HiPEAC info 72

JUNE 2024

ACACES 2024 Fiuggi

EU policy directions, from artificial intelligence to cyber resilience

Adventures in cybersecurity, from hardware to systems

Lessons in launching a cybersecurity assessment business

contents

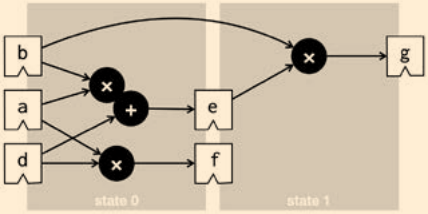
Recent EU policy developments

Cybersecurity, the NCP and the HiPEAC Vision

Digital guardians: HiPEAC vs. cyberattacks

31

How to launch a cybersecurity assessment company



33

Faster hardware with verified high-level synthesis

46

Career talk: Virginia Niculescu, Babeș-Bolyai University

Spanning the compute continuum from edge to cloud, HiPEAC (High Performance, Edge And Cloud computing) is a network of over 2,000 world-class computing systems researchers, industry representatives and students. First established in 2004, the project is now in its seventh edition. HiPEAC7 focuses on networking and roadmapping activities: bringing the computing community together in Europe, exchanging ideas, building thriving European value chains and exploring the long-term vision for computing systems.

hipeac.net @hipeac / @hipeacjobs

hipeac.net/linkedin hipeac.net/tv

Funded by the European Union

The HiPEAC project has received funding from the European Union's Horizon Europe research and innovation funding programme under grant agreement number 101069836. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Cover image: Andrii Yalanskyi / Adobe Stock

Design: www.magelaan.be

Editor: Madeleine Gray

The theme of this magazine is cybersecurity. HiPEAC has covered cybersecurity in computing since the first HiPEAC Vision document in 2007, but it is fair to say that most computer architects are more concerned about performance and power consumption than about cybersecurity. In the last decade, there have been several cases in which performance-enhancing optimizations have been exploited to set up attacks, like Spectre and Meltdown.

Unfortunately, security is of growing importance, and we are currently in a perfect storm. No modern society or large company can function properly without digital technology to support its basic functions: communications, financial transactions, mobility, energy, …. Attacking a few servers is sometimes enough to disrupt an organization – in some cases for weeks.

Modern systems based on the IoT-edge-cloud computing continuum are even more vulnerable to attacks because they expose a larger attack surface. A server in a datacentre is easier to protect (both physically and logically) than a network of sensor nodes and edge computers in a power grid, for example. The recent trend of edge artificial intelligence (AI), combined with local control for actuation, can further increase the impact of an attack because it can also compromise safety.

Finally, there is cyber warfare. Cyberspace is now generally recognized as the fifth domain of warfare, after land, sea, air and space. All major countries have created cyber armies to defend critical infrastructure, and to build capability to attack other countries. Attacks range from disrupting the critical infrastructure of a country (power grid, public transportation, emergency services, water supply), to destabilizing society with fake news. The current tensions between the European Union (EU) and Russia, China and the Middle East have increased the intensity of such attacks. Elections and big international events like the Olympic games are another opportunity to interfere in the public debate.

All the above has been going on for years, but it has accelerated with the advent of powerful AI-based tools. The tools cannot only be used to set up or defend against attacks, but also to influence public opinion, and to create unrest.

In conclusion, I believe that our community has not only the responsibility but also the moral obligation to make our systems secure by design. Cybersecurity is too important to be an afterthought, and we should put more effort in it (both in hardware and in software). I hope that this magazine will give you some inspiration.

Koen De Bosschere, HiPEAC coordinator

From the AI Act to the Cyber Resilience Act: in this article, Jan Komarek, policy offer in the European Commission's Internet of Things Unit and HiPEAC project officer, sets out recent policy developments and their implications for the computing systems community.

Powering the digital and green transitions: Recent EU policy developments

Together with the strategic research agendas of European industrial associations, the HiPEAC Vision is a key strategic document for the European Commission when defining its research and innovation programmes. In parallel with technology advancements, in recent years the Vision has also referred to the geopolitical situation surrounding Europe, which has increased competition and tensions. We are witnessing technology races between the United States and China, and Europe has to find its position. This has been articulated in the 'races' outlined in the HiPEAC Vision 2024 – namely, for the 'next computing paradigm', for artificial intelligence, for new hardware technologies, for cybersecurity, and for sustainability, with the sovereignty of Europe a key issue in each area.

These races also resonate with the European strategy for achieving digital sovereignty and competitiveness, which aims to provide the framework for digital innovation in Europe. Below are some of the main policy developments in recent months.

In December, the European Parliament and the Council of the European Union (EU) reached a political agreement on the **Artificial Intelligence Act**, the first legislation of its kind in the world. The AI Act intends to prohibit biometric identification, untargeted scraping of facial images from the internet or CCTV footage, emotion recognition or social scoring. Artificial intelligence systems that exploit the vulnerabilities of people or manipulate human behaviour to circumvent their free will are to be curbed by this legislation as well. Deployers of high-risk AI systems will have to conduct a fundamental rights impact assessment prior to putting an AI system into use.

Of course, the regulation is set up so that we keep the innovation space open. European negotiators also agreed on a series of safeguards for the use of biometric identification systems in publicly accessible spaces for law enforcement purposes under strictly defined conditions. The AI Act comes at a time when AI applications are being rolled out in many different scenarios, resulting in even more requirements for computing systems. For example, the AI inference phase in autonomous driving requires special-purpose processors that are both inexpensive and provide results in real time.

The conviction that we are stronger when we join forces at EU level is at the heart of our vision for Europe's digital transformation to be achieved by 2030. This is expressed in the **Digital Decade** policy programme. It is driving our implementation strategy in digital through its ambitious 2030 targets, which are structured around four key pillars: investing in digital infrastructures, the digital transformation of business, the digital transformation of the public sector, and, of course, digital skills. In particular, it aims to double EU share

in global production of cutting-edge semiconductors and provide for 10,000 climate-neutral highly secure edge nodes.

With the European strategy for data adopted in 2020, we are making Europe a leader in the data economy by creating a genuine single market for data. With computing power moving closer to where the data is, data legislation has a strong influence on the rules across the computing continuum. The European Commission is putting forward the legislative framework for a prospering data economy. For example, the **Data Act**, which came into force in January 2024, will make more data available for use, in particular industrial data. It will set up rules on who can use and access what data for which purpose. Furthermore, with its cloud strategy, Europe is deploying secure, interoperable, and low-carbon cloud and edge services and infrastructures across Europe.

Hardware is not forgotten, either: the European **Chips Act**, which came into force on 2023, will help reinforce the European semiconductor ecosystem, increase the security and resilience of our supply chains, and strengthen our technological sovereignty. This will help achieve both the digital and green transition. The Chips Joint Undertaking is supporting these advancements and it has already launched its first call in November with €1.7 billion of EU funding. The Chips Act is one of several initiatives announced by the Commission over the last three years that address Europe's technical sovereignty, following the initiative on batteries, and being followed this year by the initiative on reducing dependencies on third countries for critical raw materials.

Europe's sovereignty has also been supported by the European **High Performance Computing Joint Undertaking**, which has procured nine supercomputers, located across Europe. At the time of writing, two of them – Italy's Leonardo and Spain's MareNostrum – are in the top 10 of the Top500 list of supercomputers worldwide. Meanwhile, efforts to extend the secure and hyper-connected supercomputing, quantum computing, service- and data-infrastructure ecosystem will continue.

The Leonardo supercomputer is among the best in the world. Credit: CINECA

Finally, the **Cyber Resilience Act** approved by the European Parliament in March 2024 is of particular relevance to this issue of *HiPEACinfo*. This aims to address the twin problems of inadequate levels of security in many products, and the inability of consumers and businesses to currently determine which products are cybersecure, or to set them up in a way that ensures their cybersecurity is protected. To do so, the Cyber Resilience Act lays down a uniform legal framework for essential cybersecurity requirements for bringing products with digital elements to the EU market, covering the the planning, design, development and maintenance of such products.

In addition to these legislative advances, the European Commission has also been allocating funding for computing systems research. To give a few examples, recent calls in the Horizon Europe Work Programme have sought project proposals on next-generation edge computing technologies and open-source for cloud and edge computing, while last year's funding programme included calls for research projects relating to intelligent data processing. With the DISCOVER-US project launched in January 2024, the Commission also aims to further research into swarm intelligence and distributed computing in collaboration with the National Science Foundation in the United States.

Given the HiPEAC Vision's role as a major reference for the formulation of research and innovation programmes, it is likely that this will apply to the context of the next work programme for Horizon Europe, for the year 2025 and beyond. We look forward to proposals from this community which will help Europe deliver the digital and green transitions with secure computing systems, responsible AI, a prospering and privacy-respecting data economy, and designed-in-Europe computing infrastructure.

Record attendance at HiPEAC 2024 in Munich

With almost 700 registered participants from 36 countries, the 2024 HiPEAC conference had the highest attendance of any event organized by HiPEAC in its 20-year history. Taking place in Munich on 17-19 January 2024, the conference was once again an excellent opportunity to catch up with colleagues and find out about the latest computing systems research.

'We are very proud to have hosted the largest HiPEAC conference to date, with close to 700 participants this year, as well as raising the highest amount of sponsorship so far,' said Diana Göhringer (TU Dresden), general chair of HiPEAC 2024. 'The programme was also huge, covering everything from edge through cloud to high-performance computing.'

The conference keynote talks covered three compelling topics. Reetuparna Das (University of Michigan) set out how the demands of genomics applications are challenging systems to new levels of performance. Lieven Eeckhout (Ghent University) made a cogent case for the importance of sustainability in computing systems research, while Mitsuhisa Sato (RIKEN) discussed the design of the mighty Fugaku supercomputer and offered a glimpse into post-exascale computing. Meanwhile, Jan Komarek of the European Commission also updated the community on the latest policy developments, as set out in the 'Policy corner' article in this issue of *HiPEACinfo* (pp. 4-5).

The paper track featured presentations of articles published in ACM TACO, thanks to HiPEAC's journal-first model, covering topics including runtime optimization, machine learning, and emerging devices. With 32 workshops and 11 tutorials, there were also a plethora of parallel events to choose from, exploring the latest advances in edge artificial intelligence (AI), the computing continuum, next-generation communications, neuromorphic and quantum computing, high-performance computing, and much more. Participants also had the opportunity to visit the nearby Leibniz Supercomputing Centre.

The 2024 edition was testament to the ever-increasing role played by industry in the HiPEAC conference: having once again broken sponsorship levels, the conference included a vibrant exhibition of technology companies, from spinoffs to multinationals, while the industry session was an ideal opportunity for networking. Careers development and student-industry interaction were also key features of the conference, with activities including the HiPEAC Student Challenge and an Inspiring Futures careers session featuring representatives from companies.

The conference wouldn't have been complete without its social event, which this year took place in the convivial setting of Munich's famous Hofbräuhaus. During the dinner, a ceremony was held to celebrate the winners of the 2023 HiPEAC Technology Transfer Awards, who were congratulated in front of the 500-strong audience.

It goes without saying that the success of this year's conference was in large part due to the generosity of the 31 sponsors who supported the event, and to the commitment of the organizing team. For 2024, the general chair was Diana Göhringer, the programme chair was Hayden So (University of Hong Kong), and the local chair was Stefan Wallentowitz (Hochschule München University of Applied Sciences). Longstanding committee members include Sascha Uhrig (workshops chair), Daniel Gracia (industry chair), and Dezun Dong, Diego R. Llanos and Gaspar Mora Porta (all publicity chairs).

hipeac.net/2024/munich

HiPEAC24 Google Photos album
bit.ly/HiPEAC24_photos

HiPEAC24 on HiPEAC TV – including full keynote talks
bit.ly/HiPEAC24_videos

Have your say on the HiPEAC Vision 2025

The HiPEAC Vision editorial board is currently preparing the 2025 HiPEAC Vision, setting out the computing challenges that private and public investments should address to get Europe ahead in the technology race.

The HiPEAC Vision 2025 will be centred on the 'next computing paradigm' (NCP). More details on this can be found in the following paragraphs and in the HiPEAC Vision 2024 (see 'Further reading', below).

The NCP envisions the convergence of the web, cyber-physical systems (CPS), the cloud, the internet of things (IoT) and digital twins into a coherent, federated, dynamic and intelligent continuum of computing.

The resources needed to execute service requests in the NCP will be dynamically orchestrated by edge-based generative artificial intelligence (AI) in a manner that balances availability, privacy, provenance, latency, energy, decentralization, personalization and context awareness, delivering users from the all-absorbing pull of the cloud.

The NCP will be a 4D computing paradigm that elevates the 2D plane of the current web into a time-sensitive spatial (spatial hence 3D, plus time for 4D) web-based space that intertwines the physical and the digital worlds in a unified networked infrastructure where services and data move fluidly, opportunely, seamlessly and interoperably.

The HiPEAC Vision editorial board is seeking feedback from the community on this vision. Have you heard of the NCP – and if so, was it from HiPEAC or from elsewhere? Do you agree with this vision of the evolution of computing? Do you agree that it is attractive for next-generation applications, and / or strategically important for business and the economy? What are the main technical challenges involved in achieving this vision?

You can have your say by responding to this brief survey:
hipeac.typeform.com/to/bSd4xm

FURTHER READING:

HiPEAC Vision 2024 hipeac.net/vision/2024.pdf

T. Vardanega and M. Duranton. 'HiPEAC's Vision for the Future: Envisioning the next computing paradigm' on the Communications of the ACM blog, 1 May 2024
cacm.acm.org/blogcacm/hipeacs-vision-for-the-future

Vardanega, T., & Duranton, M. (2024). 'The Next Computing Paradigm: an Introduction' in the HiPEAC Vision 2024, Rationale
doi.org/10.5281/zenodo.10874349

Key cybersecurity concerns for European MetaOS projects

On 10 April 2024, the 'Meta Operating Systems (MetaOS) Workshop', organized by Jan Komarek and Rolf Riemenschneider (European Commission), took place in Brussels. The workshop brought together project representatives, experts and stakeholders to discuss ongoing work in the EUCloudEdgeIoT initiative and set out future directions for the computing continuum.

Following presentations covering the work of the six task forces under the EUCloudEdgeIoT umbrella in the morning, the afternoon session explored cross-cutting topics including platform integration, standardization, and cybersecurity.

The cybersecurity session discussed the need for cybersecurity measures at all layers from internet-of-things (IoT) devices up to central computing and storage facilities, with Jan Komarek identifying key issues for cybersecurity in MetaOS projects, particularly on the pros and cons of decentralized networks from the cybersecurity perspective.

While storing data in a single centralized location makes data protection easier and reduces the attack surface, it also presents a tempting target for attacks, as breaching the protection potentially grants access to all data. Therefore, even centralized systems often share data among multiple silos to reduce the impact of any individual breach.

The MetaOS use cases are driven by moving from centralized security management to distributed security management. The topologies add another security layer by minimizing the transmission of sensitive data by processing at the edge and by data aggregation before transmission to the cloud. Security measures at the network edge, including firewalls, access controls, and encryption, are essential to protect data as it is transmitted to the cloud. Securing communication channels between edge devices and the cloud, for example by encryption, is critical to prevent interception.

Authentication and access control mechanisms must be enforced at the edge and on IoT devices to prevent unauthorized access to sensitive data and resources. Device identity management and role-based access control are commonly used to enforce security policies. Decisions on where to place security controls — locally on devices or centralized in the cloud — can significantly impact the system's ability to respond to and recover from attacks. Cybersecurity measures are inherent to all MetaOS networks, notably those with application in critical infrastructures as well as in healthcare sector.

What is challenging for decentralized networks is to provide for resilience ensuring that critical components continue to function with uninterrupted operation even during security attacks. Validation of the resilience of application's cybersecurity measures and response procedures is carried out by simulated real-world cyberattacks, attack scenarios such as unauthorized access attempts and data breaches. Redundancy, replication and synchronization mechanisms are necessary to ensure data resilience of distributed computing.

DISCOVER-US webinar videos available online

DISCOVER US

DISCOVER-US is a coordination and support action that promotes networking between researchers in the European Union (EU) and United States (US), focusing on distributed computing and swarm computing. Activities include vision building, funded transatlantic research exchanges, and networking events.

The project also holds regular webinars, and several are now available to view on the HiPEAC TV YouTube channel, in a dedicated DISCOVER-US playlist.

- **DISCOVER-US EU-US Virtual Networking Event**, featuring six EU-funded projects and four projects funded by the National Science Foundation (NSF)
- **Meet DISCOVER-US**: an introduction to the DISCOVER-US project, its main aims and activities
- Rute Sofia on the **CODECO edge-cloud orchestration framework**
- Talks on different aspects of the **SmartEdge project**
- **Sage: A Software-Defined Sensor Network**, presented by Pete Beckman (formerly Argonnne National Laboratory, now Northwestern University)

bit.ly/DISCOVER-US_webinar_videos

EuroHPC hosting entities join forces in new project to support HPC users

Paula C. Rodrigues and Rita Moutinho, INESC TEC

EPICURE
Unlocking European-level HPC Support

The EuroHPC Joint Undertaking has been working towards taking European supercomputing to new heights by installing new machines capable of competing with their global counterparts. Now, the current and future hosts of these supercomputers have combined forces to provide support services to selected users. The result is the EPICURE project (High-level specialised application support service in High-Performance Computing), featuring 16 partners from 14 different countries, and led by CSC – IT Center for Science, in Finland.

'At the moment, most of the help provided to users is short-term support which requires basic high-performance computing (HPC) skills and competences,' says Emmanuel Ory, development manager at CSC and coordinator of the project. 'EPICURE will help users at an advanced technical level. Right now, all the different sites in Europe use different architectures, so users need to be able to move from one system to another very easily and seamlessly,' continues Ory.

The project will provide several services, including code enablement and scaling, performance analysis and benchmarking, and code optimization. To this end, the project will draw on the experience and knowledge of all partners in HPC operations and support, using training activities and hackathons to share knowledge.

EPICURE's consortium is composed by 16 organizations from 14 different countries: Barcelona Supercomputing Centre (Spain), Cineca Consorzio Unteruniversitario (Italy), IT4Innovations at VSB – Technical University of Ostrava (Czechia), Institut Informacijskih Znanosti (Slovenia), Institut Jozef Stefan (Slovenia), Luxprovide SA (Luxembourg), INESC TEC (Portugal), Sofia Tech Park JSC (Bulgaria), Universiteit Antwerpen (Belgium), Kungliga Tekniska Hoegskolan (Sweden), Akademia Gorniczo-Hutnicza IM. Stanislawa Staszica W Krakowie (Poland), Danmarks Tekniske Universitet (Denmark), Forschungszentrum Julich Gmbh (Germany), Grand Equipement National de Calcul Intensif (France) and Centre Informatique National de L'Enseignement Superieur (France).

For further information, contact: paula.c.rodrigues@inesctec.pt
and rita.moutinho@inesctec.pt

Co-funded by the European Union

EuroHPC Joint Undertaking

EPICURE has received funding from the European High-Performance Computing Joint Undertaking under grant agreement no. 101139786.

EIC Pathfinder and Acclerator Challenges 2024

European Innovation Council

The 2024 European Innovation Council (EIC) Work Programme includes over €1.2 billion of funding for proposals focusing on breakthrough technologies and game-changing innovations which are high risk and with high potential for impact and to scale up internationally.

EIC Pathfinder challenges for this year include 'Nanoelectronics for energy-efficient smart edge devices'. Meanwhile, EIC Accelerator challenges a number of potential interest to the HiPEAC community, including 'Human Centric Generative AI made in Europe', 'Enabling virtual worlds and augmented interaction in high-impact applications to support the realisation of Industry 5.0' and 'Enabling the smart edge & quantum technology components'.

For further information, including the application process and deadlines, visit the EIC website.
eic.ec.europa.eu/eic-funding-opportunities_en

"Over €1.2 billion of funding for proposals focusing on breakthrough technologies which are high risk and with high potential for impact and to scale up internationally"

AMBRA uses AI to help people with communication disabilities

Christian Pilato (Politecnico di Milano), Ambra Di Paola (Fondazione Artos), Serena Muraro (Fondazione Artos), and Roberto Marinelli (Fondazione Artos)

Augmentative and alternative ccommunication (AAC) techniques leverage images and symbols to support individuals with communication disabilities. These methods help consolidate the autonomy levels of the individuals; for example, teaching them how to behave in many situations. However, individuals can understand symbols differently, meaning that personalized approaches are necessary.

