UMR IRISA

Activity Report 2025

Team PIRAT

Protection of Information and Resistance to ATtacks

Joint team with Centre Inria de l'Université de Rennes

D1 – Large Scale an Secure Systems

Université de Rennes | ENS rennes | INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES | cnrs | Inria | Inserm | CentraleSupélec | IMT Atlantique Bretagne-Pays de la Loire École Mines-Télécom | Université Bretagne Sud ubs

Project-Team PIRAT

Creation of the Project-Team: 2024 March 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

A1.2.2. – Supervision

A1.2.8. – Network security

A1.3.3. – Blockchain

A1.3.4. – Peer to peer

A2.2.1. – Static analysis

A4.1. – Threat analysis

A4.4. – Security of equipment and software

A4.9. – Security supervision

A4.9.1. – Intrusion detection

A4.9.2. – Alert correlation

A4.9.3. – Reaction to attacks

A9.1. – Knowledge

A9.2. – Machine learning

A9.6. – Decision support

A9.8. – Reasoning

A9.9. – Distributed AI, Multi-agent

A9.10. – Hybrid approaches for AI

Other research topics and application domains

B6.5. – Information systems

B9.5.1. – Computer science

B9.10. – Privacy

Contents

Project-Team PIRAT **1**

1 Team members, visitors, external collaborators **5**

2 Overall objectives **6**

3 Research program **7**

4 Application domains **7**

5 Highlights of the year **7**
5.1 Awards . . . 7

6 Latest software developments, platforms, open data **8**
6.1 Latest software developments . . . 8
6.1.1 URSID . . . 8
6.1.2 Fos-R . . . 8
6.1.3 TADAM . . . 8
6.1.4 BAGUETTE . . . 9
6.1.5 ROSCA . . . 9
6.1.6 ShareMal . . . 9
6.1.7 Androscalpel . . . 10
6.1.8 Android Class Shadowing Scanner . . . 10
6.1.9 Android of Theseus . . . 10
6.2 New platforms . . . 10
6.3 Open data . . . 11

7 New results **12**
7.1 Axis 1 : Comprehension of Attacks . . . 12
7.2 Axis 2 : Detection of Attacks . . . 13
7.3 Axis 3 : Resistance to Attacks . . . 15
7.4 Reproducibility, reusability and open data . . . 16

8 Bilateral contracts and grants with industry **17**
8.1 Bilateral contracts with industry . . . 17
8.2 Bilateral Grants with Industry . . . 18

9 Partnerships and cooperations **19**
9.1 International initiatives . . . 19
9.1.1 Inria associate team not involved in an IIL or an international program . . . 19
9.1.2 Visits to international teams . . . 20
9.2 National initiatives . . . 21

10 Dissemination **25**
10.1 Promoting scientific activities . . . 25
10.1.1 Scientific events: organisation . . . 25
10.1.2 Journal . . . 26
10.1.3 Invited talks . . . 27
10.1.4 Scientific expertise . . . 28
10.1.5 Research administration . . . 28
10.2 Teaching - Supervision - Juries - Educational and pedagogical outreach . . . 28
10.2.1 Supervision . . . 29
10.2.2 Juries . . . 30
10.2.3 Educational and pedagogical outreach . . . 31

10.3 Popularization 31
10.3.1 Participation in Live events 31

11 Scientific production **31**
11.1 Major publications 31
11.2 Publications of the year 31
11.3 Cited publications 34

1 Team members, visitors, external collaborators

Research Scientists

- Emmanuelle Anceaume [CNRS, Senior Researcher, HDR]
- Pierre-Francois Gimenez [INRIA, ISFP]
- Yufei Han [INRIA, Senior Researcher]
- Michel Hurfin [INRIA, Researcher, HDR]
- Ludovic Me [INRIA, Senior Researcher, until Apr 2025, HDR]

Faculty Members

- Valerie Viet Triem Tong [Team leader, CENTRALESUPELEC, Professor, HDR]
- Jean-François Lalande [CENTRALESUPELEC, Professor, HDR]
- Loic Miller [CENTRALESUPELEC, Associate Professor, from Oct 2025]
- Samuel Pelissier [CENTRALESUPELEC, Associate Professor, from Oct 2025]

Post-Doctoral Fellows

- Anatolii Khalin [CENTRALESUPELEC, Post-Doctoral Fellow, until Oct 2025]
- Omnia Mohamed [CENTRALESUPELEC, Post-Doctoral Fellow, until Oct 2025]
- Fabien Pesquerel [INRIA, Post-Doctoral Fellow, until Mar 2025]

PhD Students

- Lucas Aubard [INRIA]
- Bassirou Badiane [INRIA, from Dec 2025]
- Antoine Cellier [INRIA, from Oct 2025]
- Solene Delourme [ORANGE, CIFRE, from Mar 2025]
- Fanny Dijoud [INRIA]
- Lucas Giordani [CENTRALESUPELEC, from Feb 2025]
- Jean Haurogne [ORANGE, CIFRE, from Nov 2025]
- Sebastien Kilian [CENTRALESUPELEC]
- Pierre Lledo [DGA-MI]
- Jean-Marie Mineau [CENTRALESUPELEC, until Oct 2025]
- Matthieu Mouzaoui [INRIA]
- Manuel Poisson [AMOSSYS, CIFRE]
- Vincent Raulin [CENTRALESUPELEC, ATER, until Sep 2025]
- Natan Talon [HACKUITY, CIFRE, until Mar 2025]
- Patrick Zounon [INRIA]

