samsung, Huawei, HP, Intel, Nvidia, Bentley Systems, Linux Foundation, and the Russian ones: Kaspersky Lab, Security Code, Open Mobile Platform, RusBITech, GosNIIAS, VimpelCom, Basalt SPO, Swemel.

SCIENTIFIC COLLABORATION

Long-term cooperation with ISP RAS can be organized in a form of a joint laboratory. Having permanent funding, they allow planning flexibly available resources as well as increasing competencies in the newly emerging areas of system programming and organizing the training of young specialists with the skills needed by partners.

Currently, the institute has joint laboratories with Samsung Electronics (focused on compiler technology including improving security for Android and Tizen OS) and Huawei (mainly focused on compiler technology and operating system components). Also there is a laboratory for solving continuum mechanics problems based on the Fanlight cloud platform. The lab successfully implements research projects for industrial enterprises. Finally, there is a linguistic virtual laboratory for documenting endangered languages that uses the Lingvodoc platform. The research is performed with the Institute of Linguistics (RAS), Tomsk State University and other universities and research institutes. In 2021, ISP RAS has opened the intelligent digital foresight and media data lab in collaboration with MGIMO University of the Ministry of Foreign Affairs.

CENTERS

The important mission of ISP RAS is creating and moderating multidisciplinary communities. In 2021, three such centers are launched and operating:

- World-class Research Center (WCRC) "Digital biodesign and personalized healthcare," jointly with Sechenov University, Institute of Biomedical Chemistry, Yaroslav-the-Wise Novgorod State University, Institute for Design-Technological Informatics of RAS;
- Technology center for security analysis of the Linux kernel, jointly with FSTEC of Russia and with active participation of leading Russian IT companies;
- Trusted AI center, jointly with Ministry of Economic Development, academia (MIPT, Skoltech, Medical Scientific Center and Faculty of Mechanics and Mathematics of Moscow State University, Innopolis University, Lobachevsky University, Psychology Institute of RAS, Joint Supercomputer Center of RAS) and industry (Kaspersky Lab, EC-leasing, InterProCom, Technoprom).

INTELLECTUAL PROPERTY

ISP RAS business model suggests that IP rights are either retained by the institute or transferred to an open source developer community under special agreements (e.g. with Free Software Foundation). Taking into account the specifics of this model ISP RAS developed a unique license based on the direct financing by the customer of the research and development for the licensed technology (instead of paying royalties). The customer gets non-exclusive rights for using the technology, and the institute retains the exclusive IP rights. For some cases, decisions on managing IP rights are made individually based on long-term collaboration perspectives. An example of such an exception is the collaboration with Advanced Research Foundation, which assumes transferring all IP rights to the customer.

OPEN SOURCE

One of the most important components of the created ecosystem is the widely used open source software that is absolutely necessary for modern system programming. Open source is considered as:

- a tool that provides legitimate free access to all modern technologies, including ready-to-use software products and open standards;
- an ability to ensure the institute innovative research without outsourcing contracts but interacting with global market of products and services;
- a powerful educational resource as the environment and infrastructure of international open source projects can be used to train engineers.

Scientific activity implies the result's openness and the visibility of its author, which often contradicts IT corporate policies. For ISP RAS, the openness of research results is both a work motivation and a tool for promoting the institute's technologies. Open research means that each young researcher is visible in the international IT community. Their contribution and reputation are their capital, and the institute does everything to ensure that this capital grows as quickly as possible.

EDUCATION

The ISP RAS innovation ecosystem cornerstone is educational activity, which is carried out in several directions:

- Cooperating with leading universities. Institute specialists are working on system programming chairs of MSU, MIPT and HSE. Starting from their first year students attend to system programming lectures and corresponding practical lessons. At the third year students enter the chairs for system programming and continue to attend to lectures, start to work on special seminars, get acquainted with the institute's scientific directions, participate in projects and receive a special scholarship. By the time of graduation, many students have scientific publications and become system programming experts. ISP RAS researchers are constantly updating education courses and bachelor programs. For example, in 2021 ISP RAS started academic advising to modernize the software engineering bachelor program of the Faculty of Computer Sciences in Higher School of Economics. In 2020, ISP RAS and Moscow Aviation Institute have launched a master program in the area of supercomputer modeling.

- Scholarship program. ISP RAS launched a special scholarship program for students and postgraduates of MSU, MIPT, HSE, Novgorod State University, Russian-Armenian University, etc.
- ISP RAS postgraduate study helps gaining practical experience and learning new technologies at the same time. Postgraduates are actively involved in education: they instruct seminars and practical classes for students, supervise term papers and theses. After getting a degree they usually become leaders of small research groups.
- System programming labs network. Currently, ISP RAS external labs are working in Yerevan, Veliky Novgorod, Orel, Plekhanov Russian University of Economics. The laboratories attract successful students and involve them in the development of promising technologies in close cooperation with industry.

Starting from 2017, ISP RAS are actively working with Samsung in the Samsung IoT Academy. In 2021, ISP RAS and Huawei in collaboration with MIPT has launched a two-year free educational program devoted to system programming that can be taken by students of any MIPT departments and years of education.

CONFERENCES

ISP RAS organizes a number of annual events:

1 International ISP RAS Open Conference: https://www.isprasopen.ru/en

2 OS DAY, a conference on science and practice (jointly with other organizers): https://www.osday.ru/

3 International Ivannikov Memorial Workshop: https://www.ivannikov-ws.org/en

4 International Conference “Data Science In Medicine” (jointly with other organizers): https://digital-med.ru/en

5 SYRCoSE Software Engineering Colloquium: http://syrcose.ispras.ru/

6 The “System Programming as a Key Direction for Counteracting Cyberthreats” roundtable (International Military-Technical Forum “Army”).

Ivannikov Institute for System Programming of the Russian Academy of Sciences

25 A. Solzhenitsyn st. Moscow 109004 Russia
scsec@ispras.ru

Ivannikov Institute for System Programming of the RAS 2021