A collaboration between Politecnico di Milano and Fondazione Artos, the AMBRA project is addressing these challenges by leveraging a combination of artificial intelligence (AI)-based methods to create a personalized learning environment in AAC. For example, it uses large language models to create texts or images, convolutional neural networks to automate the conversion of photos into symbols, and collaborative filtering to create personalized schedules. The AMBRA platform is cloud based and intended to be modular and easy to extend or cooperate with other AAC initiatives. For example, it leverages and extends the ARASAAC symbol library. Educators can share the learning material worldwide and further customize it to adapt to specific needs, cultures, or languages. The project is currently self-funded, leveraging support from research partners and industry to access computational resources.

Our approach aims to lower the technological, social, and cultural barriers to using AAC methods. The frontend of the AMBRA platform can run on all modern devices (such as smartphones, tablets, laptops, etc.) so that both educators and users can use it in every context, including real life.

Educators may exchange material more easily, creating a virtuous cycle that could boost the adoption of AAC methods. Having low-cost methods to create materials can encourage shops, public places, and community places to adopt such solutions and allow for a seamless transition between school and society, creating a more inclusive environment around the individuals. For example, QR codes allow users to download specific material (like a menu or a shopping list), which is later personalized based on specific user's information and symbols. This allows users to apply the autonomy concepts that have been acquired in class. AMBRA can also lower the cultural and language barriers when used with immigrants (for example using symbols from different cultures), facilitating their integration into a new community.

In conclusion, this project is intended to become a beacon to extend the use of AAC in real life. However, AMBRA's idea extends beyond its technical architecture; it embodies a vision for a more inclusive future where the boundaries of communication are pushed ever further thanks to the virtuous circle created by the combination of all these solutions, enabling society to include individuals with diverse communication needs who can express themselves with clarity and dignity in any context and without barriers. This approach can also lead to a rethink of the educator's role to be more open to worldwide collaboration in material and knowledge sharing.

Please feel free to contact us directly:
info@progetto-ambra.it

To stay updated on project news and connect with the AMBRA team, follow the LinkedIn page:
linkedin.com/company/progetto-ambra

"AMBRA [helps enable] society to include individuals with diverse communication needs who can express themselves with clarity and dignity in any context"

IT4Innovations is a new member of the iRODS Consortium

Marketa Dobiasova, IT4Innovations

IT4Innovations National Supercomputing Center (Czech Republic) has become a member of the iRODS Consortium, a group that brings together businesses, research organizations, universities and government agencies from around the world to ensure the sustainability of the iRODS software as a solution for distributed storage, transfer, and management of data.

Integrated Rule-Oriented Data System (iRODS) is open-source software that allows the storage, management and sharing of large amounts of data, including their metadata, between different organizations and platforms. It also provides a mechanism for defining rules for data storage, processing and distribution. iRODS is designed to support collaboration, interoperability and scalability of data infrastructures.

Martin Golasowski, senior researcher at IT4Innovations, summarized the benefits of membership in the iRODS Consortium: 'The demand for a comprehensive solution for fast and efficient data transfer between locations is increasing across the European scientific community. Membership of the iRODS Consortium enables us to communicate directly with the development team and provides us with access to the latest features and support in providing tools to the scientific community and beyond.'

iRODS provides a virtual file system for various types of data storage, metadata management, and a mechanism for federating geographically distant locations for data transfer. 'These features are used in the LEXIS Platform, which simplifies the use of supercomputers to run complex computational tasks through a unified graphical interface or using a specialized application interface. The transfer of large volumes of data between supercomputers and data storage is performed automatically and transparently for those using iRODS and other data management technologies,' adds Golasowski.

The iRODS software has been deployed at thousands of locations for long-term management of petabyte data in various industries such as the oil and gas industry, biosciences, physical sciences, archives, the media and the entertainment industry.

FURTHER INFORMATION:
irods.org
docs.lexis.tech

New MAR chapter of Women in High Performance Computing

This year, a new chapter of Women in High Performance Computing (WHPC) has been opened. Named MAR-WHPC, the chapter has members from all over Spain and plans to expand to include members throughout southern Europe. The MAR chapter, which was presented by Elisabeth Ortega (HPC Now!) and Sara Royuela (Barcelona Supercomputing Center) at the HiPEAC conference in Munich, is dedicated to fostering an inclusive environment and supporting women in all stages of their high-performance computing (HPC) careers.

Read more about MAR-WHPC in our interview with Sara Royuela on pp. 48-49.

linkedin.com/company/mar-whpc
@mar_whpc

"MAR-HPC is dedicated to fostering an inclusive environment and supporting women at all stages of their careers"

Coverage-guided fuzzing with virtual ECUs

Martin Ring, Bosch Research

The increasing complexity of automotive software – and its subsequent impact on safety and security – means that standards have to be followed for safe and secure software development. Fuzz testing is a cybersecurity measure that generates targeted test cases to identify security vulnerabilities in software. It is recommended by ISO/SAE 21434, as well as other standards and best-practice documents, to improve not only cybersecurity but software quality in general.

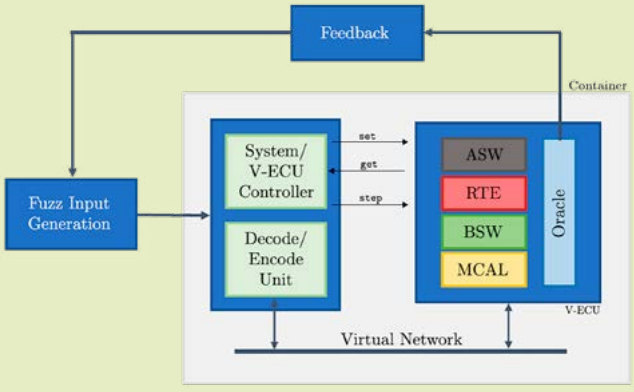
To enable fuzz testing in a stateful and embedded environment such as automotive, fuzzing based on a virtual electronic control unit (V-ECU) has been explored by Bosch. This work aims at creating a paradigm shift from a black-box fuzzing approach to a grey-box source-code fuzzing approach utilizing coverage feedback.

We use the FMI (Functional Mock-up Interface) standard to build the vECU with source-code instrumentation for coverage feedback and oracle implementation. The FMI standard defines four levels of abstraction, as follows:

- single software components
- communicating software components
- fully integrated ECUs with a changed MCAL (microcontroller abstraction layer)
- target software executed with an emulator

The fuzzer generates meaningful coverage-guided mutation data, which is communicated to the V-ECU through the FMI. In addition to fuzzing the ECU via protocol interfaces such as CAN (controller area network) and Flexray, the FMI interface facilitates the fuzzing of internal variables exposed by the FMI, thus enabling grey-boy fuzzing.

We focused mainly on the level three defined by the FMI standard, with a codebase up to >98% the same as the target software due to the increased speed (100x to 0.5x wall clock time (WCT)), precision and easier scalability. Using V-ECU based fuzz testing in continuous security testing (CST) increases the robustness of software developed by identifying security vulnerabilities and implementation bugs at a very early stage. The advantages shown by V-ECU fuzzing in terms of easy integration into the continuous integration / continuous deployment (CI/CD) pipeline and reduced costs are a convincing argument for the adoptiong of V-ECU fuzzing as a standard testing approach.



EmLogic VHDL verification course material now available

Espen Tallaksen, EmLogic

Used by almost 40% of field-programmable gate array (FPGA) designers in Europe and supported by the European Space Agency (ESA), UVVM (Universal VHDL Verification Methodology) has been the fastest-growing FPGA verification methodology for the last four years, according to Wilson Research. UVVM provides a test-bench infrastructure, randomization, functional coverage, specification coverage, verification components, scoreboards and more. This allows UVVM users to make structured, readable and reusable testbenches, thus improving quality and speeding up verification.

Embedded systems specialists EmLogic have been running courses on UVVM across the globe for many years. They are now making much of their training material available for universities to use in their education of FPGA designers.

The following modules will be made available to universities, with others potentially being released in due course:

- Basic infrastructure with logging, alert handling, verbosity control, checkers, waiters, etc.
- Bus functional models (BFM): Introduction and advanced use
- Simple, enhanced and optimized randomization
- Functional coverage

You can request access to the UVVM course material by emailing Espen Tallaksen

et@emlogic.no

emlogic.no/uvvm

EmLogic

Effective testing of cyber-physical systems

Christopher Huth, Bosch Research

Cyber-physical systems, such as control elements in power plants, traffic control systems or vehicles, connect the physical world with the digital world. This interlinking of systems with important infrastructure requires special protection.

Nowadays, safety-critical elements in software development are checked using automated test procedures, as non-automated approaches are too slow and prone to errors with many control elements and complex overall systems. Fuzz testing, in which possible program paths are tested using random inputs and decisions, has become an established technique in this area, particularly for complex software applications (see also 'Coverage-guided fuzzing with virtual ECUs' on p. 12). Compared to alternative test procedures, fuzz testing offers greater efficiency as the complexity of the application increases. The method is therefore also suitable for cyber-physical systems, although it must be adapted to the technical requirements of the systems.

The 'Effective testing of cyber-physical systems' (CPSec) project is researching how fuzz testing can be adapted for embedded cyber-physical systems. Embedded systems are a challenge for researchers as they do not use standardized operating systems and have few input options and interfaces, meaning that they have different requirements for analysis than those of conventional computer systems.

Cyber-physical systems are often characterized by nested program paths that are logically connected to external sensors and actuators. Systematic and in-depth coverage is therefore important to increase quality, speed, and cost efficiency. To this end, static and symbolic analysis methods need to be combined with fuzz testing. The strength of static methods lies in the structured analysis of source code. Symbolic methods simulate the execution of programs and take various program states into account.

The results of CPSec will be used to develop new tools for test generation on all information technology (IT) system layers. These tools will be made available as open-source software and their techniques are documented and evaluated in scientific publications.

The work is funded by the German Federal Ministry of Education and Research (BMBF, project CPSec).

FURTHER INFORMATION:
cpsec.io

CPSEC

Max Eisele receives ACM Frank Anger Memorial Award 2024

Max Eisele, a research engineer at Bosch, has been honoured with the ACM SIGSOFT Frank Anger Award 2024 for his pioneering work on 'Debugger-driven Embedded Fuzzing'. Presented annually by the Association for Computing Machinery (ACM), this student travel award promotes cross-disciplinary research between embedded systems and software engineering.

Debugger-driven software analysis is a methodology supposed to work on any system or programming language. The key ingredient is utilizing only generic debugging features of widely available debug probes to systematically analyse the system under test during runtime.

Embedded systems, such as those found in medical devices, automotive, and industrial control systems, are increasingly interconnected and vulnerable to cyber-attacks. Eisele's work aims to enhance the security of these systems by identifying potential vulnerabilities before they can be exploited.

Based on the observation that the vast majority of microcontrollers, the brains of any embedded system, provide debug capabilities, debugger-driven fuzzing is applicable at scale. For instance, Eisele showed that the limited number of hardware breakpoints on microcontrollers can be used to obtain partial code coverage feedback that can be used to guide fuzzing. Another approach uses single-stepping and watchpoints to analyse the system under test as it processes input data to learn the input specification, which can then be utilized for model-based fuzzing.

By systematically analysing and identifying potential vulnerabilities in embedded systems, Eisele's research offers a proactive approach to enhancing their security and reliability.

On behalf of HiPEAC, congratulations!

AIOTI releases paper on edge IoT industrial immmersive technologies

The Alliance for IoT and Edge Computing Innovation (AIOTI) Working Group on Research and Innovation has published a paper titled 'Edge IoT Industrial Immersive Technologies and Spatial Computing Continuum, Release 1'. Available for download from the AIOTI website, the paper provides a vision of the convergence of edge internet of things (IoT), artificial intelligence (AI), digital twins, intelligent mesh connectivity, the IoT of senses and spatial computing technologies to create the industrial real-digital-virtual continuum.

In their announcement of the publication of the paper, AIOTI states that the convergence of these technologies into industrial immersive solutions advances the integration and application of edge intelligent immersive technologies combining augmented reality (AR), virtual reality (VR), mixed reality (MR), and extended reality (XR) with concepts like metaverses, omniverse, multiverses, the next-generation spatial web, and web 4.0 as part of future virtual worlds.

AIOTI's position is that this convergence can dramatically transform industrial manufacturing, production operations, training, maintenance, design, collaboration, data visualization, mobility and logistics, improving efficiency, reducing downtime, enhancing safety, and ensuring better decision-making in industrial settings.

However, achieving this convergence results in challenges, including the robustness of hardware, data security, and effective industrial workflow integration.

FURTHER INFORMATION:
bit.ly/AIOTI_WG_Edge_IoT_immersive_spatial_1

Research on contemporary open-source honeypots

Niclas Ilg, Bosch Research

Threat intelligence is a vital part of any cyber-security strategy. Web applications, medical devices, automotive connectivity units, or (industrial) internet-of-things (IoT) devices are an entrance into personal and corporate networks. In addition to the rapidly growing number of connected devices and services, machine learning — e.g. in the form of large language models — is expected to shorten the time needed to discover security bugs and zero-day exploits. Whether based on proprietary or open-source software, these applications must be equally secure.

One building block of a successful defence-in-depth strategy is the honeypot, a decoy resource that mimics a valuable target system to entice adversaries. The attacker's strategy and movements on the honeypot system are logged and used to improve existing security measures on genuine devices. In enterprise information technology (IT), honeypots are already an established tool to get early notice of new exploits and zero-day attacks. In addition, the (industrial) IoT domain has also adopted the use of honeypots since botnet variants started attacking its devices on a large scale.

We investigated a large number of honeypots to learn how different solutions seek to deceive potential attackers. The results were published in a comprehensive survey (see 'Further reading', below) and highlight how many different approaches to honeypots there are.

Our research aims to utilize honeypots effectively in the automotive domain. While previous solutions have failed due to the characteristics of the domain (including strict budgets, highly embedded systems with low resources, very limited update capability, and low attacker exposure) we see upcoming electrical / electronic (E/E) architectures as a new chance for successful honeypot deployments. Thus, we consider how honeypots can be an additional layer of intrusion detection inside the vehicle, and how honeypots can be used to learn about novel attack vectors against automotive systems.

FURTHER READING:
doi.org/10.1016/j.jnca.2023.103737

Programming Heterogeneous Hardware via Managed Runtime Systems



Authors: Juan Fumero, Athanasios Stratikopoulos and Christos Kotselidis

This book aims to act as an entry point for developers using managed programming languages such as Java and Python to harness accelerators such as graphics processing units (GPUs) for better performance and energy efficiency. It provides an introduction to both heterogeneous execution and managed runtime environments (MREs) by discussing the current trends in computing and the evolution of both hardware and software. To this end, it first details how heterogeneous hardware differs from traditional central processing units (CPUs), what heterogeneous hardware components are and what challenges they pose to heterogenous execution. Subsequently, current solutions for programming heterogeneous MREs are described, highlighting for each existing solution the associated advantages and disadvantages.

This book is written for scientists and advanced developers who want to understand how choices at the application programming interface (API) level can affect performance and / or programmability of heterogeneous hardware accelerators, how to improve the underlying runtime systems in order to seamlessly integrate diverse hardware resources, and / or how to exploit acceleration techniques from their preferred programming languages.

bit.ly/Springer_Managing_Het_HW

TLS Cryptography In-Depth

Explore the intricacies of modern cryptography and the inner workings of TLS

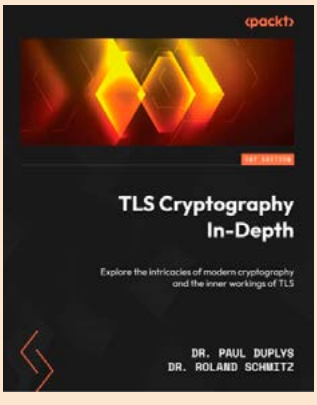

Authors: Paul Duplys and Roland Schmitz

Transport Layer Security (TLS) is the most widely used cryptographic protocol today, enabling e-commerce, online banking, and secure online communication. *TLS Cryptography In-Depth* by Paul Duplys and Roland Schmitz helps you gain a deep understanding of how and why TLS works, how past attacks on TLS were possible, and how vulnerabilities that enabled them were addressed in the latest TLS version 1.3.

Starting with the basic concepts, you'll be led step by step through the world of modern cryptography, guided by the TLS protocol. As you advance, you'll learn about the necessary mathematical concepts from scratch, with even seemingly arcane topics such as public-key cryptography based on elliptic curves explained in terms of real-world applications in TLS. You'll also find out how secret keys are generated and exchanged in TLS, and how they are used to create a secure channel between a client and a server. By the end of this book, you'll have understood the inner workings of TLS and how to configure TLS servers securely, as well as gaining deep knowledge of the cryptographic primitives that make up TLS.

The book was written with two audiences in mind. On one hand, it is written for IT professionals, cybersecurity professionals, security engineers, cryptographers, software developers, and administrators looking to gain a solid understanding of TLS specifics and their relationship with cryptography. On the other hand, it can also be used by computer science and computer engineering students to learn about key cryptographic concepts in a clear, yet rigorous way.

FURTHER READING:
bit.ly/Packt_TLS_Cryptography_In-Depth



Dimitris Gizopoulos and collaborators awarded IEEE Transactions on Computers Best Paper Award 2022

HiPEAC member Dimitris Gizopoulos and collaborators have won the best paper award for 2022 from IEEE Transactions in Computing, one of the most prestigious archival journals in computing, for joint work by researchers at the University of Athens, the Universidade Federal do Rio Grande do Sul (UFRGS) and the University of Trento.

The paper, entitled 'Soft Error Effects on Arm Microprocessors: Early Estimations versus Chip Measurements', shows for the first time how early (pre-silicon) estimates (predictions) of central processing unit (CPU) error rates (through detailed microarchitectural injection) in performance simulators (the state-of-the-art in the field: gem5) are correlated to physical error rates obtained from expensive post-silicon accelerated neutron beaming experiments on actual silicon.

Diligently tuned microarchitecture-level fault injection is a powerful approach that saves engineering resources, manufacturing costs, and delivers 'just-right' error protection for CPUs and system-on-chips (SoCs) by identifying the most vulnerable structures of the hardware and the most vulnerable pieces of the software. In the era of large-scale computing in the cloud, edge, and internet of things (IoT), with reports of silent data corruptions (SDCs) compromising datacentre operation, microarchitecture-level reliability assessment of CPUs and SoCs early in the design stages is a valuable piece of the puzzle of implementing robust computing systems.

The paper's journey began with an initial version presented at the IEEE/IFIP International Conference on Dependable Systems and Networks 2019, which was named the best paper runner-up at this top dependability conference. An expanded version of the paper, with two CPU models and more experimental setups, was then published in IEEE Transactions on Computers in October 2022 and subsequently named best paper for that year; it was that edition's 'Featured Article' and was spotlighted in IEEE Computer journal in its July 2023 issue.

The paper is the result of joint research between Dimitris Gizopoulos' research team at the University of Athens (Computer Architecture Lab, Department of Informatics and Telecommunications), which builds microarchitecture-level reliability assessment methods and tools for heterogeneous SoCs with CPUs, graphics processing units (GPUs), and domain-specific accelerators, and collaborators at the UFRGS and the University of Trento who specialize in beaming. Congratulations to the entire team: Athanasios Chatzidimitriou, George Papadimitriou, Dimitris Gizopoulos (all University of Athens), Paolo Rech (University of Trento), Pablo Rafael Bodmann and Rubens L. Rech Jr (UFGRS).

FURTHER READING:

P. R. Bodmann, G. Papadimitriou, R. L. R. Junior, D. Gizopoulos and P. Rech. 'Soft Error Effects on Arm Microprocessors: Early Estimations versus Chip Measurements' in IEEE Transactions on Computers, vol. 71, no. 10, pp. 2358-2369, 2022
ieeexplore.ieee.org/document/9616430

A. Chatzidimitriou, P. Bodmann, G. Papadimitriou, D. Gizopoulos and P. Rech. 'Demystifying Soft Error Assessment Strategies on ARM CPUs: Microarchitectural Fault Injection vs. Neutron Beam Experiments', 2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Portland, OR, USA, 2019, pp. 26-38
ieeexplore.ieee.org/document/8809532

Lieven Eeckhout wins Computer Architecture Letters Best Paper Award 2022

In related news, HiPEAC member Lieven Eeckhout was awarded the IEEE Computer Architecture Letters Best Paper Award 2022 for his paper 'A First-Order Model to Assess Computer Architecture Sustainability'. Lieven, who gave a keynote talk at the 2024 HiPEAC conference in Munich, has been pioneering the study of sustainability in computer architecture and is an active contributor to the HiPEAC Vision.