Technical Staff

- SANCHEZ Alexandre [INRIA, Engineer]
- Dorian Bachelot [CENTRALESUPELEC, Engineer, from Mar 2025]
- Malak Bazine [CENTRALESUPELEC, Engineer, from Nov 2025]
- Pol Jaouen [INRIA, Engineer, from Nov 2025]
- Yohann Rio [CENTRALESUPELEC, Engineer]
- Guillaume Vachez [CENTRALESUPELEC, from Sep 2025]
- Guillaume Vachez [CENTRALESUPELEC, from Feb 2025 until Jul 2025]

Interns and Apprentices

- Bassirou Badiane [CENTRALESUPELEC, Intern, from Mar 2025 until Sep 2025]
- Arthur Baudy [CENTRALESUPELEC, Intern, from Jun 2025 until Jul 2025]
- Malak Bazine [CENTRALESUPELEC, Intern, from Mar 2025 until Jul 2025]
- Loric Cantinat [CENTRALESUPELEC, Intern, from Jun 2025 until Jul 2025]
- Romain Debreu [CENTRALESUPELEC, Intern, from May 2025 until Jun 2025]
- Gaby Le Bideau [CENTRALESUPELEC, Intern, from Apr 2025 until Jul 2025]

Administrative Assistant

- Amandine Seigneur [INRIA]

External Collaborators

- Frederic Majorczyk [DGA, from Mar 2025]
- Frederic Majorczyk [DGA, until Feb 2025]

2 Overall objectives

The PIRAT team is interested in attack campaigns pursued by an attacker with the aim of perpetrating a malicious action on a specific element in a heterogeneous information system.

The team has three primary objectives in their research efforts:

- **Comprehension**: Understanding the scope and mechanisms of attacks by analyzing artifacts, logs, and malware.
- **Detection**: Developing advanced techniques to identify attacks as early as possible, using real-time observations like logs and network traffic.
- **Resistance** : Enhancing system resilience against attacks

The PIRAT team integrates expertise across programming, networks, AI, and distributed systems to propose solutions for the evolving cybersecurity landscape.

3 Research program

The research program of the PIRAT team revolves around the three main lines of research mentioned above:

Comprehension of Attacks

This line of research focuses on understanding the mechanisms and scope of cyberattacks. The objectives include:

- **Data Collection**: Acquiring up-to-date and representative attack data from sources like honeypots and attack-defense exercises.
- **Attack Modeling**: Developing tools and methods to build clear and scalable representations of complex attack scenarios. These models aim to assist experts in analyzing compromised systems, including large-scale infrastructures.

Detection of Attacks The PIRAT team aims to improve the detection of attacks by addressing current limitations. Our goals include:

- **Distributed Detection** : Designing collaborative systems that combine local analysis with global models, leveraging federated learning and fault-tolerant mechanisms to ensure privacy and robustness.
- **AI-Driven Detection**: Creating adaptable intrusion detection systems (IDS) powered by artificial intelligence, with a focus on explainable AI (XAI) and human-in-the-loop approaches to enhance operational efficiency and reduce false positives.

Resistance to Attacks Our goal is to improve the resilience of systems against attacks and ensure rapid recovery. Key objectives include:

- **Automatic Cyber Ranges**: Developing automated platforms for training and evaluating defensive tools.
- **Hardening Security Tools**: Strengthening AI-driven threat detection and providing irrefutable evidence for undesirable behaviors.
- **Containment Strategies** : Implementing mechanisms to isolate and mitigate the impact of attacks, restoring normal operations quickly.

Technological Development and Transfer The team emphasizes the development of reproducible tools and datasets:

- **Open-source publication** of mature tools and data for the academic community.
- **Industry collaborations** to create practical, transferable solutions.
- **Training programs and public engagement** to disseminate our research and tools.

4 Application domains

The PIRAT team's application domains focus on the practical use of their research findings in cybersecurity. These include: Infrastructure Protection, Intrusion Detection Systems including AI-driven, Security in AI and Distributed Systems, Malware Analysis and Mitigation, Education an Cyber training.

5 Highlights of the year

5.1 Awards

Emmanuelle Anceaume is the winner of the 2025 Senior Researcher Award from the GDR RSD (Distributed Networks and Systems).

Emmanuelle Anceaume and her colleagues Nicoló Rivetti, Leonardo Querzoni, Yann Busnel and Bruno Sericola have received the "Test of Time Award" of the 19th ACM International Conference on Distributed and Event-based Systems (DEBS) for their paper entitled "Efficient key grouping for near-optimal load balancing in stream processing systems".

6 Latest software developments, platforms, open data

6.1 Latest software developments

6.1.1 URSID

Keywords: Cybersecurity, Cyber attack, Virtual Machine, Cyber Range

Functional Description: URSID makes it possible to deploy multiple variants of vulnerable virtual architectures from a single attack scenario description. These architectures can be used to train security teams or students, or as a honeypot for learning and analyzing attack techniques used in the field.

URL: https://gitlab.inria.fr/pirat-public/ursid

Contact: Valerie Viet Triem Tong

6.1.2 Fos-R

Name: Forger of Security Records

Keywords: Generative Models, Cybersecurity

Functional Description: Fos-R is an AI-based synthetic network traffic generation tool. It generates high-quality data at high throughput without requiring specific hardware such as a GPU. Fos-R can also inject synthetic network traffic, for example to make a honeynet more realistic or to increase the complexity of offensive or defensive security exercises.