FURTHER READING:

L. Eeckhout, 'A First-Order Model to Assess Computer Architecture Sustainability' in IEEE Computer Architecture Letters, vol. 21, no. 02, pp. 137-140, 2022
bit.ly/L_Eeckhout_sustainability_model

ETH Zürich Digital Circuits and Systems Group win tinyML Research Symposium Best Paper Award

The Digital Circuits and Systems Group at ETH Zürich, led by HiPEAC member Luca Benini, received the best paper award at the tinyML Research Symposium 2024. The paper, titled 'Boosting Keyword Spotting through On-Device Learnable User Speech Characteristics', was co-authored by Cristian Cioflan and Lukas Cavigelli.

The paper shows that AI-based speech models can be augmented with simple user features, offering better recognition rates for different accents or voice pitches. This method is suitable for tinyML devices, as it needs only 0.7 mJ of energy to improve the accuracy results by 78% of what training the entire model achieves.

Congratulations on behalf of HiPEAC!

In memoriam: Detlef Olschewski

In January, we were saddened to hear that HiPEAC member Detlef Olschewski, the chief executive of Cleopa GmbH, had passed away unexpectedly at the age of 54. Detlef was described by his colleagues as someone who brought life, energy, and ingenious ideas to every meeting; who made work feel like home, and who spread humour and positivity everywhere he went.

Our thoughts go out to Detlef's family and friends, who sadly miss him.

Dates for your diary

HiPEAC webinars
Check the HiPEAC website to keep up to date on forthcoming dates
hipeac.net/webinars

HiPEAC 2025: High Performance, Edge And Cloud computing
20-22 January 2025, Barcelona, Spain
Workshop proposal deadline: 1 July 2024
hipeac.net/2025/barcelona
Sponsorship opportunities available
sponsorship@hipeac.net

EPS Summer School 2024
18-23 August 2024, Bertinoro, Italy
chipsacademy.eu/summer-school

Euro-Par 2024: 30th International European Conference on Parallel and Distributed Computing
26-30 August 2024, Madrid, Spain
2024.euro-par.org

CPS Summer School
16-20 September 2024, Alghero, Sardinia, Italy
cpsschool.eu

ITEM Workshop 2024: IoT, Edge, and Mobile for Embedded Machine Learning
Co-located with ECML-PKDD 2024
9-13 September 2024, Vilnius, Lithuania
item-workshop.org

2024 IEEE Nordic Circuits and Systems Conference (NorCAS)
29-30 October 2024, Lund, Sweden
Paper submission deadline 15 August 2024
events.tuni.fi/norcas2024
Further information: Jari Nurmi, Tampere University
norcas@tuni.fi

Since 2007, the HiPEAC Vision has considered the topic of cybersecurity, monitoring how increased connectivity and increased reliance on digital technology has resulted in greater vulnerability to cyberattacks. In this article, Olivier Zendra (Inria), a member of the HiPEAC Vision editorial board, sets out some of the main issues for current systems, and for the 'next computing paradigm' to come.

Threats on the horizon

Cybersecurity, the 'next computing paradigm' and the HiPEAC Vision

In today's world, where digital technologies are present in almost every area of activity of human life, it is hard to overstate the importance of cybersecurity. Cyberattacks on critical infrastructure, such as manufacturing equipment or power grids, or supply chains, many of which were developed before the most recent digital technologies were available, can have serious, even life-threatening, consequences in the real world. In December 2015, for example, a hack on Ukraine's power infrastructure left 230,000 people without power for several hours. Another striking example is that of healthcare systems, a favourite target for cyberattackers, with a spate of attacks in recent years leading to operational disruption and misery for patients.

One thing that characterizes many of these non high-tech sectors is the use of legacy technology. Manufacturing systems are purchased with the aim of running for decades. One of the main issues in the devastating 2017 Wannacry ransomware attacks on the UK's National Health Service, among others, was the use of legacy operating systems that are no longer supported by the manufacturer.

HiPEAC's vision for the future of computing, the 'next computing paradigm' (NCP), foresees an even greater role for digital technologies, with components dropping seamlessly in and out of the network, and computations taking place across the computing continuum. It envisages even further intertwining of the digital and physical worlds, with the web of humans operating in conjunction with the web of machines.

In practice, this means more connected devices and services, more variety in the kinds of devices and services connected and the kinds of data they share between them, and therefore a considerably increased attack surface. The NCP is more vulnerable to supply-chain attacks due to the multiplication of the components involved, and the large number of composable services creates numerous integration and communication points. The dynamic composition of services offers further attack opportunities.

Another defining characteristic of the NCP is its incorporation of artificial intelligence (AI), from intelligent orchestrators to context adaptation. AI is both a shield and sword in cybersecurity: it can help detect cyberattacks, but could also be used to massively scale up such attacks. Large language models (LLMs) post a whole new area for cybersecurity challenges, such as prompt poisoning, and there is terra incognita ahead. With LLMs available for anyone to access, and increasingly powerful resources being released at incredible speed, cybersecurity approaches will have to adapt to the new AI reality.

No computing system can be accepted without trust, and the NCP is no exception. Hence robust cybersecurity is not optional, which means that the significant challenges in establishing necessary levels of security must be addressed. Individual components need to be assessed on an ongoing basis, but so does the system as a whole. The scale of the challenge is, in fact, enormous, when you start to think of all the points of vulnerability in such a complex digital landscape. One such point would be the communication between devices, for example, bearing in mind how many of today's internet-of-things (IoT) devices are poorly secured.

Another is the software supply chain: contemporary software development encourages the reuse of third-party components, many of them open source, to allow developers to focus on their area of specialization and therefore promote innovation. However, security weaknesses in an extensively used library may impact all systems that use this library. This was demonstrated by the Log4j vulnerability, revealed in December 2021. Rogue developers are another issue; in a recent case, a developer gained the trust of the administrators of the xz compression library, eventually being granted commit access and maintainership – at which point he installed a carefully hidden backdoor. Even hardware is far from immune: malicious integrated circuit alterations can provide hackers with control, while attacks like side-channel analysis allow hackers to obtain data from the system.

Is there light at the end of the tunnel? The decentralization aspect of the NCP offers potential for isolating – and hence protecting – computations, and technologies such as distributed ledger technology (DLT) and the InterPlanetary File System (IPFS) can play a role in decentralization. As witnessed by this magazine's special feature on cybersecurity, and also by the contributions to the HiPEAC Vision, researchers are hard at work on creating defences across the computing stack, from hardware to applications. What is for sure, though, is that making systems more cyber resilient and protecting against attacks has to be a constant, sustained and permanent effort.

Some common types of cyberattack

- **Malware** (ransomware, spyware, trojans – including hardware trojans)
- **Denial of service (DoS) attacks**, where attackers overwhelm a network with traffic in order to degrade performance or shut down the service. These include battery denial of service attacks, which attempt to drain a device's battery.
- **'Man-in-the-middle' attacks**, where the attacker intercepts communications between two parties while making them believe they are communicating with each other.
- **Social engineering attacks** (phishing, etc.), where attackers impersonate people or services, in order to lure users to give them information, such as passwords, bank details, etc.

Credit: Arnout Fierens

FURTHER READING

HiPEAC Vision 2024
HiPEAC.net/vision

Zendra, O., & Coppens, B. (2024). 'The NCP cybersecurity challenges'. HiPEAC Vision 2024, Rationale
doi.org/10.5281/zenodo.10874985

Coppens, B., & Zendra, O. (2024). 'More data for the NCP implies more privacy risks'. HiPEAC Vision 2024, Rationale
doi.org/10.5281/zenodo.10875003

Rudametkin, W., & Zendra, O. (2024). 'The browser: the key to your privacy on the Web'. HiPEAC Vision 2024, Rationale
doi.org/10.5281/zenodo.10875019

Bertin, E. (2024). 'DLT and IPFS Technologies are Paving the Way for the Next Paradigm in Computing'. HiPEAC Vision 2024, Rationale
doi.org/10.5281/zenodo.10875036

Reimann, L. M., Sisejkovic, D., & Leupers, R. (2024). 'Integrity at Every Link: A Roadmap to Trustworthy Hardware Supply Chains'. HiPEAC Vision 2024, Rationale
doi.org/10.5281/zenodo.10875050

Kunz, W., & Stoffel, D. (2024). 'Microarchitectures as Root-of-Trust in Computing Systems – Research Needs in Formal Security Analysis'. HiPEAC Vision 2024, Rationale
doi.org/10.5281/zenodo.10875080

The expansion of edge computing is posing new challenges for cybersecurity: solutions need to offer robust security with low energy and low latency. To answer this challenge, the NEUROPULS project, funded by the European Union (EU), is taking a novel, neuromorphic approach, developing photonics-based hardware security primitives that draw on the properties of light for robust, yet lightweight, cybersecurity layers. HiPEAC caught up with NEUROPULS coordinator Fabio Pavanello (CNRS – Center for Radiofrequencies, Optics, and Microelectronics in the Alps – CROMA laboratory) to find out more.

Lighting the way to better security

The NEUROPULS photonics-based approach

What's wrong with the current security landscape?

The exponentially increasing number of edge devices is posing major challenges not only on the computing side, in terms of processing requirements, but also in terms of cybersecurity. The available surface for cyberattacks increases with the number of interconnected devices, as do the complexity and latency of authentication and encryption protocols. Current solutions that rely on, for example, storing cryptographic keys in non-volatile memory are not secure enough. As an example, hardware vulnerabilities can be leveraged to access specific memory sectors, and approaches based purely on electronic hardware security primitives present a series of weaknesses, such as being prone to machine-learning (ML) modelling and side-channel attacks, or reliability issues due to the ageing of key components, for example.

That doesn't sound good…

No, especially when you consider that applications in sectors such as banking, manufacturing, automotive and healthcare cannot function without robust cybersecurity.

Achieving robust cybersecurity isn't that simple, though, right?

Exactly. In addition to being robust to cyberthreats and reliable, security layers also need to be low power, lightweight, and low cost for large-volume applications. All these requirements set severe constraints on both the hardware and software layers of these devices, meaning that we need to develop unconventional cybersecurity approaches with superior robustness compared to current solutions.

So what's NEUROPULS doing about it?

Adopting a neuromorphic approach, which saves energy, avoiding the input / output bottleneck between memory and processing units, and which supports machine-learning (ML) algorithms, NEUROPULS targets the secure operation, communication, and integrity of edge devices. We're building security layers starting with hardware security primitives, using the unique technology developed in the project based on augmented silicon photonics platforms. This technology is readily available for our photonic accelerator, which is co-developed in the same photonic integrated circuit (PIC) as the security primitives. These primitives are the key ingredient of our approach and are based on photonic physical unclonable functions (PUFs), which have the potential for better performance compared to their electronic counterparts.

What gives photonics an advantage over electronics in this area?

The enhanced strength of our approach comes from the underlying physics of such complex photonic devices, which provide many degrees of freedom that can be exploited, making it more robust against ML or side-channel attacks. This is in stark contrast with electronic approaches, where essentially information propagates in a binary manner, thus leading to lower system complexity overall and thus potential vulnerabilities for various types of attacks.

In the software layer, we can then leverage key properties of photonic PUFs, such as high operation speed (Gbit/s) to enable lightweight security protocols. For example, a lightweight remote software attestation mechanism periodically ensures that the device has not been tampered with, protecting against unauthorized software, particularly malware, malicious software modifications and associated risks, by using the photonic PUF as the root of trust.

Any other cybersecurity protections?

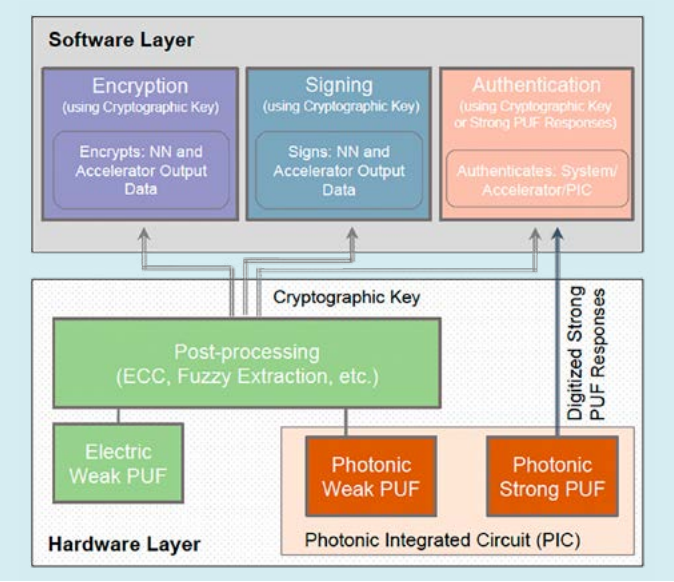
We also use encryption to protect sensitive neural-network information such as weight values and data in input / output, as well as to secure communication with third parties using the error-corrected keys generated by the photonic PUF. Finally, various lightweight authentication protocols are under investigation in the NEUROPULS project to assess the genuine identity of a given device within the network without requiring fully stable PUF responses.

What are the particular challenges of ensuring robust security for edge computing? I assume energy is going to be at a premium, for example…

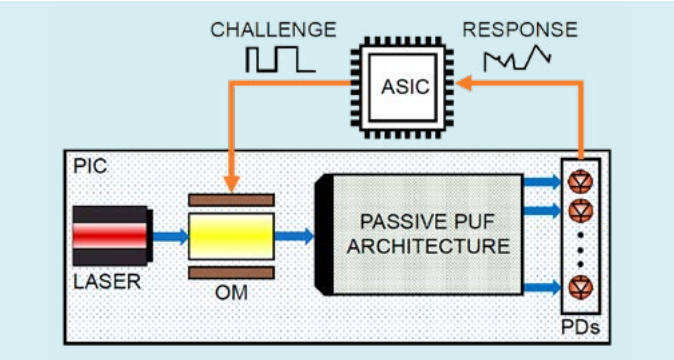
Precisely: in edge computing, the main goal is to securely process substantial amounts of data close to where these data are generated. This means that you need lightweight accelerators that can deal with massive amounts of data using little energy and with low latency, while also being protected by robust security layers. To address this problem, NEUROPULS takes advantage of the properties of light and of materials such as phase-change materials (PCMs), as well as III-V materials, to develop an ultra-low-power (sub-pJ/MAC) accelerator which at the same time presents highly robust security layers.

Sounds groundbreaking – but how would it work with existing systems?

We are co-developing processor interfaces for seamless integration in current systems. We're using RISC-V core architectures to drive our accelerator, which allow us to develop a prototype without licensing constraints. In parallel, a gem5-based simulation platform is under development to enable accelerator scaling and performance profiling, while incorporating the novel security layers.



Hardware-software communication flow for security services (Credit: 'Security layers and related services within the Horizon Europe NEUROPULS project', DATE 2024)



PUF operation considered in NEUROPULS (Credit: 'Security layers and related services within the Horizon Europe NEUROPULS project', DATE 2024)

I'm sold. What are the next steps towards commercialization, and how could this technology change the current cybersecurity landscape?

We aim to bring this technology, which is still at the proof-of-concept phase, to a much more mature technology readiness level. This would make it a realistic, scalable alternative to current insecure, power-hungry solutions for edge computing applications.

The advent of novel security primitives such as PUFs based on photonics rather than more conventional CMOS technologies, coupled with protocols which exploit their unique properties, could affect the overall supply chain and penetrate market segments that have so far been dominated by electronic solutions. This would allow us to build edge devices and sub-systems that are far more resilient to cyberthreats and fluctuations. For example, safety-critical applications such as autonomous driving could greatly benefit from security layers that outperform current solutions and that are integrated into low-power and low-latency photonic accelerators.

neuropuls.eu

NEUROPULS has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 10170238. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

FURTHER READING:

F. Pavanello et al., 'Security layers and related services within the Horizon Europe NEUROPULS project', 2024 Design, Automation & Test in Europe Conference, 2024 - arxiv.org/abs/2312.09383

F. Pavanello et al., 'NEUROPULS: NEUROmorphic energy-efficient secure accelerators based on Phase change materials aUgmented siLicon photonicS', 2023 IEEE European Test Symposium (ETS), 2023 - doi.org/10.1109/ETS56758.2023.10173974

N. Marastoni and M. Ceccato, 'Remote Attestation of IoT Devices using Physically Unclonable Functions: Recent Advancements and Open Research Challenges', CPSIoTSec '23: Proceedings of the 5th Workshop on CPS&IoT Security and Privacy, 2023 doi.org/10.1145/3605758.3623502

C. Odysseas et al., 'Gem5-MARVEL: Microarchitecture-Level Resilience Analysis of Heterogeneous SoC Architectures', 2024 IEEE International Symposium on High-Performance Computer Architecture (HPCA), 2024 doi.org/10.1109/HPCA57654.2024.00047

Digital guardians

How HiPEAC takes on cyberattacks

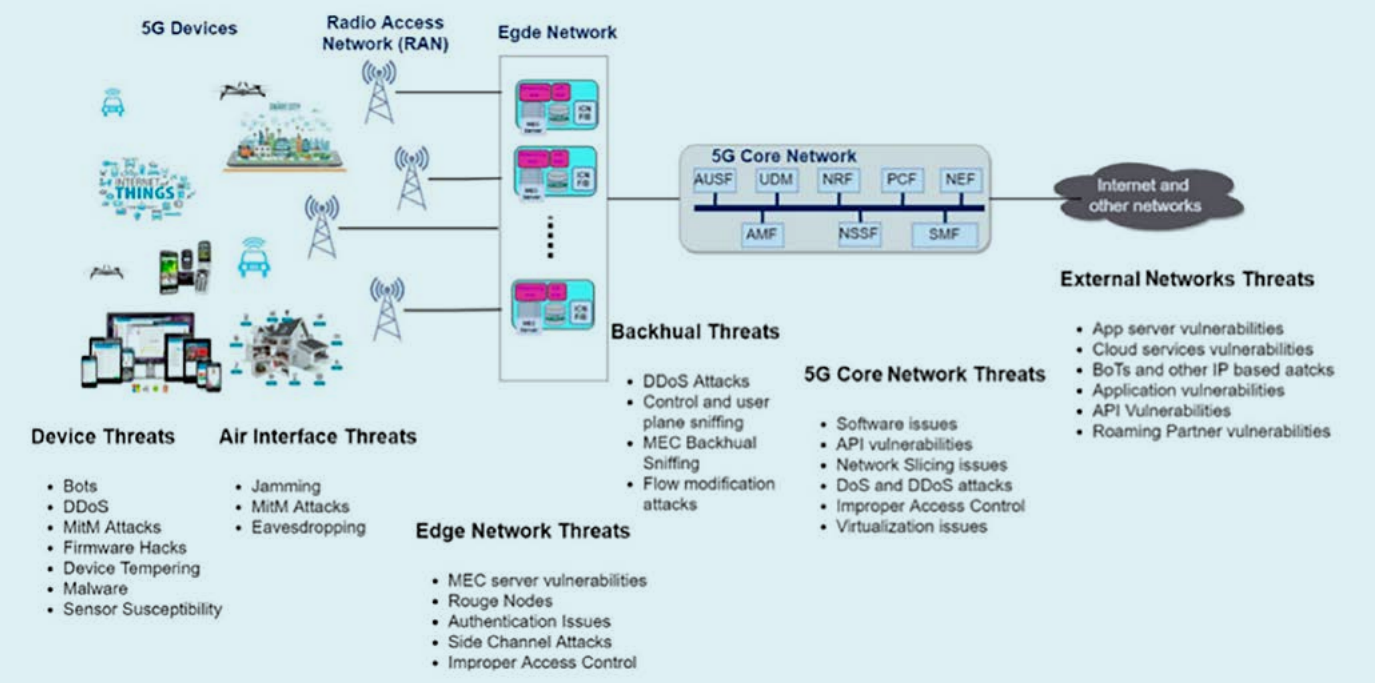
As society becomes increasingly reliant on digital technology, cybersecurity is becoming more important than ever. In this article, we learn about solutions from the HiPEAC community, from tackling 6G security to using machine learning in the fight against cyberattacks, and from decentralized zero-trust architectures to hardware security modules.

CONFRONTING TELECOMMUNICATIONS THREATS HEAD ON: SAFE-6G'S TRUSTWORTHY NETWORKS

The shift from 4G to 5G and now to 6G in mobile networks has led to an exponential increase in the number of connected devices and data transferred. This has resulted in a significant increase in security risks for the telecommunications industry, which showed a 51% increase in attacks in 2021 and is now the third most vulnerable sector. Successful attacks on the telecommunications network can potentially expose information from millions of customers, highlighting the need for increased security measures.

Insider threats and phishing are major risks for the telecom industry. The sector also deals with multiple third-party entities, and weak cybersecurity maturity levels of these vendors can give hackers a backdoor entry into the network, while the adoption of internet-of-things (IoT) devices has further increased the threat surface. Cloud computing, which has its own security challenges, is being increasingly adopted to support telecom operations, while protocols used by telecom companies, including Signalling System 7 (SS7) and Diameter, are also vulnerable to cyberattacks.

'Distributed denial-of-service' (DDoS) attacks – reported by ENISA to be more than 70% of telecom security incidents in 2021 – aim to disrupt services and cause significant financial losses. Given that DDoS attacks happen mostly on authentication procedures between the devices, radio access network (RAN) and core levels of the network, most 6G-oriented research projects



focus on using artificial intelligence (AI)-based threat detection and mitigation components ranging from the lower layers of the RAN to the highest layer of the packet core.

Another approach is to enhance the multi-step authentication component of the packet core by higher encryption levels and by incorporating AI in the encryption. Defined by Ericsson as AI-native methodology, this approach states that all network components should be AI-based, coupled and in communication between layers through a cross-layer data infrastructure with a higher-level life-cycle model that will operate as a 'machine-learning-as-a-service' (MLaaS) platform for the AI models to continue learning with time. Within this approach, multiple AI-based models are included to identify, classify and mitigate threats.

This is the starting point for SAFE-6G, a European research project (HORIZON-JU-SNS-2023-STREAM-B-01-04) which pioneers a holistic research approach, situating a native trustworthiness framework atop the open and distributed user service node (USN) / network service node (NSN)-based 6G core. SAFE-6G leverages explainable artificial intelligence / machine learning (known as (X)AI/ML) techniques to coordinate user-centric safety, security, privacy, resilience, and reliability functions, strategically optimizing the level of trust (LoT) as a pivotal key value indicator (KVI) for 6G, while addressing the specific trust requirements and data governance policies specified by each user / tenant throughout the entire 6G lifecycle, encompassing onboarding, deployment, operation, and decomposition.

Recognizing that security measures come at a cost to usability and agility, the proposed trustworthiness framework aims for a balanced approach, emphasizing a security-by-design methodology and leveraging cognitive coordination technologies, including intent-based trustworthiness driven by AI and ML techniques. This comprehensive approach is critical for establishing a resilient and trustworthy 6G system, ensuring the responsible evolution of network architectures.

SAFE-6G objectives in a nutshell

- Establish **6G trustworthiness requirements**, considering safety, security, privacy, resilience, and reliability, to design a cognitive trustworthiness framework for 6G architectures across the edge-cloud continuum,
- Design, implement and evaluate a **cognitive coordination framework** of the distributed AI techniques that ensure the level of trust in each 6G system,
- Design and develop the **functions (safety, security, privacy, resilience, and reliability)** necessary for **trustworthiness in a 6G ecosystem**, considering the full lifecycle,
- Follow a **cloud-native** paradigm for the design and development of SAFE-6G framework components,
- Ensure **interoperability** through compatibility with meta operating systems currently under development, such as that in the aerOS Horizon Europe project,
- **Verify and validate the SAFE-6G framework** using **metaverse-based pilots**.

Funded by the European Union

SAFE-6G has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under grant agreement no. 101139031.

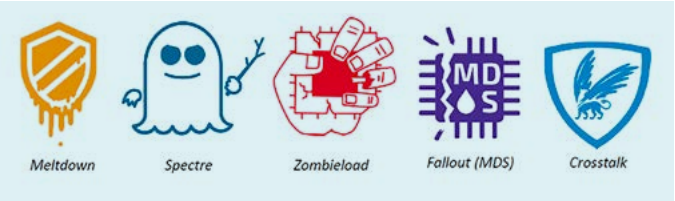
DYNAMIC MALWARE AND HARDWARE ATTACK DETECTION THROUGH MACHINE LEARNING

David Andreu, Marti Alonso, Beatriz Otero, Juan José Costa, Enric Morancho (all Universitat Politècnica de Catalunya-Barcelona Tech (UPC)), Stefano Di Carlo, Cristiano Pegoraro Chenet, Alessandro Savino (all Politecnico di Torino) and Ramon Canal (UPC)

In today's digital era, technology has transformed our lives, yet it also invites malicious actors to exploit vulnerabilities. Among the most prominent threats is malware, designed to infiltrate and harm computer systems. As technology dependence increases, so does the urgency for effective malware detection and prevention. Traditional signature-based methods, while somewhat effective, are struggling to keep pace with evolving malware. Cyber-criminals continuously are refining their tactics, producing variants that evade detection. Proactive measures are essential. Machine learning, a subset of artificial intelligence (AI), offers a potent defence. By training algorithms to recognize patterns and behaviours indicative of malware, cybersecurity experts gain a powerful tool in combatting these threats.

The VITAMIN-V project, funded under European agreement 101093062, aims to develop a robust infrastructure for a secure RISC-V cloud environment. At HiPEAC 2024, we presented ground-breaking work introducing two novel AI-based methods for malware detection. One method targets software-based

Members of the VITAMIN-V team at HiPEAC 2024



Examples of hardware cyberattacks

malware, while the other addresses hardware-based attacks such as Meltdown or Spectre. These methods prioritize low computational costs, high accuracy, and resilience to new threats, leveraging the pattern-recognition capabilities of machine learning.

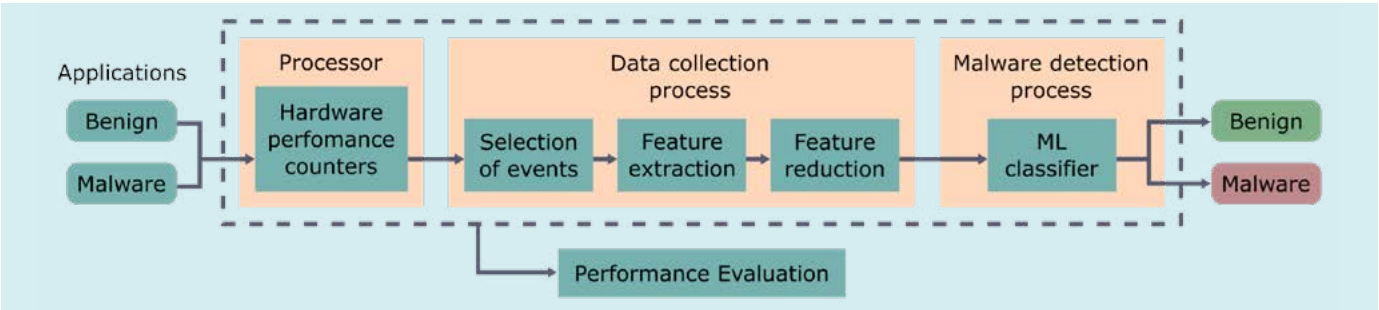
The first method, which focuses on software-based malware detection, utilizes dynamic opcode traces as its cornerstone. Opcodes represent instructions executed by a computer's central processing unit (CPU). Analysing the sequence of opcodes during program execution provides valuable insights into program behaviour, enabling researchers to distinguish normal activity from activity that is potentially malicious. Dynamic opcode traces offer distinct advantages over static analysis techniques. Unlike static analysis, which examines code in its static form, dynamic analysis considers actual program execution, leading to a more accurate understanding of behaviour. Moreover, by concentrating on opcodes rather than specific signatures, dynamic analysis is better equipped to detect previously unseen, zero-day malware.

For this detection method, a comprehensive malware dataset was compiled from the VirusShare database. Over 600 malware executables, along with an equivalent number of benign programs from Ubuntu's binaries folder, were processed to generate the dataset. Execution of both benign software and malware attacks was segmented into groups of 10,000 executed instructions, with each group representing a sample. These samples comprise the opcode histogram of each instruction group. The dataset was then utilized to train and test a Support Vector Classifier (SVC) detection model. Initial experiments yielded promising results, achieving a detection accuracy of 98% after fine-tuning the classifier's parameters.

While machine learning-based malware detection using dynamic opcode traces shows significant potential, it is crucial to consider the specific architecture on which these techniques are applied. While the x86 architecture, prevalent in personal computers, has been extensively studied and has produced impressive results, ongoing technological advancements mean that these techniques need to be validated on different architectures.

The second method, which focuses on hardware-based attack detection, utilizes hardware performance counters (HPCs) as its foundation to differentiate attacks from benign software. HPCs are specialized registers present in most modern computers that monitor various events occurring within the microprocessor. By understanding common hardware-based attack techniques, such as exploiting out-of-order and speculative execution, and placing significant stress on the microprocessor's cache memories, optimal HPCs can be monitored to identify the presence of malicious programs.

As with the opcode traces method, dynamic execution analysis offers significant advantages over static file analysis and can also prevent new hardware attacks by identifying behaviours akin to existing ones. For this detection method, the tracked HPCs include branch instructions, branch misses, L1 data-cache load misses, and last-level cache (LLC) load misses. The first two counters aim to detect attackers' efforts to manipulate the branch-prediction unit to exploit speculative execution, while the cache-related HPCs focus on the cache-stressing methods commonly used in hardware attacks, such as FLUSH+RELOAD.



VITAMIN-V's hardware and AI-based malware detection framework

Traces from various hardware attacks were collected, along with an equal number of samples from the same counters obtained by running benign programs, such as benchmarks or conventional software like bzip2 and ffmpeg. This data was used to train a support vector classifier (SVC) with a Gaussian kernel. Subsequently, the model's parameters were finetuned to achieve optimal accuracy in both detecting the presence of an attack and classifying it accurately. This approach achieved an accuracy of over 99% in both tasks.

As this method is based on hardware events rather than instructions, it is more straightforward to implement on new architectures, provided the necessary counters are available in the microprocessor. While the tests and results were conducted on an x86 architecture, current ongoing work is analysing RISC-V architectures.

In conclusion, the threat of malware continues to escalate as our digital landscape expands. Traditional signature-based detection methods and attack-specific mitigations are insufficient in combatting this evolving menace. Machine learning-based malware detection, utilizing dynamic opcode traces or HPC measures, offers an effective and enduring solution. By training models on labelled datasets and analysing program behaviour, we can accurately detect and prevent the infiltration of existing malware and prepare to counter future attacks.

FURTHER INFORMATION:

VITAMIN-V website vitamin-v.eu

One-minute video on dynamic malware detection
youtu.be/7QwR9ifyfmE

One-minute video on hardware attack detection
youtu.be/Oh8jxFbnlvc

VITAMIN-V has received funding from the Horizon Europe research and innovation programme under grant agreement no. 101093062.

DECENTRALIZED POLICY ENFORCEMENT IN ZERO-TRUST ARCHITECTURES

Lars Creutz and Guido Dartmann, Trier University of Applied Sciences

Zero-trust architectures (ZTA) have become increasingly popular in recent years as organizations seek to protect their sensitive data and systems from cyber threats. At the core of these architectures is the principle of least privilege, which grants only the minimum level of access required for an individual or system to perform a specific task. While this approach has proven effective in limiting the damage caused by security breaches, it also poses significant challenges when it comes to enforcing policies and regulations.

In this article, which is based on our paper 'Decentralized Policy Enforcement in Zero Trust Architectures' which we presented at the Symposium on Security in Future Networks of IEEE FNWF23 (see 'Further reading', overleaf), we explore the limitations of centralized policy enforcement in ZTA and argue that a decentralized architecture can be helpful in addressing these challenges.

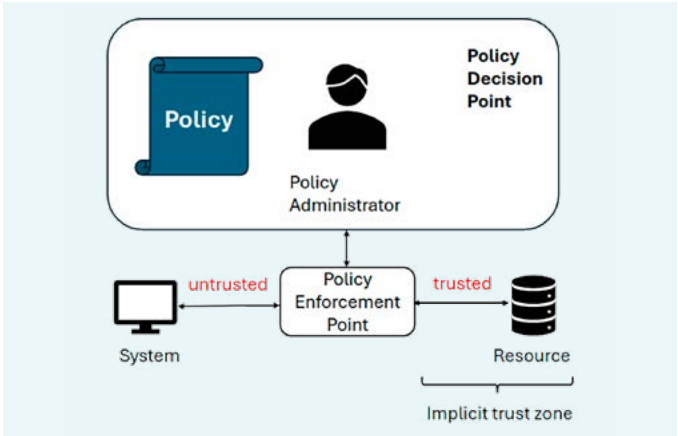


Figure 1: Centralized ZTA

The policy enforcement point (PEP) is a central point within a ZTA that enforces security policies across the network (see Figure 1). It acts as a gatekeeper, inspecting and controlling all traffic in the network. The PEP is responsible for authenticating users and entities, verifying their permissions, and enforcing security policies based on zero-trust principles (see Rose et al. in 'Further reading', overleaf).

Being a central point in the network, the PEP can be vulnerable to several attacks. For example:

- **A distributed denial of service (DDoS)** attack is a type of cyber-attack where an attacker floods a network or system with traffic to overwhelm it. In the context of the PEP, a DDoS attack could target the central point of enforcement, causing denial of service and potentially disrupting the entire network (see Teerakanok et al. in 'Further reading', overleaf).
- **Lateral movement within a network** refers to the ability of an attacker to move laterally across the network after gaining initial access. In the case of the centralized ZTA, lateral movement can occur from a compromised PEP or resource which results in the exploitation of the implicit trust zone.

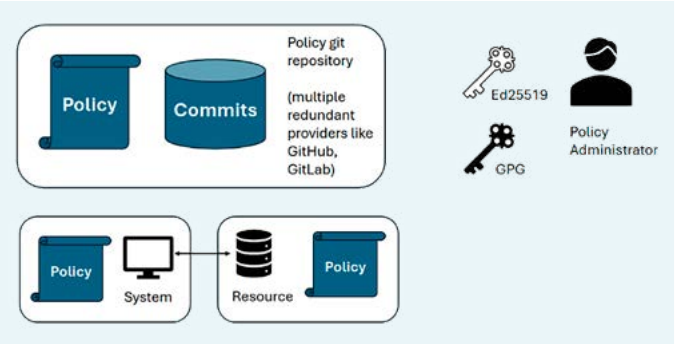


Figure 2: The authors' decentralized approach to a ZTA

Our approach differs by decentralizing policy enforcement and using a more distributed architecture. Security is improved by eliminating the central PEP and enforcing the policy in a decentralized way across each resource in the network, which not only reduces the described attack vectors but also motivates secure implementation of resources and services in true zero-trust fashion, while not hiding them inside an implicit trust zone.

We describe and distribute the policy inside a git repository using publicly available infrastructure like GitHub and GitLab to prevent vendor lock-ins and promote the usage of open-source software. Resources periodically check for updates from the policy administrator and perform a pull operation on their local repository. We use two signature methods to increase the security of our approach: first, the policy file (inside the repository) is signed using Ed25519; second, the git commit itself is signed using GPG. The separation of those keys allows close monitoring of the changes to the policy with two distinct parties responsible for the policy change.

To test our approach, we have simulated an edge-cloud scenario, which is currently being implemented in the form of a demonstrator on real hardware. Within the ZTA, ventilation events are transmitted to a central cloud service from edge-cloud devices. The entire network architecture corresponds to our previous description of a decentralized ZTA. We have also used cyber-physical contracts to implement authentication at the resources. Secure encrypted authentication tokens are transmitted in the contracts and control access to the system's resources without the use of a central PEP.

l.creutz@umwelt-campus.de
g.dartmann@umwelt-campus.de

FURTHER READING

Creutz, L. & Dartmann, G.: 'Decentralized Policy Enforcement in Zero Trust Architectures', 2023 IEEE Future Networks World Forum (FNWF), Baltimore, MD, USA (accepted for publication)

S. Rose, O. Borchert, S. Mitchell, and S. Connelly, 'Zero trust architecture', National Institute of Standards and Technology, 2020
doi.org/10.6028/NIST.SP.800-207

S. Teerakanok, T. Uehara, and A. Inomata, 'Migrating to zero trust architecture: Reviews and challenges', Security and Communication Networks, 2021
https://doi.org/10.1155/2021/9947347

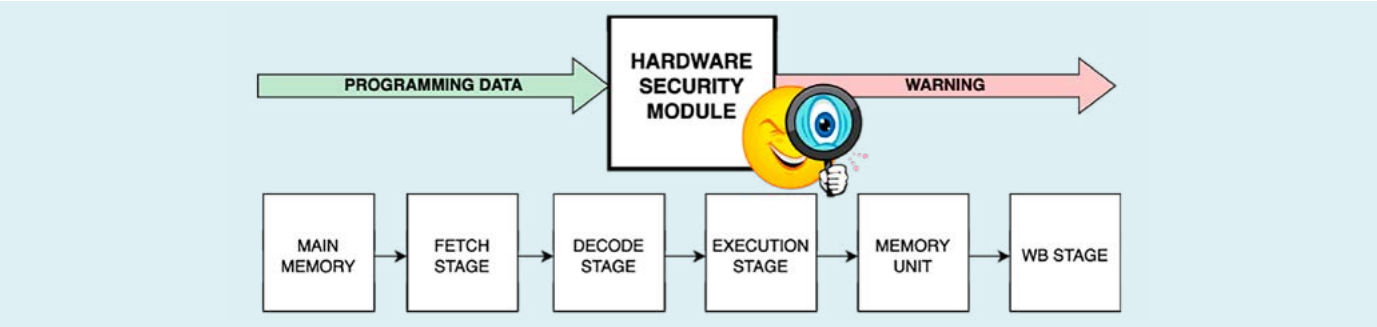
HARDWARE SECURITY MODULES DETECTING HARDWARE TROJAN HORSES AND SIDE-CHANNEL ATTACKS

Alessandro Palumbo, CentraleSupélec, Inria

Software-exploitable hardware trojan horses (HTHs) allow attacks to execute malicious software to modify running software or to acquire root privileges. In 2018, the Rosenbridge backdoor was found in a commercial VIA Technologies C3 processor, as documented by Christopher Doumas (see 'Further reading' below): a specific sequence of instructions allowed the attack to activate such a backdoor and enter supervisor mode.

As technology strives for enhanced performance, processors have evolved to include components such as multiple cores, caches, specialized acceleration units, and speculative execution, increasing both the systems' throughput and complexity. These additions often expose unexpected vulnerabilities that pose new challenges. For example, the timing differences introduced by caches or speculative execution may be exploited to leak information and private data or to detect activity patterns. As a result, a system deployed with trustable components may still be vulnerable despite having dedicated security modules.

Side-channel analysis exploits unintended information leakage by analysing timing information, power consumption, thermal footprint, or the electromagnetic emanation of computing systems while executing security primitives to extract information about the data processed and then using these to infer sensitive information. Attackers can use their own programs to access the memory (and secrets) of other programs and the operating system without having physical access to the victim's



The protection implementation approach

microprocessor. Well-known examples of microarchitectural side-channel attacks include Spectre and Meltdown, which have served as foundational models for numerous other ones.

Observing runtime behaviour with hardware security modules

It is extremely challenging for software to protect against vulnerabilities so close to the hardware. Hardware security modules (HSMs) operate at the circuit level and are therefore well suited to detecting and defending against low-level attacks.

Hardware trojan horses

In a 2021 paper by Palumbo et al. (see 'Further reading'), the authors set out how an HSM can be implemented in a microprocessor to deal against HTHs, looking at the behaviour of the protected system. During the safe program's installation, the HSM works in configure mode. Then, during program execution at runtime, the HSM switches to query mode. After each fetched instruction, the HSM checks whether the accessed address and the fetched instruction comply with the previously configured information. During the configuration phase, the HSM stores information about legitimate instructions and their addresses by hashing the instruction-address tuples. In the query phase, the HSM checks the tuples of the instruction fetched and the address required by the core to monitor if there are differences concerning the previous phase.

Side-channel attacks

To detect side-channel attacks, an HSM can look at microprocessor features (i.e. hardware performance counters, power consumption and temperature traces, timing, etc.). A 2022 paper by Ankan et al. presents a HSM based on count-min sketches for detecting microarchitectural side-channel attacks in microprocessor-based systems.