Release Contributions: First public version

URL: https://fosr.inria.fr

Publications: hal-04886774, hal-04871198, hal-04871298

Contact: Pierre-Francois Gimenez

Participant: Pierre-Francois Gimenez

6.1.3 TADAM

Name: TADAM

Keywords: Timed automata, Machine learning

Functional Description: Timed Automata (TA) are formal models capable of representing regular languages with timing constraints, making them well-suited for modeling systems where behavior is driven by events occurring over time. Most existing work on TA learning relies on active learning, where access to a teacher is assumed to answer membership queries and provide counterexamples. While this framework offers strong theoretical guarantees, it is impractical for many real-world applications where such a teacher is unavailable. In contrast, passive learning approaches aim to infer TA solely from sequences accepted by the target automaton. However, current methods struggle to handle noise in the data, such as symbol omissions, insertions, or permutations, which often result in excessively large and inaccurate automata. TADAM is a novel approach that leverages the Minimum Description Length (MDL) principle to balance model complexity and data fit, allowing it to distinguish between meaningful patterns and noise. TADAM is significantly more robust to noisy data than existing techniques, less prone to overfitting, and produces concise models that can be manually audited.

Release Contributions: First public version.

URL: https://github.com/Fos-R/TADAM

Publication: hal-04886774

Contact: Pierre-Francois Gimenez

Partner: CISPA Helmholtz Center for Information Security

6.1.4 BAGUETTE

Keywords: Malware, Data mining

Functional Description: Malware analysis consists of studying a sample of suspicious code to understand it and producing a representation or explanation of this code that can be used by a human expert or a clustering/classification/detection tool. The analysis can be static (only the code is studied) or dynamic (only the interaction between the code and its host during one or more executions is studied). The quality of the interpretation of a code and its later detection depends on the quality of the information contained in this representation. To date, many analyses produce voluminous reports that are difficult to handle quickly. BAGUETTE is a graph-based representation of the interactions of a sample and the resources offered by the host system during one execution. BAGUETTE helps automatically search for specific behaviors in a malware database and efficiently assists the expert in analyzing samples.

URL: https://gitlab.inria.fr/vraulin/baguette-verse

Contact: Vincent Raulin

6.1.5 ROSCA

Name: Robust and Scalable Correlation of Alerts

Keywords: Alert correlation, Intrusion Detection Systems (IDS), Anomaly detection

Functional Description: ROSCA is a transparent solution that can handle a large volume of security alerts to reduce analyst fatigue, delayed responses, and missed attacks. It is suited for in-house adaptation or re-implementation by Security Operations Centers (SOCs). It is grounded in the MITRE ATT&CK kill chain model, and automatically aggregates and correlates alerts based on their shared attributes, enabling the construction of contextualised attack cases. Each case is assigned a score reflecting its threat level, before being presented to analysts within a prioritised queue. Our method handles multi-stage attack patterns and supports rapid processing through a robust, noise-tolerant scoring mechanism designed for interpretability and operational integration. We validate the effectiveness of ROSCA on real-world alert data and compare it against the MATE framework, demonstrating superior prioritisation accuracy and more reliable identification of critical alerts.

URL: https://gitlab.inria.fr/resist/watch-rosca-alert-prioritisation/

Publication: hal-05351162

Contact: Remi Garcia

Participants: Remi Garcia, Abdelkader Lahmadi, Pierre-Francois Gimenez, an anonymous participant

6.1.6 ShareMal

Keywords: Cybersecurity, Dynamic Analysis, Malware

Functional Description: Malware analysis requires special precautions to avoid infecting analysis environments and networks. Sharing and storing malware is also a significant issue due to the critical nature of this field and the large amount of data involved. ShareMal aims to provide a comprehensive platform that serves as a one-stop shop for all matters related to malware (storage, processing, dataset creation, analysis service experimentation, etc.). It provides access to information held on each piece of malware and allows analysis services to be performed without having to handle the samples directly. The platform consists of a website, an extensible and scalable analysis pipeline, a Python library for programmatic interaction with the API, and a unique deployment and scaling solution.

URL: https://www.lhs-rennes.fr/platforms/sharemal/

Contact: Dorian Bachelot

Participants: Dorian Bachelot, Alexandre Sanchez

6.1.7 Androscalpel

Keywords: Android, Bytecode

Functional Description: Androscalpel is a Rust program that helps to manipulate Android application bytecode (dalvik).

URL: https://gitlab.inria.fr/pirat-public/android/androscalpel

Contact: Jean-François Lalande

6.1.8 Android Class Shadowing Scanner

Keyword: Android

Functional Description: Android Class Shadowing Scanner detects if an Android application is in a situation that may lead to class spoofing.

URL: https://gitlab.inria.fr/pirat-public/android/android_class_shadowing_scanner

Contact: Jean-François Lalande

6.1.9 Android of Theseus

Keyword: Android

Functional Description: Android of Theseus is an implementation of the method presented in chapter 5 of the thesis 'The Woes of Android Reverse Engineering: from Large Scale Analysis to Dynamic Deobfuscation', by Jean-Marie Mineau. The idea is collecting dynamic data like reflection calls and dynamic code loading using Frida, then patch the application to include this data statically. The application can then be analyse with any static analysis tools taking an application as input.

URL: https://gitlab.inria.fr/pirat-public/android/android-of-theseus

Contact: Jean-François Lalande

6.2 New platforms

Smart and Secure Room platform

Participants: Jean-François Lalande, Anatolii Khalin.

The Smart and Secure Room platform is one of the platforms of the FIRRST (Federation of Research Infrastructures in SecuriTy of Rennes). It is intended to explore attacks targeting IoT environments and energy management in smart buildings. In 2024, Anatolii Khalin worked on detecting cyberattacks that could target such a physical system. In particular, smart buildings taking autonomous decisions about energy production and consumption could be the target of an attacker. We have conducted a campaign for simulating an attacker stealing energy when the smart room is producing and consuming solar energy. We propose new methods to detect such attacks and we explore attacks that may be undetected [15].