Suggestions for further reading, discussing the approaches and vulnerabilities more in detail, follow.

alessandro.palumbo@centralesupelec.fr
palessumbo.github.io
team.inria.fr/sushi

FURTHER READING

C.Doumas, 'Hardware Backdoors in x86 CPUs' https://bit.ly/44IUwQk

K. Arıkan et al. 'Processor security: Detecting microarchitectural attacks via count-min sketches'. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 30(7):938-951, 2022

A. Palumbo et al. 'Is your FPGA bitstream Hardware Trojan-free? Machine learning can provide an answer'. Journal of Systems Architecture, 128:102543, 2022

A. Palumbo et al. 'A lightweight security checking module to protect microprocessors against hardware trojan horses'. 2021 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)

S. Ribes et al. 'Machine Learning-Based Classification of Hardware Trojans in FPGAs Implementing RISC-V Cores'. 2024 International Conference on Information System Security and Privacy (ICISSP)

A. Palumbo et al. 'Improving the Detection of Hardware Trojan Horses in Microprocessors via Hamming Codes'. 2023 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)

A. Palumbo et al. 'Built-in Software Obfuscation for Protecting Microprocessors against Hardware Trojan Horses'. 2023 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)

P. R. Nikiema et al. 'Towards Dependable RISC-V Cores for Edge Computing Devices'. 2023 IEEE International Symposium on On-Line Testing and Robust System Design (IOLTS)

L. Cassano et al. 'Is RISC-V ready for Space? A Security Perspective'. 2022 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)

Medical implants hold significant therapeutic promise, but they also represent a significant security risk, with potentially devastating consequences. In this article, Muhammad Ali Siddiqi (Delft University of Technology / Lahore University of Management Sciences) and Christos Strydis (Erasmus Medical Center / Delft University of Technology) sketch out some of the main issues.

Brain gain or brain drain?

The risky rise of connected neural implants

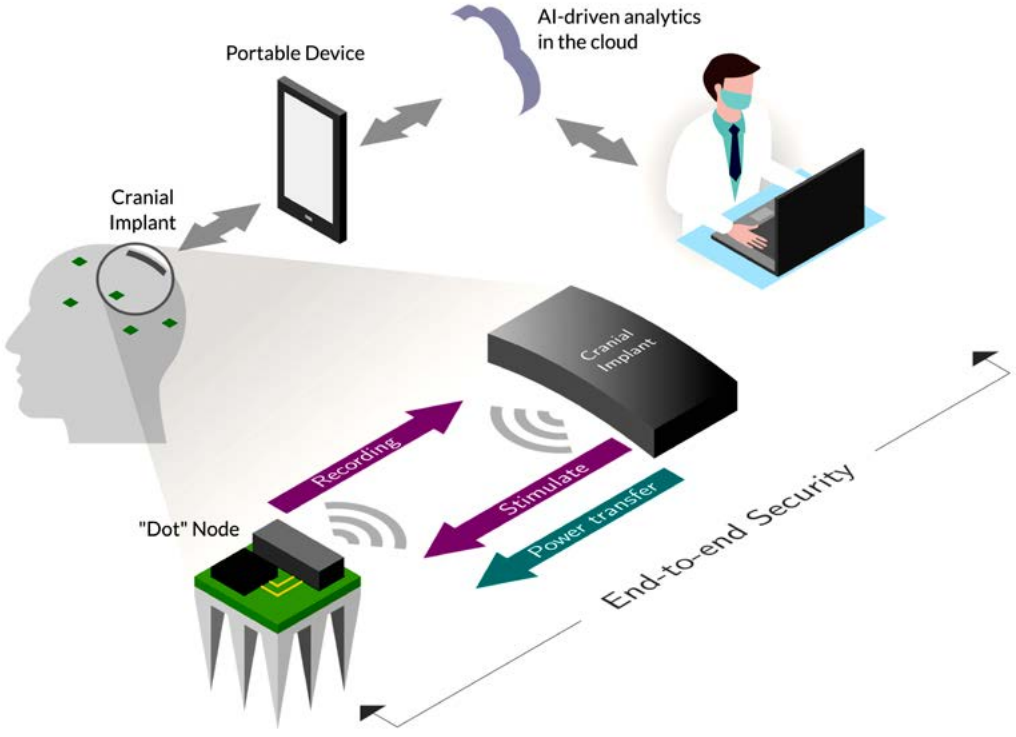
Christos Strydis (left), Muhammad Ali Siddiqi (right)

Implant technology has come a long way from its humble beginnings as resource-constrained devices focused on restoring aspects of health, such as heart pacing or Parkinson-tremor suppression, with limited connectivity to the outside world. Over the last two decades, we've witnessed a remarkable evolution in the design of implants, leading to a paradigm shift in their capabilities and potential.

First, modern implants now boast significantly enhanced computing power onboard. This advancement has enabled the tackling of more sophisticated medical applications, such as retinal implants, expanding the scope of conditions that can be effectively treated with implant technology.

Moreover, wireless connectivity has become a hallmark of contemporary implants, facilitated by ongoing miniaturization trends. This connectivity not only allows for seamless communication between implants and external devices but also opens up avenues for integrating in-body data with external sources, such as wearables and cloud-based analytics platforms. This integration aligns closely with the HiPEAC Vision's 'next computing paradigm' (NCP), fostering a cohesive continuum of computing that spans cyber-physical systems, the internet of things, and artificial intelligence; see figure below.

However, this increased connectivity also presents new challenges, particularly in terms of security and privacy, as we outlined in our paper 'IMDfence' (see relevant work in 'Further reading', overleaf). Recognizing the importance of safeguarding implant systems against cyber threats and data misuse has become paramount, especially in light of past attacks targeting medical devices.



An envisioned next-generation smart neuromodulation system highlighting the asymmetry of the devices involved along the communication path

> *"Implants are vulnerable to BDoS and 'man-in-the-middle' attacks"*

Due to their limited battery life and restricted charging windows (e.g. during sleep), implants are vulnerable to 'battery denial-of-service' (BDoS) attacks. In a BDoS attack, attackers overwhelm the implant with malicious communication requests, rapidly draining its battery even though they are successfully denied. Furthermore, implants are vulnerable to textbook 'man-in-the-middle' attacks since traditional security solutions like PKI (public-key infrastructure) are not suitable due to implants' resource constraints and often lack of direct internet connectivity. As a result, novel solutions leveraging low-level hardware are required to address these vulnerabilities effectively.

However, a key insight is that implants' unique location, which in some ways is a limitation, does offer a security advantage. Body-coupled communication (e.g. ultrasound-based SecureEcho – see 'Further reading', below) leverages the in-body environment for secure pairing, boosting implant security against the aforementioned attacks while conserving limited resources. By adopting such proactive approaches to security, manufacturers can bolster trust in implant technology and ensure the safety and wellbeing of patients relying on these devices.

The future of implants promises exciting advancements. One groundbreaking evolution is NeuroDots (see 'Further reading', below), a departure from single-site brain stimulation; see the 'dot'-node detail in the figure above. With NeuroDots, we envision next-generation implants targeting multiple connector hubs within brain networks. This approach could offer a more precise and localized treatment for various neurological disorders by addressing the root cause – the maladaptive interactions between these networks.

However, NeuroDots' capabilities come with an inherent security challenge. Their potential to alter brain function as well as their enhanced out-of-body connected functionalities represent a significant expansion to their attack surface. Malicious actors could exploit vulnerabilities, particularly those stemming from the asymmetrical distribution of resources among the devices involved, to manipulate individuals, potentially even altering their personality. This underlines the critical need for significantly enhanced implant security to manage the immense power of these devices.

In conclusion, connected implants present a double-edged sword. Their advanced functionalities hold tremendous promise for revolutionizing healthcare. However, robust security measures are essential for ensuring their safe and ethical use. By fostering innovation in both implant technology and cybersecurity (see related work in the Horizon Europe project SEPTON, available at the link below), we can unlock the transformative potential of connected implants and improve the quality of life for patients worldwide.

c.strydis@erasmusmc.nl
m.siddiqi@lums.edu.pk

FURTHER READING

Siddiqi, M. A., Doerr, C., & Strydis, C. (2020). 'IMDfence: Architecting a secure protocol for implantable medical devices'. IEEE Access, 8, 147948-147964
ieeexplore.ieee.org/document/9165063

Siddiqi, M. A., Beurskens, R. H., Kruizinga, P., De Zeeuw, C. I., & Strydis, C. (2021). 'Securing implantable medical devices using ultrasound waves'. IEEE Access, 9, 80170-80182
ieeexplore.ieee.org/document/9440455

De Ridder D, Siddiqi MA, Dauwels J, Serdijn WA, Strydis C. NeuroDots: From Single-Target to Brain-Network Modulation: Why and What Is Needed? Neuromodulation. 2024 Apr 16:S1094-7159(24)00051-5. doi: 10.1016/j.neurom.2024.01.003. Epub ahead of print. PMID: 38639704.

SEPTON project septon-project.eu

HiPEAC Vision 2024 hipeac.net/vision

> *"Malicious actors could exploit vulnerabilities to manipulate individuals, potentially even altering their personality"*


Ensuring robust cybersecurity means that system design has to take security into account from the outset. Here, Francesco Regazzoni (University of Amsterdam) and Paolo Palmieri (University College Cork) set out some of the main challenges.

Resilience from day 1

An overview of research working towards secure-by-design systems

Security by design is a critical paradigm in developing systems that are resilient to cyber threats from their inception. As research towards security by design advances, it faces several existing challenges as well as new developments. Here we highlight the main research directions and challenges in the area.

One primary challenge is privacy protection. With increasing data breaches and stricter regulations like the General Data Protection Regulation (GDPR), systems must incorporate robust privacy features from the outset. This involves not just encrypting data but ensuring minimal data collection and enabling user control over their information. Advanced techniques such as differential privacy and homomorphic encryption are being explored, which allow computations on encrypted data without revealing the underlying data.

Side-channel resistance is another critical area. Side-channel attacks exploit indirect information leakage, such as power consumption or electromagnetic emissions, to break cryptographic systems. Traditional security measures often overlook these vulnerabilities. Research focuses on developing hardware and software techniques to mitigate side-channel risks, such as masking and constant-time algorithms, which aim to obscure or eliminate the patterns that attackers could exploit.

Achieving crypto agility, which is the ability to switch cryptographic algorithms seamlessly in response to new vulnerabilities or advancements in cryptography, is also increasingly seen as crucial with respect to security by design. Current systems often lack such flexibility, making them vulnerable if an algorithm is compromised, and making the change to secure algorithms costly, complex and lengthy. Research is focused on developing frameworks and architectures that can support dynamic algorithm changes without significant performance hits or requiring extensive system redesigns.

The integration of security measures seamlessly into complex systems is also challenging. Security by design demands that security is not an afterthought but a core component of the system architecture. This requires interdisciplinary collaboration, as integrating security measures can affect system performance, usability, and cost. Researchers are working on creating standardized protocols and tools that facilitate easier integration of security features without compromising other system aspects.

Automation is a needed to manage the complexity of today's systems and to to create devices that are secure by construction. Nevertheless, existing design tool chains should continue to provide current functionalies and to maintain existing levels of reliability. As a result, design methodologies and tools capable of handling security should be integrated as much as possible into state-of-the-art design flows. Research on this topic spans a wide range of approaches, including domain-specific languages for security, tools for automatic application of countermeasure against specific attacks, and tools for verification and assertion of certain security requirements.

Overall, while security by design aims to create inherently secure systems, it must continually evolve to address dynamic and multifaceted challenges such as the ones presented above, ensuring that new developments keep pace with emerging threats and technological advancements.

Orbik

A spinoff from the Basque technology centre IKERLAN, ORBIK Cybersecurity was launched to provide tailored product cybersecurity assessment services to clients in the embedded systems field, with applications in safety-critical sectors such as manufacturing and energy management. HiPEAC caught up with ORBIK Chief Executive Salvador Truijillo, winner of a HiPEAC Technology Transfer Award, to find out more about this venture.

'Recognizing a market demand for specific product cybersecurity solutions, we seized the opportunity to establish a new player'

What drew you to the field of industrial product cybersecurity?

We have worked with original equipment manufacturers incorporating embedded electronics for some time. Recent years have witnessed an increase in connectivity, requiring enhanced cybersecurity measures. Future legislation is likely to mandate such protections.

What are some primary cybersecurity threats faced by companies and public authorities?

Increased attacks on industrial control systems across sectors like energy, transportation, health, and manufacturing pose significant risks to business continuity, societal resilience, reputation, and human safety. Compliance and regulatory pressures are also mounting.

Why is cybersecurity particularly critical for embedded systems in an industrial setting, such as manufacturing?

With more digital equipment being integrated into industry 4.0 for advanced functionalities related to data and digital services, the potential business impact of cybersecurity breaches is magnified. We are deploying novel services specifically for manufacturers on cybersecurity assessment and management of vulnerabilities specific for manufacturers. These services are based on ISA/IEC62443 and following the European Union's Cyber Resilience Act and other regulations that are coming and that will apply to manufacturers. We recently got accreditation to provide ISO17025-based IEC62443 testing services.

How did ORBIK Cybersecurity originate?

ORBIK Cybersecurity emerged as a spinoff from the IKERLAN research centre, leveraging 50 years of research and development (R+D) experience in embedded electronics. Recognizing a market demand for cybersecurity solutions in our region, we seized the opportunity to establish a new player in the field.

What are some of the primary pain points faced by your customers, and how does ORBIK address these challenges?

We collaborate with clients to assess the product cybersecurity status of their embedded systems, guiding them through compliance and certification processes. We also liaise with certification bodies and provide tailored operational services, including vulnerability management for embedded systems. We are launching a new service, myorbik.com, for product cybersecurity management including vulnerability management with specific features based on EU technology.

What were the main challenges in setting up the company?

The challenges we had to navigate included market identification, securing funding, establishing branding, assembling talent, product development, digital marketing, and business development. The journey of building a startup is multifaceted and demanding.

What key lessons have you learned during this journey, and what advice would you offer to aspiring entrepreneurs in this field?

While we're still in the early stages of our launch, the importance of planning and execution has been underscored. Persistence, diligence, and staying true to one's vision are crucial. To fellow entrepreneurs, I encourage you to pursue your ideas with determination. Plan meticulously and execute diligently, and success will follow.

orbik-cybersecurity.com

High-level synthesis (HLS) tools help automate the design of custom hardware accelerators, but they can be unreliable. To counter this, Yann Herklotz (now a post-doc at EPFL, formerly a PhD student at Imperial College London) and his then supervisor John Wickerson (Imperial College London) came up with Vericert, the first HLS tool that is suited to safety- and security-critical settings. In this article, they explain why Vericert was necessary and how it has been improved since its original inception.

Faster hardware with verified high-level synthesis

Offloading computation to custom hardware accelerators is a great way to improve performance these days, but designing these accelerators manually in Verilog is an arduous task. This has led to the development of HLS tools like AMD Vitis HLS, the Intel HLS compiler, LegUp, and Bambu HLS, all of which promise to automatically generate Verilog designs from a specification written in a high-level language like C.

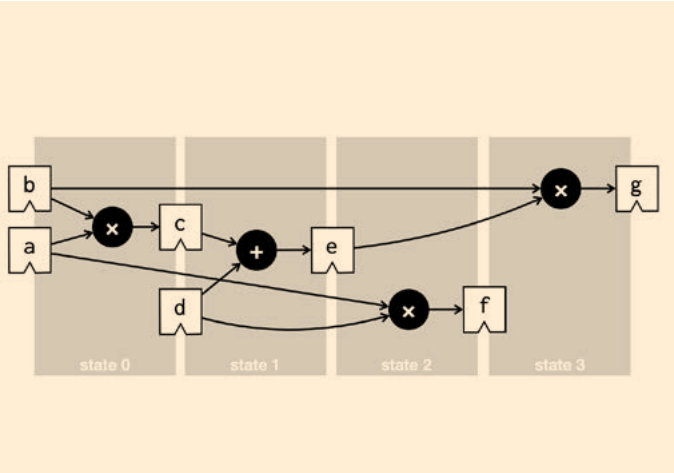
Unfortunately, the current crop of HLS tools has been shown to be a little unreliable – for instance, random testing uncovered miscompilations in all four of the above-mentioned tools. In the context of safety- or security-critical applications like cryptography, what this means is that although well-tested (and often formally verified) software implementations are available, and although custom hardware has advantages over general-purpose processors in terms of providing a well-isolated computing environment, current HLS tools cannot be relied upon to turn the former into the latter.

This is what motivated us to develop Vericert, a new HLS tool that is programmed in the Coq proof assistant, and comes with a computer-checked proof of its own correctness. That is, it is mathematically guaranteed only to produce a Verilog design that is behaviourally equivalent to the C program it is given.

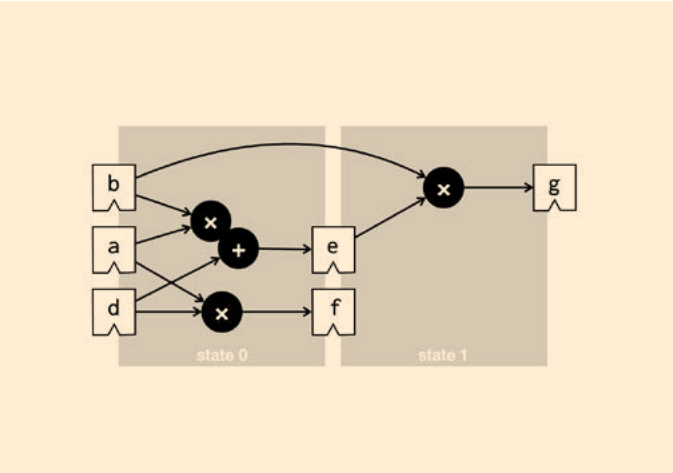
"The current crop of HLS tools has been shown to be a little unreliable ... Vericert comes with a computer-checked proof of its own correctness"

We published the first prototype of Vericert at OOPSLA in 2021. We acknowledged at the time that although Vericert had a strong 'correctness' story, its performance compared to existing, unverified HLS tools left much to be desired. Its main weakness was that it generated highly serial hardware – hardware that only performed a single operation per clock cycle.

Now, in work presented at PLDI 2024, we have addressed this weakness by adding a scheduling optimization. Vericert is now able to examine the list of operations that must be performed, and schedule them into clock cycles, reordering independent operations where necessary. It aims to pack as many operations into each clock cycle as possible, while keeping the combinational delay in each clock cycle short enough so that the hardware can run at a desirable clock frequency.



The original version of Vericert would allocate each operation to its own clock cycle



The new version of Vericert aims to pack several operations into each clock cycle

Thanks to this ability to schedule more work into each clock cycle, Vericert is now able to generate hardware designs that are about twice as fast as before. There are still more optimizations that need implementing before Vericert can truly compete with the unverified HLS tools, but our experiments indicate that the optimizations that Vericert does support seem to perform comparably to those in the unverified tools.

Many HLS tools have some sort of scheduling pass, so the challenging part of our work was not working out how to implement scheduling, but working out how to verify it. We used a 'translation validation' approach. This means that a verified checker takes a schedule produced by an unverified scheduler, and only allows the HLS process to complete if the schedule is deemed valid. Since the checker is proven only to validate correct schedules, this gives just as strong a correctness guarantee as verifying the scheduler itself, but means that the scheduler can be freely re-implemented without affecting the correctness proof.

Our scheduler works on hyperblocks, which are basic blocks whose instructions are all guarded by a predicate. Hyperblocks can represent groups of basic blocks connected by non-looping control-flow edges. An instruction only executes if its predicate is true. This means that our scheduler needs to take these predicates into account when reordering instructions. For instance, it must note when two instructions have mutually exclusive predicates and are hence independent. Therefore, our schedule-checker needs to be able to reason about Boolean formulas like 'p1 & p2 ==> False'. To do this, we use a verified SAT solver. In future work, we plan to investigate whether SAT and SMT solving can be brought into a verified compilation pipeline in an even more substantial way, as it may offer quite a general approach for building proven-correct optimizations.

In the nearer term, the next target for Vericert is to add support for loop pipelining, which aims to allow several iterations of a loop to overlap with each other. This is one of the most important HLS optimizations; once implemented, we hope that Vericert will really start to compete with the unverified HLS tools in terms of performance, not just correctness.

"Thanks to this ability to schedule more work into each clock cycle, Vericert is now able to generate hardware designs that are about twice as fast as before"

With the US Government calling for developers to avoid languages like C and C++, which don't offer inherent memory protection, there is greater focus on memory-safe languages. In this article, Codasip Chief Executive Ron Black sets out why memory safety and design provenance and traceability are key considerations in the company's cybersecurity offering.

Secure-by-design systems need memory safety and design provenance

On 24 April 2024, the UK Government's Science, Innovation and Technology Committee held an inquiry into the cyber resilience of the UK's critical national infrastructure. The enquiry featured evidence from, among others, HiPEAC member John Goodacre, the director of the Innovate UK digital security by design (DSbD) challenge, and Richard Grisenthwaite, executive vice president and chief architect at Arm. During this hearing, the witnesses discussed two extremely important cybersecurity issues: memory safety and the provenance of intellectual property (IP).

Memory safety

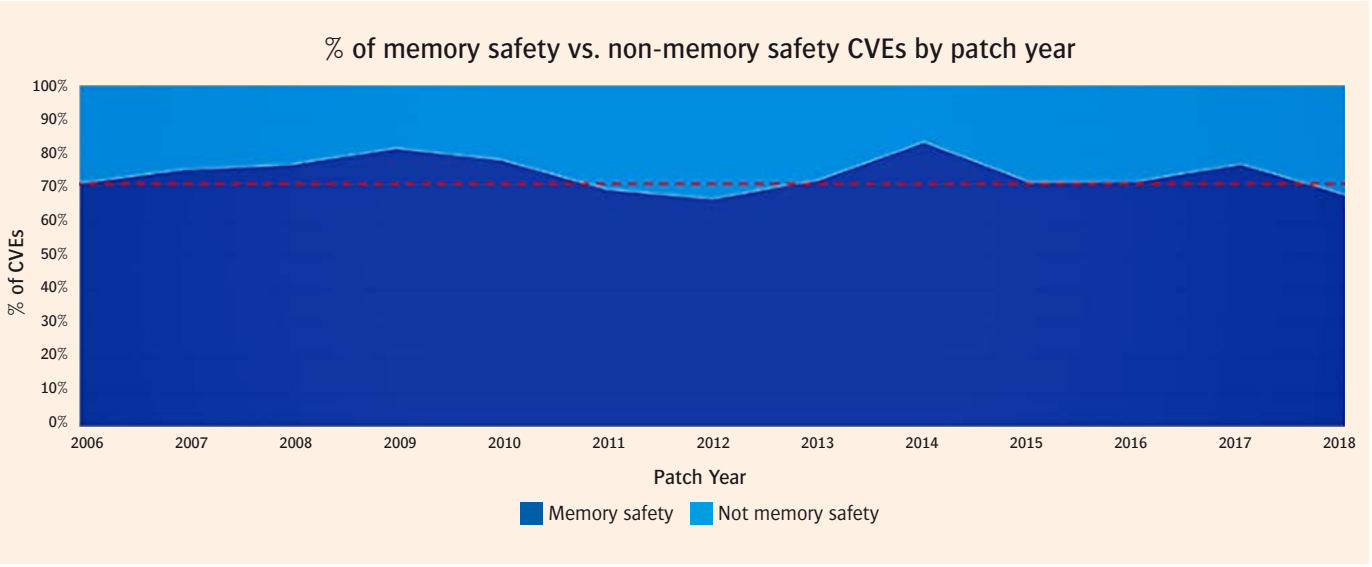
As discussed during the hearing, memory-safety issues represent roughly 70-80% of cyber issues – known as common vulnerabilities and exposures, or CVEs – being tracked by the industry. The number of CVEs has grown exponentially over the last twenty years, with the percentage of memory-safety CVEs remaining roughly constant. The reason is primarily related to the fact that most software is written in languages like C and C++, which do not provide inherent memory protection. To complicate the problem even more, software is not normally developed monolithically but by integrating pre-developed software from third parties, including open source, where anyone can contribute malicious changes.

Published CVE records

As a rough estimate, there are over one trillion lines of code in use today. The software industry has improved over the last few decades, especially regarding verification – the part of the development process that checks for bugs and corrects them. However, as verification will never be perfect (and let's be honest, nor will developers), there will always be bugs for hackers to exploit.

It's no surprise, then, that in February 2024 the White House Office of the National Cyber Director (ONCD) in the US released a report calling for the adoption of memory-safe programming languages. However, as noted in the 24 April hearing, companies have been slow to address memory-safety issues even where solutions are readily available. Although cyberattacks cause roughly $10 trillion to be lost worldwide every year, the direct impact on individual companies is sufficiently small that many do not choose to protect their customers. This situation could be compared to the automotive industry: it took decades for basic safety features to become standard in all vehicles, and it was only in response to regulations that every manufacturer ended up implementing them.

The good news is that pressure is mounting for digital products to be secure by design, and that innovative technologies have been put forward to help in this effort. CHERI – or Capability Hardware Enhanced RISC Instructions – developed by the University of Cambridge, extends instruction set architectures (ISAs) with new architectural features to enable fine-grained memory protection and highly scalable software compartmentalization. Until recently, although research platforms deploying CHERI were available, there were no commercial offerings. This has changed with Codasip's launch of a commercially available CHERI processor for license and the company's pledge to make its entire portfolio of processors available with CHERI variants. In parallel, Codasip is working with the University of Cambridge to ensure CHERI is standardized and available to everyone.



Percentage of CVEs caused by memory-safety issues - Credit: Matt Miller, Microsoft Security Response Center

Design provenance and traceability

Another topic discussed during the hearing was design provenance, which we believe must also include traceability. By provenance, we mean the origin of the design, including knowledge the specific designers; by traceability, we mean changes to the design over time, including knowledge of the specific designers who made the changes. Additional information regarding the design, such as when, where and with which tools changes were made should also be collected.

As John Goodacre and Richard Grisenthwaite explained during the 24 April hearing, most semiconductor chips today are complete systems in themselves, containing billions of transistors. Given the incredible complexity, chips are not designed monolithically, transistor by transistor; instead, they are assembled from pre-designed intellectual property (IP) blocks, such as processors, memory, on-chip interconnects between IP blocks, and chip-to-chip interconnects such as USB. Within the IP industry for semiconductors, companies such as Arm and Codasip make processor IP blocks, while companies like Cadence and Synopsys make memory and chip-to-chip interconnects. However, as Professor Goodacre and Mr Grisenthwaite discussed, some IPs are more prone to cyber issues than others, with processors being the most important and the most problematic.

When one licenses IP blocks, one does not know who actually designed the IP, nor its possible history of modification. Consequently, when the inevitable bugs are found, it is not possible to irrefutably determine who made the errors. Most bugs will be accidental, but it is also possible that nefarious actors could have inserted malicious circuitry to appear as an accidental bug. We believe that provenance and traceability will increase in importance as cyberattacks increase in frequency and are increasingly used in military conflicts. Indeed, as an article in *The Economist* recently noted, '[t]he cyberwar in Ukraine is as crucial as the battle in the trenches.'

Fortunately, Codasip is also addressing the problem of provenance and traceability with a new software tool using blockchain technology to irrefutably log the processor design process and create a record of provenance with traceability. This new software tool is currently being demonstrated to customers in a pre-release version.

At Codasip, we are proud to be at the vanguard of processor cybersecurity solutions, addressing the major issues of memory safety and IP provenance and traceability. We would be happy to discuss these developments with the HiPEAC community – just get in touch if you want to know more.

This article is adapted from an open letter to the UK Science, Innovation and Technology Committee, available on the Codasip website
bit.ly/Codasip_open_letter_cyber_resilience_May24

"Although cyberattacks cause roughly $10 trillion to be lost worldwide every year, the direct impact on individual companies is sufficiently small that many do not choose to protect their customers"

Headquartered in Munich, TASKING provides software development tools for safety-critical real-time architectures. In this article, Dr. Alexander Herz, software engineering manager at TASKING, tells us more about the company.

Software development for real-time critical systems with TASKING

COMPANY: TASKING
MAIN BUSINESS: software development tools for safety-critical systems
LOCATION: Munich, Germany
WEBSITE: tasking.com

TASKING's embedded software development solutions provide an industry-leading ecosystem for the entire development and verification process of safety-critical software up to the highest safety levels in different industry domains.

Trusted by over 30,000 engineers worldwide, our complete portfolio of integrated compilers, debuggers and target libraries enable our users to easily create reliable, safe and high-performance applications across various industries, including Tier-1 automotive suppliers, leading car manufacturers, and adjacent markets.

Our novel static analysis tools Safety Checker and Inspector make the process of developing safe software more cost effective by identifying access violations and parts of customers' source code that might be affected by a known compiler bug automatically before the code is executed (or even submitted into source control) and without the need to create high test coverage.

In 2022, TASKING expanded its offerings through the acquisition of iSYSTEM, enhancing our suite with advanced debugger, tracers and software analysis tools. These additions complement our existing portfolio, reinforcing our commitment to delivering the industry-leading development ecosystem for high-quality, functionally safe embedded software solutions.

From ISO26262- and ISO21434-compliant compiler tool sets and libraries certified by the TÜV safety-testing laboratory, to hardware debuggers like BlueBox and feature-rich integrated development environments (IDEs) like WinIDEA, we provide comprehensive support for various architectures, including Infineon's TriCore-based AURIX, NXP's Cortex-M/R based S32K and many more.

Using the TÜV-certified TASKING tools, together with the TASKING Qualified C-Library and the enhanced debugging and tracing capabilities from iSYSTEMS, qualification of customers' final application binary is made simple.

In order to provide the necessary performance required for the growing demand of artificial intelligence (AI) and other complex workloads in automotive applications, we provide optimizing C/C++ compilers for specialized accelerator cores available on many modern system-on-chips (SoCs) which are usually only addressable via specialized low level assembly language. Using our Profiler, customers can easily measure which parts of their source code use hardware resources inefficiently and adapt their core for best performance.

Moreover, we offer versatile math libraries like LAPACK for AI and other linear algebra computations, even on embedded systems without dedicated accelerators.

FURTHER INFORMATION:
tasking.com
linkedin.com/company/tasking-inc

"TASKING's portfolio of integrated compilers, debuggers and target libraries enable our users to easily create reliable, safe and high-performance applications"

Innovation Europe

In this edition we learn about publicly funded projects promoting interoperability for secure data processing and storage, as well as dealing with cyber threats. The eFlows4HPC project celebrates its close with a suite of services for the high-performance computing (HPC) sector, while EMPYREAN introduces a distributed computing paradigm, AVITHRAPID powers drug discovery and FALCON promotes more efficient aircraft.

Funded by the European Union

HANDLING HEALTH DATA WITH CONFIDENCE

How SECURED enables secure processing, anonymization and synthetic generation

Securely and privately storing and processing data has grown increasingly difficult in an era where data is both more easily generated and more valuable than ever. This is especially true with regard to health data, which is highly private and sensitive, and therefore subject to stringent data protection regulations like the General Data Protection Regulation (GDPR).

The SECURED project contribution to this challenge is to scale up the secure processing, anonymization and synthetic generation of health data. In particular, SECURED is targeted at providing a platform and architecture for the secure, trusted, efficient, decentralized and cooperative processing of health data. For this purpose, a number of techniques are being investigated, optimized and tested, in the domains of secure computation, federated and unbiased learning, data anonymization (as well as anonymization assessment via pre-emptive de-anonymization), and the generation of new, high-quality and private synthetic data.

The techniques are being implemented as part of the 'SECURED Innohub', which will provide a set of tools and services required for the secure interconnection of EU health data hubs, and will also enable other stakeholders, such as the health data analytics research community, health-application innovators such as e-health small / medium enterprises (SMEs), and end users to contribute to innovative data-based solutions without infringing the privacy rights of patients. The SECURED Innohub will be developed in a modular way, and will be expandable by external stakeholders, ensuring the sustainability of the project results beyond the end of SECURED. Ultimately, this will enable health datasets to be leveraged to their full potential across Europe, while retaining their security and privacy.

The SECURED approach follows two parallel, independent yet interacting, flows to innovation: the data flow and the processing flow. SECURED will showcase the technologies developed in four health-related use cases provided by partner hospitals and

Members of the SECURED project consortium

health stakeholders, namely: real-time tumour classification; telemonitoring for children; synthetic data generation for education; and access to genomic data.

COORDINATOR: Francesco Regazzoni (University of Amsterdam)
PROJECT NAME: SECURED: Scaling Up secure Processing, Anonymization and generation of Health Data for EU cross border collaborative research and Innovation
START/END DATE: 01/01/2023 – 31/12/2025
KEY THEMES: cybersecurity, privacy, machine learning, big data, healthcare, personalized medicine
PARTNERS: **Netherlands**: University of Amsterdam (coordinator), Erasmus University Medical Center; **Belgium**: Circular Economy Foundation, KU Leuven, NXP Semiconductors; **Cyprus**: Catalink; **France**: Thales SIX; **Greece**: Athena Research Center, Institute of Computer and Communication Systems (National University of Athens); **Hungary**: Budapest University of Technology and Economics, Semmelweiss University; **Ireland**: University College Cork; **Italy**: University of Sassari; **Spain**: ATOS Spain, Barcelona Supercomputing Center, Biomedical Research Foundation of the Niño Jesús University Children's Hospital, Josep Carreras Leukaemia Research Institute
BUDGET: €6,999,723.75

secured-project.eu
linkedin.com/company/secured-project
@SecuredEU
facebook.com/securedeuproject
youtube.com/@securedeuproject

SECURED has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement ID 101095717.

SAFE4SOC: DEFINING A CYBER AND PHYSICAL INCIDENT-DETECTION FORMAT TO FACILITATE CYBERSECURITY COLLABORATION

To fight against the exponential growth of cybercrime, organizations and, at a higher level, nations, have to cooperate and share their security information. SOCs (security operation centres) need to share their threat intelligence information, including across borders. The Safe4Soc (Security Alert Format Exchange For SOC) project aims to facilitate fast cyberthreat detection-sharing between different security-monitoring entities.

The overall objective of the project is to enable regulation compliant and accurate security-information sharing. To meet this objective, the project tackles three main challenges, as follows:

- define privacy-friendly and secure information-sharing mechanisms
- provide a structured information scheme to ensure semantic-rich exchanges, and
- develop artificial intelligence (AI) tools to enhance information sharing and ensure its relevance for the communicating entities.

The first step towards effectively sharing detection information is to ensure compliance with regulations (the GDPR, the NIS Directive, …). Information handled by SOCs is by nature sensitive. Safe4Soc will leverage and further develop a data-sharing gateway that can protect data in transit while ensuring that the required regulations are applied to the information exchange.

The second step is to structure and standardize this information and to define a common format. Detecting cyberthreats requires the deployment and operation of many sensors of different types, as well as the correlation of the information they produce to identify complex attacks or incident scenarios. A common format is the basis of interoperability for collaboration and ensures that shared information is interpreted by all parties in the same way. The format IDMEFv2 (Incident Detection Message Exchange Format), which describes both cyber and physical incidents, will be defined and proposed as an RFC ('Request for Comments' document) at the IETF (Internet Engineering Task Force).

The third step is to make sense of the information exchanged, on both sides. To do this, the Safe4Soc project will leverage advanced AI algorithms to improve detection and transform raw data into a form that is suitable for sharing (e.g. graphs). It will also develop algorithms to enable practical information sharing, ensuring that the receiving SOC can reuse the information for additional detection and remediation.

A Safe4Soc use case would be a national security entity monitoring the overall (cyber and physical) security of its critical operators. Each operator would keep its own monitoring system, and major alerts would also be sent to the central entity. Thus, Safe4Soc service providers can propose a national situational awareness centre. National centres could also centralize their information at a higher-level European situational awareness centre. Gathering alerts from critical operators may also help detect massive simultaneous and / or combined attacks on vital infrastructure – a type of attack that could be a major component of future hybrid wars. The field of defence and the armed forces, including during European coalitions, and particularly in the face of new dangers facing the European Union, is also a use case where the Safe4Soc project could provide relevant unified monitoring solutions.

PROJECT NAME: Safe4Soc: Standard Alert Format Exchange for Security Operation Center
START/END DATE: 01/01/2024 – 31/12/2026
KEY THEMES: cyber-physical systems, cybersecurity, incident detection, standards, cross-border SOC
PARTNERS: **France**: Institut Mines Telecom (coordinator), CEA, Teclib; **Germany**: Fraunhofer, Nicos; **Italy**: EHT; **Spain**: VicomTech; **Lithuania**: NRD Cyber Security, Vytautas Magnus University
BUDGET: € 7,101,101.36 (EU contribution: € 3,550,550.68)

safe4soc.eu
idmefv2.org
linkedin.com/company/safe4soc
linkedin.com/groups/13006336

Safe4Soc has received funding from the European Union's Digital Europe Programme under grant agreement number 101145846

EOSC-ENTRUST: DRIVING EUROPEAN INTEROPERABILITY FOR SENSITIVE-DATA ACCESS AND ANALYSIS

EOSC ENTRUST
European Network of Trusted Research Environments

The EU-funded EOSC-ENTRUST project launched on 1 March 2024 with the aim of enhancing European interoperability for sensitive data access and analysis. Led by ELIXIR and the EUDAT Collaborative Data Infrastructure, the project brings together partners from 15 European countries and will run for three years, with an EU contribution of €4.2 million.

'EOSC-ENTRUST brings together two major data-centric infrastructures, EUDAT and ELIXIR. With our experience and international networks, we will address the growing multidisciplinary demand for secure digital environments by orchestrating European-wide coordination and interfaces with other EU organizations and initiatives like EOSC and the European Data Spaces,' said Yann Le Franc, head of the EUDAT secretariat.

EOSC-ENTRUST will build a European network of trusted research environments (TREs) for sensitive data and develop a standard blueprint, or reference architecture, for federated data access and analysis. TREs provide secure platforms for handling sensitive datasets, enabling data privacy and legal compliance in research involving confidential information. Europe's landscape of TREs is currently fragmented, presenting challenges in access and management for researchers and providers. The variety of systems and procedures complicates research efforts, and there is a pressing need for a unified approach to managing federated access across differing technologies and governance frameworks.

EOSC-ENTRUST brings together providers of operational TREs from 15 European countries with a shared goal to implement, validate and promote their capabilities through a common European framework using shared standards and a common legal, operational and technical language.

'We are delighted to launch the EOSC-ENTRUST project and look forward to working with Europe's TRE providers and wider sensitive-data community to produce an interoperability blueprint to enable biomolecular and biomedical research involving sensitive data,' said Peter Maccallum, ELIXIR's chief technical officer and coordinator of EOSC-ENTRUST.

The project will create a reference architecture for interoperability based on the European Open Science Cloud (EOSC) Interoperability Framework to address legal, organizational, technical and semantic interoperability aspects. It includes driver projects in genomics, clinical trials, social science, and public-private partnerships to test and refine this blueprint and facilitate secure data analysis through federated workflows. Targeted outreach activities will expand the provider network and develop policy papers and guidelines to create a long-term operational TRE framework within EOSC. EOSC-ENTRUST will work closely with the SIESTA and TITAN projects, funded in the same call, to provide trusted environments for sensitive data management in EOSC.

PROJECT NAME: EOSC-ENTRUST: A European Network of TRUSTed research environments
START/END DATE: 01/03/2024 – 28/02/2027
KEY THEMES: research infrastructure, cloud computing, data, privacy, interoperability
PARTNERS: **Germany**: ELIXIR Hub (EMBL-EBI) (coordinator), GESIS - Leibniz-Institut für Sozialwissenschaften; **Belgium**: Sciensano, Vlaams Instituut voor Biotechnologie (VIB); **Czech Republic**: Masaryk University, VSB - Technical University of Ostrava; **Denmark**: Danmarks Tekniske Universitet (DTU); Eston **Estonia**: University of Tartu; **Finland**: CSC – IT Center for Science, EUDAT Collaborative data infrastructure, Finnish Institute for Health and Welfare, Turku University of Applied Sciences; **France**: ECRIN (European Clinical Research Infrastructure Network); **Greece**: GRNET – National Infrastructures for Research and Technology; **Hungary**: Tárki Foundation; **Luxembourg**: Luxembourg National Data Service (LNDS); **Netherlands**: Stichting Health-RI, SURF; **Norway**: CESSDA ERIC, NTNU - Norwegian University of Science and Technology, Sigma2 AS, University of Bergen, University of Oslo; **Portugal**: BioData.pt; **Slovenia**: University of Ljubljana; **Spain**: Barcelona Supercomputing Center, Centre for Genomic Regulation (CRG); **Sweden**: Uppsala University; **UK**: Health Data UK, University of Dundee, University of Essex, University of Manchester, University of Nottingham
EU CONTRIBUTION: € 4,218,809.75

eosc-entrust.eu
linkedin.com/company/eosc-entrust
@eosc_entrust

The EOSC-ENTRUST project has received funding from the European Union's Horizon Europe programme under grant agreement number 101131056.