Poneypot platform

Participants: Dorian Bachelot, Bassirou Badiane, Yufei Han, Yohann Morel, Alexandre Sanchez, Guillaume Vachez, Valerie Viet Triem Tong.

The Poneypot platform is a multi-instance, high-interaction honeypot designed for collecting network and system traffic generated by malicious actors' activity. The project aims to collect a recent, real dataset (network and system) to better understand the behavior of attackers. It relies on HopLab's infrastructure, hosted at the LHS of Rennes, and uses URSID to automate the deployment of each honeypot. PoneyPot also integrates a centralized monitoring solution and an archive mechanism to save infected virtual machines for post-mortem forensic analysis. In December 2025, a new experiment was launched to assess a critical vulnerability in the React web framework and demonstrate the platform's maturity through the volume of generated data and the time required to deploy the honeypot. PoneyPot is also connected to other platforms, such as ShareMal, to store and analyse retrieved payloads from infected honeypots.

6.3 Open data

We released open source datasets with our publications:

- For the paper [15], we released the dataset containing the energy consumption of the system under attack: https://zenodo.org/records/15297511.
- For the paper [4], we released the APK list and the analysis of the artifacts that we detected in these APKs: http://dx.doi.org/10.5281/zenodo.15846481.
- For the paper [16], we released the network flow and system logs of the data collected during the BreizhCTF event of 2024: https://casinolimit.inria.fr/, https://doi.org/10.5281/zenodo.15278062. Additionally, we released the source code of the challenge and the tool for manipulating the dataset. The release of this dataset follows the recommendations of Crowder et al. about the reuse practices when publishing a cybersecurity dataset [32].
- For the paper [17], we released the network flow and system logs of a one month experiment containing a stealth attack: https://dedale.inria.fr.
- For the paper [21], we released a new dataset for evaluation unsupervised intrusion detection system on SQL attacks, and most notably SQL injection attacks: https://zenodo.org/records/15744477.
- The DARPA OpTC dataset is an interesting dataset to design and evaluate intrusion detection systems (IDS), especially IDS based on machine learning. However, we discover some mismatches affecting unique identifiers in the dataset. Moreover, there is no official precise labeling of the dataset. We provide a new corrected version of the dataset and a labeling of the dataset at the host and network levels in https://correctedoptc.inria.fr/

7 New results

7.1 Axis 1 : Comprehension of Attacks

Participants: Emmanuelle Anceaume, Lucas Aubard, Dorian Bachelot, Bassirou Badiane, Yufei Han, Jean-François Lalande, Jean-Marie Mineau, Guillaume Vachez, Valerie Viet Triem Tong, Pierre-Francois Gimenez.

Unveiling Ethical Risks in Blockchain Layer 2 Ecosystems The blockchain ecosystem has long been praised for its ability to promote decentralization, transparency, and financial autonomy. However, it has also become a fertile ground for ethical concerns, particularly regarding the systematic deception and exploitation of users due to severe information asymmetries. That is to say, many users engage with blockchain applications, without fully grasping the underlying risks and trade-offs. A good illustration of this issue can be seen in the complex landscape of blockchain Layer 2 (L2) scaling solutions. Many (L2) protocols differ significantly in their design, security guarantees, and governance models. Users may assume that all L2 solutions ofer similar levels of decentralization and trustlessness, while in reality, some require substantial trust in centralized operators. This lack of clarity leads to uninformed decision-making and, in many cases, financial loss. At the core of this problem lies information asymmetry, a well-documented economic phenomenon in which one party in a transaction has access to significantly more or better information than another. Empirical Layer 2 rollups improve throughput and fees, but can reintroduce risk through operator discretion and information asymmetry. In [3]([29]), we ask which operator and governance designs produce ethically problematic user risk. We adapt Ethical Risk Analysis to rollup architectures, build a role-based taxonomy of decision authority and exposure, and pair the framework with two empirical signals, a cross sectional snapshot of 129 projects from L2BEAT and a hand curated incident set covering 2022 to 2025. We analyze mechanisms that affect risks to users' funds, including upgrade timing and exit windows, proposer liveness and whitelisting, forced inclusion usability, and data availability choices. We find that ethical hazards rooted in L2 components control arrangements are widespread: instant upgrades without exit windows appear in about 86 percent of projects, and proposer controls that can freeze withdrawals in about 50 percent. Reported incidents concentrate in sequencer liveness and inclusion, consistent with these dependencies. We translate these findings into ethically grounded suggestions on mitigation strategies including technical components and governance mechanisms.

Defending One-Party Hijacking Attacks in Vertical Federated Learning Vertical Federated Learning (VFL) is susceptible to various one-party hijacking attacks, such as Replay and Generation attacks, where a single malicious client can manipulate the model to produce attacker-specified results, thereby compromising its reliability in real-world deployments. In [5], we first uncover the underlying mechanisms of these attacks and observe that successful attacks induce significant discrepancies in the embedding-label associations across different clients. We establish a theoretical framework demonstrating how these discrepancies can serve as reliable indicators for detecting hijacking attempts. Building upon this insight, we propose VFLMonitor, a robust defense mechanism that leverages these embedding-label discrepancies to detect and mitigate hijacking attacks. Specifically, VFLMonitor identifies suspicious queries by analyzing differences in label estimations from multiple clients and applies a majority voting rule to correct or filter out these malicious queries. Moreover, VFLMonitor introduces a novel regularization strategy during training to reduce intra-class variance in embeddings, thereby enhancing their discriminative power and improving defense effectiveness. Extensive experiments were conducted on 5 real-world datasets against 2 different attack types under 3 attack scenarios. The results demonstrate that VFLMonitor can effectively identify and exclude potential hijacked requests in all types of one-party hijacking attacks, while maintaining a meager false positive rate for legitimate queries.

Class loaders in the middle: confusing Android static analyzers. When executing a mobile application, Android executes either the classes provided by the developer or the ones provided by the operating system. The dynamic linking and loading of the different classes is a complex task that may be exploited by an attacker. In particular, if the developer adds a class whose name collides with another class of Android, they

may confuse a reverse engineer. In [4], we explore the possible collisions that can occur between classes defined multiple times at different locations, i.e., multiple times in the APK file or, at the same time, in the APK and the operating system. We highlight three attacks that we call shadow attacks. In particular, we show that static analysis tools used by a reverse engineer choose the shadow implementation for most of the evaluated tools, and output a wrong result. In particular, the flow analysis of Androguard or Flowdroid can be fooled by an attacker. In a dataset of 49 975 applications, we also explored if shadow attacks are used in the wild and found that most of the time, there is no malicious behavior behind them. The main results are that 23.52 % of applications shadow a class of the SDK and 3.11 % a hidden class of the system.

Benign Traffic Comprehension Timed Automata (TA) are formal models capable of representing regular languages with timing constraints, making them well-suited for modeling systems where behavior is driven by events occurring over time. Most existing work on TA learning relies on active learning, where access to a teacher is assumed to answer membership queries and provide counterexamples. While this framework offers strong theoretical guarantees, it is impractical for many real-world applications where such a teacher is unavailable. In contrast, passive learning approaches aim to infer TA solely from sequences accepted by the target automaton. However, current methods struggle to handle noise in the data, such as symbol omissions, insertions, or permutations, often resulting in excessively large and inaccurate automata. In [11], we introduce TADAM, a novel approach that leverages the Minimum Description Length (MDL) principle to balance model complexity and data fit, allowing it to distinguish between meaningful patterns and noise. We show that TADAM is significantly more robust to noisy data than existing techniques, less prone to overfitting, and produces concise models that can be manually audited. We further demonstrate its practical utility through experiments on real-world tasks, such as network flow classification and anomaly detection.

Botnet Attack Analysis Botnets are large-scale networks of compromised devices that enable attackers to launch coordinated cyberattacks such as DDoS, credential theft, cryptojacking, and malware propagation. Their rapid propagation and stealth techniques make early detection and timely response particularly challenging. Cyber Threat Intelligence (CTI) is essential for mitigating such threats, but its production is still predominantly manual, requiring analysts to interpret raw logs and this process is too slow, resource-intensive, and difficult to scale against automated botnets. In [9], we propose a novel approach to automate botnet CTI generation directly from honeypot-captured intrusions. We rely on highinteraction honeypots to capture various botnet samples. The collected data is then analyzed using large language models (LLMs), guided by structured prompts constructed from previously observed botnet actions mapped to the MITRE ATT&CK framework. We first perform manual analyses of real botnet sessions to construct structured datasets of tactics, techniques, and procedures (TTPs) for each botnet intrusion. These resources are then used for prompt engineering, enabling LLMs to transform raw system and network logs into structured CTI reports through in-context learning (ICL). Preliminary results demonstrate that LLMs can generate coherent and actionable reports, which can help in understanding the operating modes of botnets and in developing effective countermeasures.

7.2 Axis 2 : Detection of Attacks

Participants: Jean-François Lalande, Anatolii Khalin, Yufei Han, Pierre-Francois Gimenez, Fanny Dijoud, Michel Hurfin, Frédéric Majorczyk, Matthieu Mouzaoui, Patrick Zounon
.

Privacy-Preserving Intrusion Detection In distributed networks, participants often face diverse and fast-evolving cyberattacks. This makes techniques based on Federated Learning (FL) a promising mitigation strategy. By only exchanging model updates, FL participants can collaboratively build detection models without revealing sensitive information, e.g., network structures or security postures. However, the effectiveness of FL solutions is often hindered by significant data heterogeneity, as attack patterns often differ drastically across organizations due to varying security policies. To address these challenges, in [10], we introduce PROTEAN, a Prototype Learning-based framework geared to facilitate collaborative

and privacy-preserving intrusion detection. PROTEAN enables accurate detection in environments with highly non-IID attack distributions and promotes direct knowledge sharing by exchanging class prototypes of different attack types among participants. This allows organizations to better understand attack techniques not present in their data collections. We instantiate PROTEAN on two cyber intrusion datasets collected from IIoT and 5G-connected participants and evaluate its performance in terms of utility and privacy, demonstrating its effectiveness in addressing data heterogeneity while improving cyber attack understanding in federated intrusion detection systems (IDSs).

Detecting Energy Theft Attacks With the rapid development of charging infrastructure for Electric Vehicles, the risks of cyber-physical attacks, including energy theft are growing. The attack detection results of energy theft are usually validated on real open access data of charging sessions, however, the attacks themselves are artificially introduced. To address this issue, we present in [15] a three-week experiment on a real testbed including the production and consumption of energy with realistic energy theft attacks occurring in the system. The energy setup emulates a charging bike station where users can charge bikes at different levels of state of charge and at different durations of charging sessions. The attacker is one of the users who steals energy from the system for its own bike and can override the reported consumption of power. We propose a method for detecting such attacks based on the total production/consumption power balance. The full dataset of the three-week experiment is published with this work for reproducibility purposes.