EFLOWS4HPC DELIVERS A EUROPEAN HPC WORKFLOW PLATFORM

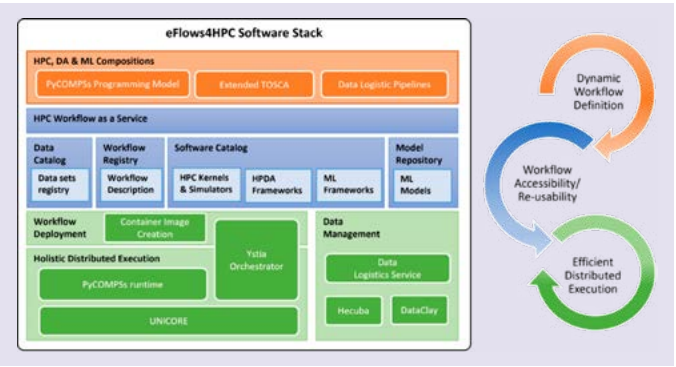
Finalized in February 2024, the three-year eFlows4HPC project has delivered a comprehensive workflow platform and accompanying services to facilitate the integration of high-performance computing (HPC) simulation and modelling with big-data analytics and machine-learning techniques. With a budget of €7.6 million and a consortium of 16 partners from seven countries, eFlows4HPC made significant advances in defining complex workflows, resulting in enhanced workflow efficiency and increased overall development and deployment productivity.

Scientific workflows are a key tool in modern science, allowing researchers to grapple with the computing and data analysis needs of current scientific applications. However, prior to eFlows4HPC, developers lacked the tools necessary to enable the development of complex workflows involving HPC simulations and modelling with data analytics and machine learning.

The two main outcomes of eFlows4HPC, high-performance computing workflows as a service (HPCWaaS) and the project's open-source software stack, are fundamental in supporting the development, deployment and execution of complex workflows. HPCWaaS helps users and developers manage complex workflows by simplifying the lifecycle steps, while the eFlows4HPC software stack is applicable across numerous scientific and industrial applications requiring HPC, artificial intelligence (AI) and big-data techniques.

The eFlows4HPC workflows were put to the test in three areas: digital twins in manufacturing, advanced climate modelling, and urgent computing for natural hazards such as earthquakes and tsunamis.

According to eFlows4HPC coordinator Rosa M. Badia (Barcelona Supercomputing Center), 'the results mark a significant milestone, demonstrating the versatility of the eFlows4HPC software stack in leveraging HPC resources. This translates into a notable decrease in time-to-solution for application developers who need to combine traditional HPC with AI or big data in a single workflow'. She added: 'We are shaping the future of HPC workflows and empowering researchers to tackle emerging challenges that require complex workflows.'



Joining forces with European initiatives to multiply impact

As reported in *HiPEACinfo* 68, to ensure wide adoption of eFlow4HPC methodologies by scientific and industrial communities, the project consortium teamed up with multiple initiatives. On the applications side, joint activities were organized with HPC centres of excellence such as ESiWACE (climate and weather simulations) and PedMedCoE, while the project also supported the ChEESE centre of excellence by developing complex workflows for earthquake simulation.

Similarly, eFlows4HPC methodologies will be used for the development of workflows in DT-GEO, which is building a digital twin for geophysical extremes, and the CAELESTIS project for next-generation aircraft. SKA Regional Centres, the regional support network for the international initiative to provide square-kilometre array telescopes, have also shown interest in evaluating the HPCWaaS methodology.

Meanwhile, in collaboration with the European Processor Initiative (EPI), eFlows4HPC optimized kernels on ARM-based and RISC-V architectures for machine-learning operators that can replace the compute-intensive simulations in application workflows. The optimized kernels were then migrated to the EPI, while the ConvLIB library, a key exploitable result of eFlows4HPC, was used to harness the vector processing unit in the EPI and significantly accelerate the performance of the convolution.

FURTHER READING

eFlows4HPC website eflows4hpc.eu

Key exploitable results eflows4hpc.eu/key-exploitable-results

Software repository eflows4hpc.eu/software

HPCWaaS step-by-step guide youtu.be/2nCUjHoA3s0

eFlows4HPC received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement no. 955558. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Spain, Germany, France, Italy, Poland, Switzerland, Norway. It also received funding from MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR (PCI2021-121957).

INTRODUCING EMPYREAN'S HYPER-DISTRIBUTED PARADIGM FOR THE IOT-EDGE-CLOUD CONTINUUM

The conventional way of dealing with internet-of-things (IoT) data is to push them to the cloud or utilize edge resources, forming an IoT-edge-cloud continuum. This continuum is usually implemented as a monolithic pipeline that cannot efficiently serve hyper-distributed and artificial intelligence (AI) / machine learning (ML)-based applications. It is clear that more local decisions and a collective logic that leads to system-wide welfare optimality are needed.

This is where EMPYREAN steps in, proposing a new hyper-distributed computing paradigm that encompasses heterogeneous IoT devices and computing, storage and connectivity resources that may belong to different providers at different segments of the continuum (see figure below). EMPYREAN builds federations of collaborative resources, to be referred to as 'IoT-edge associations' or simply 'associations', that will be created and be operated autonomously and seamlessly using the EMPYREAN AI-enabled management scheme. This association-based continuum utilizes distributed, cognitive and dynamic AI-enabled decision-making to balance computing tasks and data locally, both within an association as well as between federated associations, in a decentralized, multi-agent manner and across central computing environments. In this way, it optimizes resources, and provides scalability, resiliency, energy efficiency and quality of service.

The EMPYREAN consortium is developing technologies for 'association-native' application development and deployment technologies that contribute to the entire application lifecycle and interoperability. This will be used to power AI-driven value extraction of high volume and dynamic IoT data generated at the edge of the network from multiple sources.

The innovative capabilities of the EMPYREAN platform are being demonstrated through three use cases representing device- and data-rich applications in advanced manufacturing, smart agriculture and warehouse automation. All three use cases involve continuous federated learning from distributed data. Also, a South Korea-based use case in smart factories will showcase the use of the EMPYREAN platform in an international setting, promoting collaboration with trusted regions to ensure a guaranteed level of interoperability and portability.

PROJECT NAME: EMPYREAN: Trustworthy, cognitive and AI-driven collaborative associations of IoT devices and edge resources for data processing

START/END DATE: 01/02/2024 – 31/01/2027

KEY THEMES: hyper-distributed, collaborative, IoT, edge, interconnection, cognitive, trust, artificial intelligence, machine learning (ML), autonomous, ML workloads, manufacturing, agriculture, warehousing, robots

PARTNERS: **Greece**: Institute Of Communication and Computer Systems, Nubis Idiotiki Kefalaiouchiki Etairia; **Israel**: NVIDIA Mellanox Technologies Ltd; **Denmark**: Chocolate Cloud APS; **Spain**: Universidad de Murcia, Ideko S Coop; **France**: Zettascale Technology SARL, Ryax Technologies; **Germany**: Nec Laboratories Europe GmbH; **Belgium**: Institute for Agricultural, Fisheries and Food Research, Tractonomy Robotics BVBA

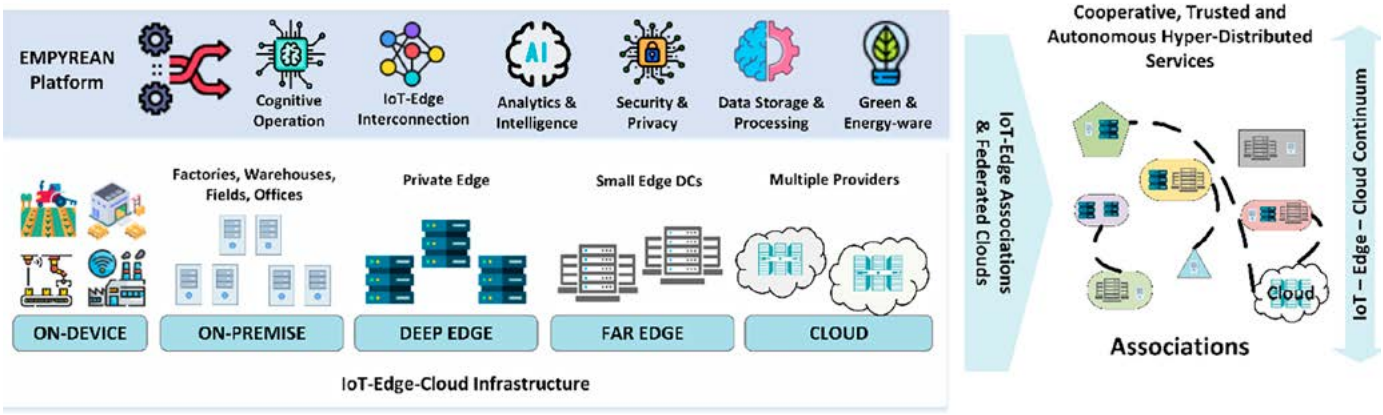
BUDGET: €4,673, 541.70

empyrean-horizon.eu

linkedin.com/company/empyrean-project

@empyrean_he

EMPYREAN has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement number 101136024



The EMPYREAN platform, utilizing IoT, multiple edge levels and federated cloud resources, will empower the 'everything as a service' notion for the IoT-edge-cloud continuum

FLYING HIGH: FALCON'S QUEST FOR QUIETER, SMARTER AIR TRANSPORT

Direct aviation emissions accounted for 3.8% of total CO_2 emissions and 13.9% of the emissions from transport in the EU in 2017, making aviation the second biggest source of greenhouse-gas emissions after road transport. In addition, although the noise emissions of each aircraft have decreased approximately by 75 % over the last 30 years, the growing amount of air traffic means that many EU citizens are still exposed to high noise levels.

Intensified research and innovation activities are therefore needed to reduce the impact and emissions of aviation (CO_2 and non-CO_2, noise, manufacturing) in order for the EU to reach its policy goals of net-zero greenhouse-gas emissions by 2050. One of the main levers to decrease CO_2 emissions is to reduce the airframe structural weight.

In response, FALCON's ambition is to develop a hybrid approach combining cutting-edge numerical and experimental methods to analyse fluid-structure interaction and better predict and control unsteady aerodynamic loads, thus improving the aeroelastic properties and sustainability of aerostructures and reducing the related aerodynamical noise.

This will ultimately contribute towards upscaling the design capabilities of the European aircraft industry while advancing the digital transformation of the European supply chain.

Building upon three industrial testcases and tight links with key European partnerships such as Clean Aviation, four specific objectives will be pursued:

- To **capture the essential fluid-structure interaction phenomena** occurring in realistic aeronautical conditions combining experiments and simulations through the establishment of a reference set of industrial test cases.
- To **simulate the aeroelasticity and related noise emissions** occurring in realistic aeronautical conditions thanks to high-fidelity and high-performance lattice Boltzmann method frameworks.
- To **increase the use and access of highly scalable, high-performance computing frameworks** for industrial fluid-structure interaction applications, while obtaining cost-efficient and timely results.
- To **control the aeroacoustics and aeroelastic instabilities** originating from fluid-structure interaction using multi-fidelity optimization.

FALCON will take the industrial leadership of aeronautical industry in Europe forward, aligning the outcomes of the project towards noise and fuel reductions, as well as towards sustainability improvements of global air transport.

PROJECT NAME: FALCON: Foreseeing the next generation of Aircraft: hybrid approach using Lattice-boltzmann, experiments and modelling to optimize fluid/structure interactiONs
START/END DATE: 01/01/2024 – 31/12/2027
KEY THEMES: sustainable aviation, fluid structure interaction, lattice Boltzmann, aeronautics, aeroacoustics, aeroelasticity, aircraft design improvements
PARTNERS: **France**: Université d'Aix Marseille (coordinator), Centre National de la Recherche Scientifique, Ecole Centrale de Marseille, CS Group, Euronovia, AIRBUS Operations SAS, MSC Software SARL; **Germany**: Karlsruher Institut fuer Technologie, Deutsches Zentrum für Luft- und Raumfahrt, AIRBUS Operations GMBH; **Spain**: AIRBUS Operations SL; **Belgium**: MSC Software Belgium; **Czech Republic**: Technical University of Ostrava; **United Kingdom**: AIRBUS Operations Limited
BUDGET: € 4,878,345.25

falconproject.eu
linkedin.com/company/falconproject
@falconeuproject

FALCON has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101138305

Members of the FALCON consortium

ENSURING THE SUPPLY OF ANTIVIRALS FOR PANDEMIC RESPONSE WITH AVITHRAPID

If the COVID pandemic taught the world anything, it was that it is essential to prepare for the possibility of future pandemics of infectious diseases. Clearly, part of this preparation effort must focus on ensuring that there is a supply of drugs to help minimize the impact of future outbreaks. As a result, the EU has dedicated €50 million to investigating broad-spectrum antiviral therapies for infectious diseases for epidemic potential. Of this, €7.5 million has been allocated to the AVITHRAPID project, which is being carried out by a consortium of 18 research institutions and companies from eight European countries, with further funding provided by the Swiss State Secretariat for Education, Research and Innovation (SERI).

Coordinated by Germany's Fraunhofer-Gesellschaft, AVITHRAPID aims to combat and mitigate the impact of future pandemics, drawing on cutting-edge technologies including molecular modelling, biochemical and cell-based assays, X-ray crystallography, and more. The consortium is working on establishing a pipeline of preclinical drug candidates, focusing on small molecules whose bioactivity against various viral targets has been already shown and which can be rapidly translated into novel antiviral agents against emerging infectious diseases with pandemic potential. In addition, the consortium will conduct a Phase 2a clinical trial for a small molecule developed against Zika virus, thereby translating AVITHRAPID research into therapeutic solutions.

As a broader objective, AVITHRAPID aims to identify and develop potential drug candidates against other pathogens, such as SARS-Cov-2 and hepatitis viruses, using artificial intelligence and molecular simulation techniques to expedite the drug-development process. Beyond combatting infectious diseases, the project also aims to catalyse investment and promote partnerships within the pharmaceutical industry. By creating a European 'ecosystem' to tackle infectious diseases within pandemic potential, AVITHRAPID aims to enhance Europe's knowledge and preparedness for future pandemics.

Inaugurated on 1 January 2024, with an end date of 30 June 2028, AVITHRAPID will therefore improve our capacity to contain the threat posed by pandemic infectious diseases.

PROJECT NAME: AVITHRAPID: Antiviral Therapeutics for Rapid Response Against Pandemic Infectious Diseases
START/END DATE: 01/01/2024 – 30/06/2028
KEY THEMES: drug discovery, pandemic preparedness, medicinal chemistry, infectious diseases
PARTNERS: **Germany**: Fraunhofer-Gesellschaft (coordinator); **Czech Republic**: VSB – Technical University of Ostrava; **France**: University of Tours, National Institute of Agricultural Research (INRA); **Italy**: Tuscia University, EuResist Network, Elettra Sincrotrone Trieste, University of Naples Federico II, Lazzaro Spallanzani National Institute for Infectious Diseases, University of Siena, University of Calgari, Tor Vergata University of Rome, Dompé Farmaceutici; **Latvia**: Latvian Institute Of Organic Synthesis; **Netherlands**: First Health Pharmaceuticals; **Portugal**: Institute of Molecular Medicine; **Switzerland**: Swiss Tropical and Public Health Institute, Chelonia
BUDGET: € 7,459,253.75

avithrapid.eu
linkedin.com/company/avithrapid
@avithrapid

AVITHRAPID is funded by the European Union under grant agreement no. 101137192.

Members of the AVITHRAPID consortium

With unstructured data growing at 50% or more per year, demand for object-storage solutions that offer robust security is high. In this interview technology journalist Stuart Cording ('The Electronics Reporter') spoke to Sascha Uhl, a solution architect at HiPEAC 2024 sponsor Cloudian, to find out more.

'Cloudian makes it easy to expand the data lake'

What does Cloudian offer its customers?

Cloudian was founded with a clear vision: to bring cloud-native storage, made familiar by Amazon Web Services (AWS), to on-premise datacentres within enterprises. This means our customers can use the same standard AWS S3 application programming interface (API) locally and with limitless scalability. Our customers' storage needs range from small environments with a couple of terabytes up to hundreds of petabytes using S3 and offering S3 storage. Cloudian is a software vendor, so we offer both hardware and software, but we also partner with Lenovo and HPE so our customers can choose the kind of hardware they want.

What are the main challenges your customers face?

Customers are usually looking for solid, secure and future-proof systems when it comes to storage. Gone are the times when you bought storage silos – different kinds of storage systems for different applications or use cases. While the needs of customers are different with every use case, our strength is that we can adjust the storage and work actively with our customers to provide the perfect environment for their needs.

Cloudian's team, seen here at HiPEAC 2024, work to ensure their storage solutions meet each customer's needs

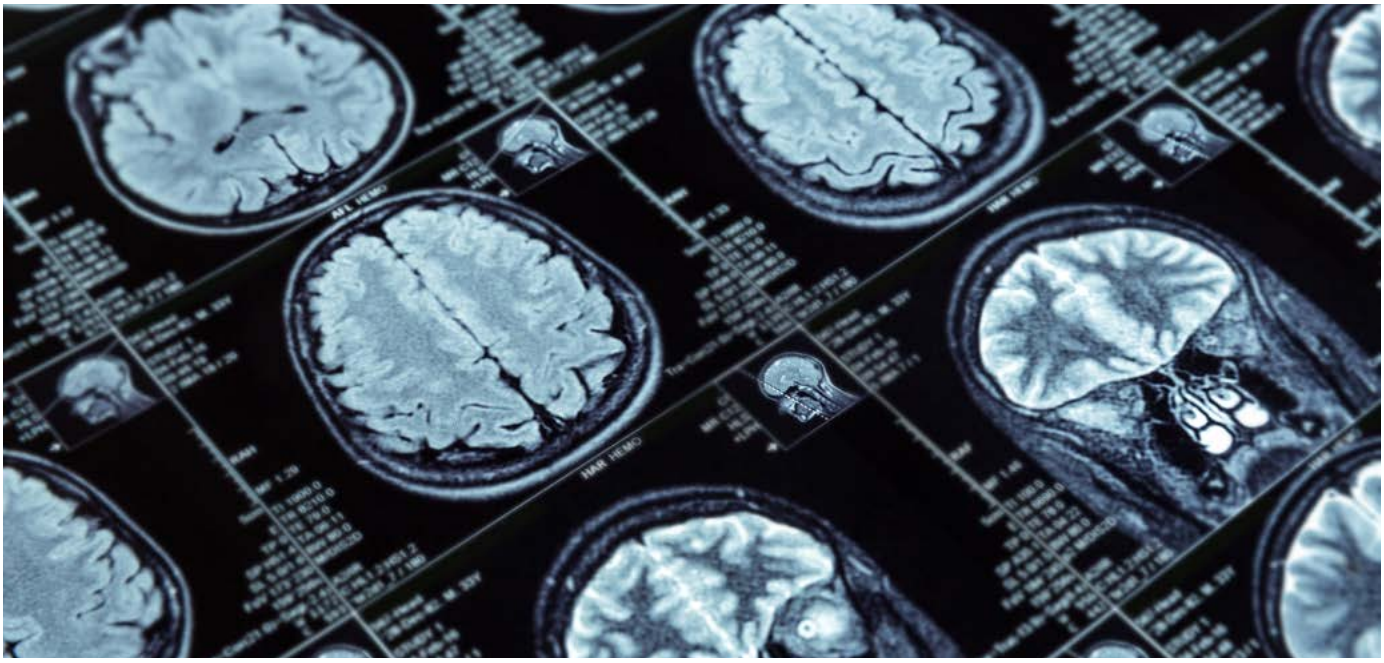
In many cases, customers also demand a high standard of security, because we are often talking about sensitive data, or data destined to be stored long term in archive. There are also projects involving artificial intelligence (AI), machine learning, with use cases where sensitive data is stored. The main thing is that we understand that every customer is different and we try and design a system together with them to ensure it fits their needs.

Most organizations have plenty of storage on offer with services integrated into Office 365, Google Drive or AWS. How is Cloudian different?

A lot of companies actually had their cloud-first approach, but many are now starting to build their own storage in their own datacentres. With Cloudian we offer the same APIs that they used in the cloud and they can even build their own applications, and the same APIs in the local data centre. Customers may have multiple datacentres, maybe operate over multiple regions – they might even be a global company that operates over several continents. The data needs to fit into the data strategy here and we help customers build solutions which fit their requirements while offering a technical solution that is the same or even better than that which they got from the cloud.

Storage is only half the challenge though; the other is using that storage effectively. How can you help organizations improve, for example, their data backup processes?