Security Alert Correlation In large organisations and complex infrastructures, the overwhelming volume of security alerts often results in analyst fatigue, delayed responses, and missed attacks. Security Operations Centers (SOCs) typically rely on black-box commercial solutions, offering limited transparency into their alert classification mechanisms and lacking the flexibility for in-house adaptation or re-implementation. To address these limitations and improve situational awareness, in [12], we propose ROSCA, an efficient alert prioritisation method grounded in the MITRE ATT&CK kill chain model. The proposed approach automatically aggregates and correlates alerts based on their shared attributes, enabling the construction of contextualised cases. Each case is assigned a score reflecting its threat level, before being presented to analysts within a prioritised queue. Our method handles multi-stage attack patterns and supports rapid processing through a robust, noise-tolerant scoring mechanism designed for interpretability and operational integration. We validate the effectiveness of ROSCA on real-world alert data and compare it against the MATE framework, demonstrating superior prioritisation accuracy and more reliable identification of critical alerts.

Overlapping IPv4, IPv6, and TCP data IPv4, IPv6, and TCP have a common mechanism allowing one to split an original data packet into several chunks. Such chunked packets may have overlapping data portions and, OS network stack implementations may reassemble these overlaps differently. A Network Intrusion Detection System (NIDS) that tries to reassemble a given flow data has to use the same reassembly policy as the monitored host OS; otherwise, the NIDS or the host may be subject to attack.

In [7], we provide several contributions that enable us to analyze NIDS resistance to overlapping data chunks-based attacks. First, we extend state-of-the-art insertion and evasion attack characterizations to address their limitations in an overlap-based context. Second, we propose a new way to model overlap types using Allen's interval algebra, a spatio-temporal reasoning. This new modeling allows us to formalize overlap test cases, which ensures exhaustiveness in overlap coverage and eases the reasoning about and use of reassembly policies. Third, we analyze the reassembly behavior of several OSes and NIDSes when processing the modeled overlap test cases. We show that 1) OS reassembly policies evolve over time and 2) all the tested NIDSes are (still) vulnerable to overlap-based evasion and insertion attacks.

In [8], we propose PYROLYSE, an audit tool that exhaustively tests and describes the reassembly policies of various IP and TCP implementation types. This tool ensures that implementations reassemble overlapping chunk sequences without errors. The second contribution is the analysis of PYROLYSE artifacts. We first show that the reassembly policies are much more diverse than previously thought. Indeed, by testing all the overlap possibilities for $n \leq 3$ test case chunks and different testing scenarios, we observe 15 different behaviors out of 23 tested implementations depending on the protocol. Second, we report eight errors impacting one OS, two NIDSes, and two embedded stacks, which can lead to security issues such as NIDS pattern-matching bypass or DoS attacks. A CVE was assigned to a NIDS error. Finally, we show that

implemented IP and TCP policies obtained through chunk pair testing are usually inconsistent with the observed triplet reassemblies. Therefore, contrary to what they currently do, NIDSes or other network traffic analysis tools should not apply $n = 2$ pair policies when the number of overlapping chunks exceeds two.

The DARPA OpTC Dataset In [18], we address the challenges of using the DARPA OpTC dataset for intrusion detection system (IDS) research. While the dataset offers a rich combination of network and host data, its usability is hindered by a lack of an official event labeling and errors in logs data. We propose a two-fold solution: first, we identify and correct errors in the dataset, and second, we design and implement a comprehensive labeling methodology for attack-related events at both the network and host levels. Our corrected dataset, along with the labeling scripts, has been made publicly available to support reproducibility and further research. Additionally, we assess the impact of these labels and corrections on the effectiveness of graph-based machine learning IDS methods.

From CTI Reports to Cyber Security Knowledge Graphs In [23], we propose a pipeline to automatically transform CTI reports (*i.e.*, text documents) into a graph-based Cyber Security Knowledge Graph (CSKG) using Large Language Models (LLMs) for security entity recognition.

Adversarial Attacks against NIDS In the context of Network Intrusion Detection Systems (NIDS), we study adversarial attacks which consist of manipulating the network traffic to degrade IDS performance by either performing evasion or inducing alarm fatigue. More precisely, we study the vulnerability of Graph Neural Networks (GNN)-based NIDS under problem-space constraints, where attackers can only create new communications towards specific IP addresses, while ensuring consistency with network protocols.

7.3 Axis 3 : Resistance to Attacks

Participants: Emmanuelle Anceaume, Pierre-Francois Gimenez, Yufei Han, Jean-François Lalande, Ludovic Mé, Jean-Marie Mineau, Manuel Poisson, Natan Talon, Valerie Viet Triem Tong
.

Secure Representation of a PoW-based Blockchain In [19]([30]), we present the first non-interactive, succinct, and secure representation of a PoW-based blockchain that operates under variable mining difficulty while satisfying both completeness and onlineness properties. Completeness ensures that provers can update an existing NIPoPoW by incorporating a newly mined block, whereas onlineness ensures that miners can extend the chain directly from a NIPoPoW. The time complexity for both the prover (to update a NIPoPoW with a new block) and the verifier is logarithmic in the number of blocks of the underlying PoW blockchain. The communication complexity required for synchronization is polylogarithmic in the length of the blockchain. We prove the correctness of our scheme in the presence of a 1/3-bounded PPT adversary.

From Risk to Resilience Data Reconstruction Attacks (DRA) pose a significant threat to Federated Learning (FL) systems by enabling adversaries to infer sensitive training data from local clients. Despite extensive research, the question of how to characterize and assess the risk of DRAs in FL systems remains unresolved due to the lack of a theoretically-grounded risk quantification framework. In [22], we address this gap by introducing Invertibility Loss (InvLoss) to quantify the maximum achievable effectiveness of DRAs for a given data instance and FL model. We derive a tight and computable upper bound for InvLoss and explore its implications from three perspectives. First, we show that DRA risk is governed by the spectral properties of the Jacobian matrix of exchanged model updates or feature embeddings, providing a unified explanation for the effectiveness of defense methods. Second, we develop InvRE, an InvLoss-based DRA risk estimator that offers attack method-agnostic, comprehensive risk evaluation across data instances and model architectures. Third, we propose two adaptive noise perturbation defenses that enhance FL privacy without harming classification accuracy. Extensive experiments on real-world datasets validate our framework, demonstrating its potential for systematic DRA risk evaluation and mitigation in FL systems.

Robust Malware Detectors Malware analysis involves analyzing suspicious software to detect malicious payloads. Static malware analysis, which does not require software execution, relies increasingly on machine learning techniques to achieve scalability. Although such techniques obtain very high detection accuracy, they can be easily evaded with adversarial examples where a few modifications of the sample can dupe the detector without modifying the behavior of the software. Unlike other domains, such as computer vision, creating an adversarial example of malware without altering its functionality requires specific transformations. In [14], We propose a new model architecture for certifiably robust malware detection by design. In addition, we show that every robust detector can be decomposed into a specific structure, which can be applied to learn empirically robust malware detectors, even on fragile features. Our framework ERDALT is based on this structure. We compare and validate these approaches with machine-learning-based malware detection methods, allowing for robust detection with limited reduction of detection performance.

Resistance Against Injection-based Attacks Many systems are controlled via commands built upon user inputs. For systems that deal with structured commands, such as SQL queries, XML documents, or network messages, such commands are generally constructed in a "fill-in-the-blank" fashion: the user input is concatenated with a fixed part written by the developer (the template). However, the user input can be crafted to modify the command's semantics intended by the developer and lead to the system's malicious usages. Such an attack, called an injection-based attack, is considered one of the most severe threat to web applications. Solutions to prevent such vulnerabilities exist but are generally ad hoc and rely on the developer's expertise and diligence. In [2, 6] we addresse these vulnerabilities from the formal language theory's point of view. We formally define two new security properties. The first one, "intent-equivalence", guarantees that a developer's template cannot lead to malicious injections. The second one, "intent-security", guarantees that every possible template is intent-equivalent, and therefore that the programming language itself is secure. We thoroughly analyze the decidability of these properties for the most common grammar classes. We conclude by highlighting the technical implications of these results for various settings and tools.

7.4 Reproducibility, reusability and open data

Participants: Pierre-Francois Gimenez, Maxime Lanvin, Jean-François Lalande, Jean-Marie Mineau, Ludovic Mé, Manuel Poisson.

Reproducibility and reusability in computer science experiments become a requirement for research works. Reproducibility ensures that results can be confirmed by using the same dataset and software of previous papers. Reusability helps other researchers to build new approaches with distributed software artifacts.

CasinoLimit: An Offensive Dataset Labeled with MITRE ATT&CK Techniques Cybersecurity exercises are a common way to train and evaluate the skills of cybersecurity professionals. These exercises also provide a unique opportunity to generate datasets with realistic attack traces on non-sensitive systems. Nevertheless, the collected logs are unlabeled, and deciding which logs are related to pentesters is a difficult problem. In [16], we present a novel methodology to label efficiently both system and network logs using MITRE ATT&CK techniques. To demonstrate the effectiveness of our approach, we introduce CasinoLimit, a dataset generated from a pentest exercise that has been played by 114 participants where we collected 540 GB of attack data. We apply our methodology to accurately label these logs with a semi-automatic approach: labels are inferred from the shell sessions and propagated to the network sessions, and eventually corrected by a junior analyst. An expert analyst has manually reviewed all the labels that have been computed to ensure the quality of the labeling process. The results of the pentest exercise are deeply discussed. We show the variability of players' behaviors and that players can be distinguished by their command line habits. In addition, the high level of granularity of labels coupled with the number of participants enables multiple other applications. With this paper, we release the full dataset and the associated labeling tool, Manatee, which can be used to browse the logs and labels. To support the generalization of our approach, we made it possible to load other datasets with this.

DEDALE: a New Dataset and Testbed for Evaluating the Detection of APT attacks among Network and System Logs High-quality data is required to design and evaluate Intrusion Detection Systems (IDSes). However, the currently publicly available datasets are often considered unrealistic and unrepresentative of advanced attacks. Moreover, many errors were recently identified in some network intrusion detection datasets. Even well-known and widely used datasets such as CICIDS2017 have recently been criticized for their poor quality. Another issue with those datasets is that they tend to become quickly outdated. In [17], we propose and share a new testbed named RESCOUSSE and a new dataset named DEDALE. RESCOUSSE can be used to generate new datasets; it is based on an existing testbed that we improved deeply. In particular, we fix the errors and biases we identified in the currently available datasets. DEDALE is a new dataset generated with RESCOUSSE. The dataset lasts one month: the first two weeks contain only benign activities, a discrete APT attack is spread over eight days from the third week, and the remaining six days are free of attacks and can be used to evaluate the false positive rate. The dataset contains both network and system logs, which allows the design and evaluation of Host and Network IDSes, as well as correlation techniques. We label both the system and network logs and check that no obvious biases exist in the network logs. Since RESCOUSSE is open source, it allows other researchers to replicate DEDALE and correct some errors that they may find if needed.

Superviz25-SQL: high-quality dataset to empower unsupervised SQL injection detection systems The digitalization of public and private services has led to more sophisticated and serious cybersecurity threats. Among them, SQL injection attacks leverage user inputs to remotely execute malicious actions on a database, such as data exfiltration and deletion, or privilege escalation. They are regularly classified as one of the most prominent threats to web services. Intrusion detection systems are widely used to detect such injection attacks and react to them, but it is difficult to assess their actual effectiveness and compare them because of a lack of high-quality datasets. Current SQL injection detection datasets lack diversity, are poorly documented, and the generated samples are not representative of real-world infrastructures. In [21], we present a new dataset Superviz25-SQ , whose design is structured around four quality dimensions: realism, diversity, benchmarking capabilities and the presence of good documentation. We examine the dataset diversity using lexical, syntactic and semantic metrics, and demonstrate that its size is sufficient to evaluate data-intensive detectors. Finally, we provide nine classical and state-of-the art SQL injection detection pipelines as baselines for future works.