Actually, backup and archiving was, for many years, part of our bread-and-butter business, so we've done many projects in that area, although now our business is moving towards primary storage use cases like AI, machine learning, and so on. With backup use cases, the most important thing is security. Security does not only mean that data must be protected on the systems. It goes very deep, actually, because when you think about how backup vendors advertise their solutions, they often talk about immutability – but immutability at the end of the day is not provided by the backup vendor. The immutability comes from the storage system underneath.



Hospitals represent a typical archival use case for Cloudian, storing image data from their MRI or CT scanners

So we provide the APIs for these immutability features, which are then leveraged by Veeam, Commvault and others to build a system for backup storage. This capability and the built-in security for protecting data is key for our customers. But here for example the S3 API is the only API which communicates between the backup application and Cloudian. Cloudian itself is a Linux-based system and part of the solution for us is that we protect that system as well. So everything is locked down, there is no shell access and so on. For normal users we can guarantee that the data is still protected and can't be altered, changed, deleted and so on in the system itself.

Let's dig into that topic of data security in a little more detail. How is the customer's data actually protected?

When data is written to the storage, it is locked. So in the past we talked about warm storage [i.e. storage of data that is accessed less frequently], and actually we bring that to the object storage piece. This means that, once data is written to the object storage, the data will be locked and is then unchangeable for a given time frame, which is specified in the backup solution usually.

On the other side, we also ensure that attackers cannot alter or delete the data from our service. This is done by the special hardening which applies to our nodes on one side. It allows for the other backup applications to see object storage as an air-gapped solution because we have the free protocol which provides high security here, like an air gap in the days of tape, but as part of a modern storage system which can be used for multiple applications and use cases.

From the customer's perspective, we can lock it down completely and provide full management and maintenance for them, with hyper care support services that allow the customer to see Cloudian just as a black box.

What kind of businesses use Cloudian?

We have customers in a lot of verticals: for example, the banking sector, such as the Swiss institution PostFinance, a long-term customer, which is currently using Cloudian for some machine-learning use cases in a primary storage test. And interestingly enough, they are one of the customers where we are seeing an increase in consumption every year. So the whole data lake is growing, and Cloudian makes it very easy actually to expand that existing data lake.

We also have public institutions like the Austrian parliament and the University of Hamburg, which also have multiple use cases. The Austrian parliament, for example, uses Cloudian to publish court videos to citizens, which is a nice use case. There's also the typical archival use case for medical institutions like hospitals, who have to store image data from their MRI or CT scanners, for example, and keep this data protected long term for compliance reasons. And Cloudian is a good place to do that.

CLOUDIAN®

"Cloudian brings cloud-native storage to on-premise datacentres within enterprises"

From doing a PhD before computer science was mainstream at her university to launching an interdisciplinary centre and master's specialization, Virginia Niculescu is used to breaking new ground. In this interview she talks us through her career to date and why international networking is important for personal development.

Career talk: Virginia Niculescu

How did you first become interested in computer science?
Even from the first years of school I liked mathematics very much; solving logical problems seemed more like enjoyable play than hard work. Learning about the surrounding world was interesting, but I liked finding structures and extracting abstractions from it. During high school we didn't learn anything about computers or programming, but I found out that at the Faculty of Mathematics of Babeş-Bolyai University there was a new section – computer science – that seemed extremely promising for the future, since the best students were selected for it. This was how my path in this challenging domain began. I graduated in 1994 with top grades in this specialization, so going further for a PhD came naturally. Since then, I've progressed in my academic career to become an associate professor in computer science at the same university.

What are some highlights of your career so far?
The research domain of my PhD was parallel computing for numerical methods, an area that connects mathematics with computer science. My focus on high-performance computing (HPC) started at that time, and it was quite challenging because at our university there were not many people that worked in this area – even my PhD supervisor was a professor in mathematics who had just started to do research in parallel computation. Interaction and collaboration with researchers from other countries was also difficult, and for a long period of time I worked mostly alone.

Another difficulty was the lack of proper equipment: we didn't have a real HPC infrastructure until 2015. Before that we tried to cover this need by connecting workstations to have some ad hoc kind of parallel systems. Besides my interest in abstractions, this was another reason for going toward more theoretical approaches of parallel computing such as models, patterns, etc.

There were two important milestones in my activity related to parallel computing and HPC in general. Both are related to my attempts to enlarge the numbers of collaborators that I could work with and to increasing the interest of the students in the HPC domain.

First, the creation of the Research Center of Modeling, Optimization and Simulation (MOS), where I proposed a team formed of mathematicians and computer scientists. We work within the Faculty of Mathematics and Computer Science, giving us the opportunity to work together in areas that involve both kinds of expertise. The research centre was founded as part of a European infrastructure project through which our university finally succeeded in having a powerful computing cluster, too. I have been part of the management team for this computing infrastructure since 2014.

For the students I proposed a new master's specialization in high-performance computing and big-data analytics (HPC&BDA), starting in 2015, which I have coordinated since then. The information technology (IT) industry in Cluj-Napoca has developed quite a lot over the last few decades, but most firms are oriented towards web development or other technologies that do not necessarily need HPC. This situation has changed recently, but at the beginning it was a challenge to convince students that this specialization would be good for their future.

An important step was taken when we obtained the support of a large company, Bosch, that very early saw the potential of this specialization; since 2020, we have had a very good collaboration and support agreement for this master's degree offered by Bosch. Putting together these two domains, HPC and BDA, proved to be aligned to present trends, with artificial intelligence (AI) needing more and more HPC, but also HPC having the opportunity to be enhanced through AI.


Virginia is based at Babeș-Bolyai University (left) in the Romanian city of Cluj-Napoca (right)

While HPC was initially much more connected to computational science, it is now is going to be very much connected to AI. This need for HPC that comes from AI transforms it from a 'niche domain' into one that is very close to the mainstream. This leads to the appearance of new domains, such as high-performance data analytics, and at the same time boosts classical ones, such as modelling and simulations.

What is your day-to-day work like?

Since we may have classes at any time during the day (morning, afternoon, or evening), my schedule is not very regular. This involves great flexibility, which is great to have, but at the same time the lack of regularity makes me think sometimes that I never finish my workday. Our flexible programme also allows us to do our research work from home while having well-established hours for classes, consultations, research meetings and organizational activities. However, since starting the MOS research centre in 2015, I try to go and work there to promote research collaboration with colleagues and students.

How do you think networks such as HiPEAC can support their communities?

A network like HiPEAC is of great importance for the community. If I had had the opportunity to join such a network when I was a PhD student, the advantages would have been significant; they are now available for the students who join it.

You could say that in computer science (CS) collaboration is mandatory. If in mathematics there may be some fields where researchers may have good results even if they work independently, in CS and for sure in HPC this would be very difficult, if not impossible.

Another thing worth considering is the question about the degree to which success is influenced by opportunities to work at other universities abroad. After my PhD I was offered the chance to go to a university in the UK as a post-doc, but I didn't take up the opportunity for personal reasons. After that, I thought that I had missed the chance to enter a really high-level research stream. However, a network such HiPEAC offers opportunities to be up to date on research in the domain, to be aware of new trends and collaboration opportunities. It facilitates access to HPC infrastructure and to international research projects.

For me, joining an international research project is the main focus that would allow me, my collaborators, and students to be part of important research that could make a difference. I really believe that we are a valuable human resource that may contribute to such a development.

Finally, I would like to say something to the women interested in working and researching HPC. Perhaps the most important thing is to have trust in themselves – they have good potential and value that just needs to be harnessed. I gained this trust over time, too much time that could have been used more efficiently. Nonetheless, valuable experience could be gained also in small and medium research projects and there are a lot of rewarding activities that could be followed in this challenging domain.

Networks such as HiPEAC offer opportunities to connect and stay up to date

Inspired by the possibilities of global networking, Sara Royuela, an established researcher in the Predictable Parallel Computing Group at Barcelona Supercomputing Center, teamed up with colleagues to set up a Women in High Performance Computing (WHPC) chapter in Barcelona, named MAR-WHPC. To celebrate the International Day of Women and Girls in Science, we asked Sara why we need greater diversity in computing, how the MAR-WHPC chapter came about, and what the challenges ahead are.

'WHPC provides a "safe space" where people can share experiences'

What's all this fuss about gender diversity in computing? Can't we just carry on as we are?

We really need to improve gender diversity, and in fact all kinds of diversity, in computing research. Without it, we're potentially wasting half of the brilliant minds out there in the world. I believe we're lucky in the field of scientific research, because people tend to be open to these discussions. However, we still need to draw more attention to this topic, and this is a particularly important task for those in privileged positions.

Why did you decide to establish a WHPC chapter in Barcelona?

The decision really has its origins in an initiative which started at Barcelona Supercomputing Center (BSC). Some years ago, the Life Sciences Department at BSC – which has an almost equal gender balance – created a group named Bioinfo4Women, which both supports female researchers in their careers and highlights research issues relating to women. Inspired by this, researchers in the Computer Sciences Department at BSC set up our own group, Women in Computer Sciences (WiCS). Today, every research department at BSC has its own group focusing on issues relating to gender.

We made efforts to recruit members to WiCS, such as asking our contacts if they wanted to join and holding open days with students from the local university. However, everyone is very busy and obviously our main focus as engineers has to be on our research careers – publishing papers, attending conferences, etc. – so it was hard to keep momentum going. Then we attended a WHPC workshop at ISC HPC and we thought that, given WHPC's global reach, establishing a chapter in Barcelona would be a good way to attract more people to help share the load and make the group more dynamic. And that's what has happened. As an example, the MAR-WHPC chapter now has many members from HPC Now!, a local company, and the chapter is becoming consolidated.

That all sounds very positive. So is MAR-WHPC going to be confined to Barcelona?

Well, we now have members from all over Spain and we'd actually like to expand it to become a southern Europe chapter. WHPC has strong representation in the United States, for example, but not so much in the south of Europe. So it would be great to broaden our membership with people from many non-represented institutions from the rest of Spain and other southern European countries.

What does WHPC provide for people seeking to improve the gender balance in computing and raise awareness about gender-related issues?

WHPC provides several things. First, visibility, including providing role models. This is hugely important: role models allow other women to believe that they too can do this, that they belong in this field, while visibility helps them gain recognition, which is important to advance their careers. On a personal note, I remember being really inspired early by Mary Hall, when she was a teacher at the HiPEAC summer school, ACACES, back in 2011.

Second, WHPC provides networking opportunities. And third – and this might sound a bit strange –WHPC provides a 'safe space', that is a platform where people feel comfortable enough to share experiences, complaints and advice honestly and openly.

This is making me think of a panel at a womENcourage event organized by ACM Women, which participants described as 'group therapy'.

Exactly. It's important to provide the conditions in which people feel comfortable so that they can share these experiences.

> *"The key is to ensure that we integrate more diverse viewpoints at the point of creating technology"*

What do you think are the main challenges ahead of us in achieving a greater gender balance, and more diversity generally, in computing?

One key thing is to get rid of all the unnecessary layers and social pressures that hinder girls and women from embarking on careers in scientific research. Probably the greatest challenge is that, by the time they get to secondary school, most girls have lost interest in technology-related topics. We need to do a huge amount of outreach so that they can choose what they really want to do, with a genuine understanding of all the options available to them. As an example, I've arranged 'work experience' visits to BSC so that teenagers can get an idea of what it's like to work in high-performance computing research, but I've asked for the participants to be female.

What about potential tensions with people feeling that women are getting 'preferential treatment'? For example, if they feel that a speaker has been chosen for the fact that they are female rather than for the quality of their work?

This is another major challenge: getting men and boys to understand that feminism is for them, as well – it's about dismantling stereotypes, so that they don't have to conform to some kind of 'alpha male' template either. Once this has been achieved, I think people will stop seeing someone as 'the female speaker', for example, and instead see the individual in question.

Do you think things are getting better or worse? One example that comes to mind is that the application of artificial intelligence (AI) often seems to negatively affect women – for example, generative AI applications which reduce women to sexualized commodities.

Well, if you think of the data that AI is trained on (i.e. the internet), it's not surprising… I think the key is to ensure that we integrate more diverse viewpoints at the point of creating technology. By including different kinds of people, you can increase the range of ideas available. Something that gives me hope is that women seem to be entering the field of AI in large numbers. As an example, at the Universitat Politècnica de Catalunya, the master's degree in AI offered by the informatics faculty currently has a cohort of around 50% women and 50% men. Many of the students seem to come from diverse backgrounds, so this is encouraging.

Sara with Elisabeth Ortega of HPC Now! at HiPEAC 2024

Reasons to be cheerful! Finally, I can't let you go without asking whether you have any advice for girls or women who may become future researchers…

I'd just like them to be reassured that they are able to do what they want to do. They should surround themselves with people who empower them, and stay away from people who belittle them, like all of us should. It's also important for them to challenge themselves and get out of their comfort zone.

Follow MAR-WHPC on social media:
linkedin.com/company/mar-whpc
twitter.com/mar_whpc

In February, Laura Menéndez Gorina (Barcelona Supercomputing Center) took over the role of HiPEAC Jobs manager from Federico Iori. In this interview, Laura tells us a bit about her background, the importance of careers development, and her plans for HiPEAC Jobs.

'HiPEAC's most valuable asset lies in shaping the next generation of specialists'

Hi Laura! Can you tell us a bit about yourself?

The truth is that my professional background is quite particular: I graduated in translation and interpretation studies from the Autonomous University of Barcelona (UAB) and then did a PhD in literature and cultural studies at Stanford University. There I specialized in Iberian and Latin American literatures, with a focus on digital humanities, arts, and urbanism.

Over the course of almost ten years in the United States, however, my interests expanded significantly: being in close contact with the world of Silicon Valley and based at a university with a strong emphasis on science, technology, engineering and mathematics (STEM) and entrepreneurship, I progressively approached the field of computer sciences. Serving as professor and mentor of STEM students was also very enriching in learning about the ins and outs of careers in computing systems. At the Barcelona Supercomputing Center, I've come to delve deeper into the interplay between research, policymaking, and industry, as well as meeting people behind some of the major advances in the field. Without skilled researchers we wouldn't have those advances, and so I am excited to help bridge the gap between recruiters and candidates.

What is your impression of HiPEAC so far?

I believe that HiPEAC carries out a crucial mission in providing a solid network of collaboration between researchers, industry, and policymakers to help Europe strengthen its position in computing systems. HiPEAC is already well known for envisioning tomorrow's technologies; in my view, its most valuable asset lies in shaping the next generation of computer architecture and computing systems specialists through networking, training, and professional development opportunities to stay abreast of the latest advances in the field.

How can HiPEAC help with career development and recruitment?

The resources that HiPEAC offers to students and junior researchers, industry stakeholders, and research centres are manifold. The jobs portal and careers centre provide recruiters

and applicants with a unique, centralized focus on matching job opportunities. A series of in-person events further promote this interaction between academy, research centres, and industry. In particular, during the HiPEAC conference and ACACES summer school, students and jobseekers have the opportunity to meet with cross-sector experts and representatives in the field and find out for themselves about current trends and business needs. In addition to reinforcing contacts with companies and institutions, these events also allow members and participants to network and cooperate, from building common career opportunities to laying the foundations for future European projects.

Any plans for HiPEAC Jobs you'd like to share?

We are working on a number of improvements and novelties, and we are very excited about them. Firstly, we are updating our website and hope to soon be able to roll out a new feature for a more proactive job search approach using artificial intelligence (AI) to identify the best matches. We are also planning several brand-new initiatives for ACACES 2024 and HiPEAC 2025. Our main focus is on providing students with more opportunities, but also promoting startups and spinoffs. Stay tuned to the HiPEAC website and social media channels for forthcoming announcements.

hipeac.net/jobs

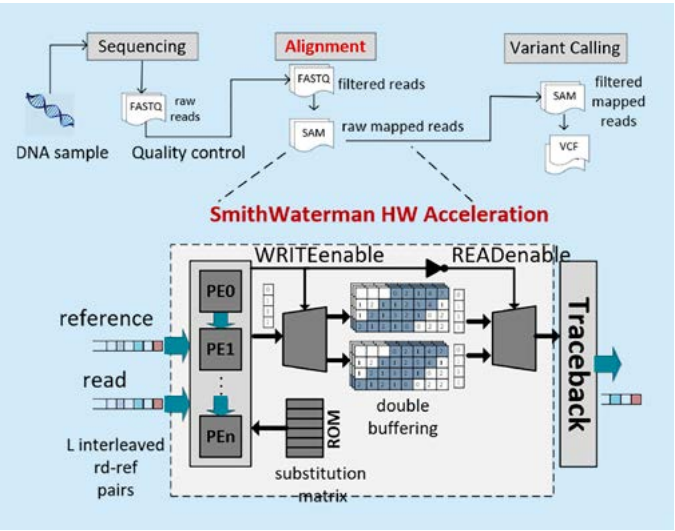
With their huge datasets and hefty computational demands, genomics applications often experience performance bottlenecks. In this article, Konstantina Koliogeorgi explains how her research harnesses field-programmable gate array (FPGA) acceleration to help solve this issue.

Three-minute thesis

NAME: Konstantina Koliogeorgi
RESEARCH CENTRE: National Technical University of Athens
SUPERVISOR: Dimitrios Soudris
DATE DEFENDED: 02/05/2023
THESIS TITLE: Hardware Acceleration Techniques for Computation and Data Intensive Machine Learning and Bioinformatic Applications

Featured research: FPGA acceleration of healthcare analytics

Advancements in the healthcare domain are progressing faster than ever and the analytics are becoming increasingly demanding on modern systems. Machine learning (ML)-based and genomic analytics particularly suffer from severe performance bottlenecks due to the enormous dataset size and the computational intensity of the required algorithms. High-level synthesis (HLS) FPGA-based acceleration is often leveraged to offload bottlenecks to hardware devices for acceleration. This thesis explores efficient architectures for healthcare analytics by building powerful FPGA accelerators and addressing real integration aspects, such as system-wide communication and accelerator-call overheads.



A dataflow architecture for accelerating short-read alignment on FPGAs

The first part of the thesis targets the acceleration of an electrocardiogram (ECG) support vector machine (SVM)-based arrhythmia detection application that has real-time requirements for execution on internet-of-things (IoT) and embedded devices. To this end, the thesis proposes a two-level systematic design framework for HLS to quickly deliver an accelerated SVM classifier. The first level proposes source-code structure optimization techniques that fully reveal the data- and instruction-level parallelism of the SVM. On the second level, an extensive automated design space exploration (DSE) of the HLS directives creates a set of pruning guidelines based on memory layout. The result is a highly compact design space, only 3% of the initial space, that includes the majority of the Pareto optimal points.

The thesis then focuses on acceleration for short-read alignment, one of the major bottlenecks of a typical genomic pipeline. We first present a novel high-throughput dataflow FPGA accelerator for the alignment bottleneck, accompanied by radical software restructuring to an open-source aligner to minimize data transfer and accelerator-call overheads during integration. We then leverage pre-alignment filters and a data-driven approach to generate an architecture of multiple input-specific, heterogeneous and resource-efficient accelerators that achieves greater speedup without losing accuracy. The multi-dataflow heterogeneous system delivers average speedups of x2 over state-of-art multi-accelerator FPGA solutions.

Konstantina's PhD supervisor, **Dimitrios Soudris**, commented: 'Konstantina's thesis has introduced our team to the field of genomics and the design of architectures optimized for healthcare applications and genomic analysis. The promising results and the need for optimized execution of such workflows encourages us to leverage techniques from the HPC world, computer architecture and VLSI to contribute to novel and efficient solutions. A very promising research path that combines the two branches of this thesis is hardware acceleration of genomic pipelines that rely on the use of deep learning methods to detect patterns within very complex DNA data.'

Funded by the European Union

SPONSORSHIP OPPORTUNITIES

HiPEAC conference

20-22 January 2025
Barcelona

Europe's largest computing systems research event (600+ participants, 80+ companies)

Expand your business network and meet new clients

Excellent recruitment opportunity: HiPEAC Jobs Fair and STEM Student Day

Media coverage and enhanced visibility options

Tailored sponsorship plans

	PLATINUM €10,000	GOLD €5,000	SILVER from €2,500	BRONZE €1,500
Logo on HiPEAC website and communications	✓	✓	✓	✓
Conference passes	6	3	2	1
Industry session presentation	30 min	20 min	10 min	✗
Industry exhibition booth	✓	✓	✓	✓
Privileged booth location and customized options	✓	✓	✗	✗
Eligibility to sponsor specific activities	✓	✓	✗	✗
Year-round HiPEAC Jobs support	✓	✓	✗	✗
Additional promotional opportunities	✓	✓	✗	✗

sponsorship@hipeac.net

bit.ly/HiPEAC25_sponsorship