Synthetic Network Traffic Generation Network data can be difficult to collect due to privacy and confidentiality reasons. For these reasons, network datasets are typically created with controlled environments called testbeds. However, these datasets are regularly criticized for their limited size, class imbalance, obsolescence, and lack of actual user activity. Following the rapid development of generative artificial intelligence, new methods have been applied to synthetic network traffic generation without emulation or simulation. The systematic literature review presented in [13] assesses the current state of synthetic network traffic generation for intrusion detection systems.

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

DGA

Participants: Pierre-Francois Gimenez, Yufei Han, Vincent Raulin, Valerie Viet Triem Tong, Alexandre Sanchez.

Vincent Raulin's PhD focuses on using machine learning approaches to boost malware detection and classification based on dynamic analysis traces by extracting feature representations with the knowledge of malware analysis experts. This representation aims at capturing the semantics of the program (i.e., what resources it accesses, what operations it performs on them) in a plateform-independent fashion, by replacing

the implementation particularities (system call number 2) with higher-level operation (opening a file). This representation could notably provide semantic explanation of malware activity and deliver explainable malware detection/malware family classification.

DGA

Participants: Jean-François Lalande, Pierre Lledo, Frederic Majorczyk.

Pierre Lledo is working on the evaluation of simulators of attacks. After selecting open source simulators, we built different topologies of network and we evaluated the robustness of the different simulators on these topologies. The scalability is evaluated on different parameters: the number of hosts, the number of subnets, the diversity of attacks. We also evaluated the correctness of the simulators on a fixed topology where attacker should follow a precise order in the steps of attacks.

AMOSSYS

Participants: Manuel Poisson, Gilles Guette, Valerie Viet Triem Tong.

Manuel Poisson has started a thesis in collaboration with Amossys. Manuel Poisson is interested in identifying operational attack scenarios in an information system.

Hackuity

Participants: Natan Talon, Gilles Guette, Yufei Han, Valerie Viet Triem Tong.

Natan Talon started his PhD in October 2021 in the context of a collaboration with the company Hackuity. The main objective of this thesis is to be able to assess whether an information system is likely to be vulnerable to an attack. This attack may have been observed in the past or inferred automatically from other attacks.

Orange

Participants: Jean-François Lalande, Pierre-Francois Gimenez, Valerie Viet Triem Tong.

Two phD thesis started in 2025 with Orange: Solène Delourme and Jean Haurogne. We started to explore how AI methods (RAGS, LLM, agents) can help for tasks such as the investigation of security incidents of SOCs and malware analysis.

8.2 Bilateral Grants with Industry

DGA

Participants: Fanny Dijoud, Pierre-Francois Gimenez, Michel Hurfin, Frederic Majorczyk.

The objective of Fanny Dijoud's thesis is to design an intrusion detection system to analyze system logs. The approach is based on the use of several AI models to observe anomalies in the different provenance graphs that are built. The proposed solution takes into account the fact that these graphs are dynamic and heterogeneous. This PhD has started in november 2023.

ANSSI

Participants: Lucas Aubard, Gilles Guette, Ludovic Me.

Lucas Aubard started his PhD in October 2022 in the context of a collaboration between Inria and the ANSSI. The objective of this thesis is to improve the existing knowledge on reassembly policies, to design mechanisms to automate IDS configuration and to improve the application of these policies within IDS/IPS to increase their detection capabilities in specific contexts such as cloud computing.

DGA

Participants: Antoine Cellier, Pierre-Francois Gimenez, Frederic Majorczyk, Gilles Guette.

Antoine Cellier is financed by the DGA through the Pôle d'Excellence Cyber (PEC) since October 2025. Antoine works on synthetic benign network traffic and system logs generation with AI models.

DGA

Participants: Yufei Han, Valerie Viet Triem Tong.

Helene Orsini's PhD thesis is financed by DGA since October 2021. Her thesis project focuses on adversarially robust and interpretable machine learning pipeline for network intrusion detection systems. She will study how to automate the feature engineering phase to extract informative features from non-structured, categorical and imperfect security reports / logs. Furthermore, she will investigate how to make the machine learning pipeline resilient to intentional evading techniques in network intrusion behaviors.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Inria associate team not involved in an IIL or an international program

SecGen

Participants: Pierre-Francois Gimenez.

Title: Security Data Generation

Duration: 2023 -> 2025

Coordinator: Mario Fritz (fritz@cispa.de)

Partners:

- CISPA (Allemagne)

Inria contact: Pierre-Francois Gimenez

Summary: SecGen aims at improving network traffic generation to make it usable for intrusion detection systems learning and evaluation. Indeed, existing datasets are highly criticized because of their age, their non-representativeness and the errors they contain. The CIDRE team (now PIRAT) is already involved in a thesis on network data generation, and CISPA will bring their expertise in data mining (for the identification of temporal patterns specific to network data that our experts will be able to analyze) and in anomaly detection based on deep learning.

Participants: Bassirou Badiane, Pierre-Francois Gimenez, Yohann Morel, Valerie Viet Triem Tong.

DeceptIA

Title: Deception Technologies for Honeypots with Intelligence and Adaptability

Duration: 2025 -> 2027

Coordinator: Hans Dieter Schotten

Partners:

- DFKI (Allemagne)
- University of Tokyo (Japon)
- Osaka Metropolitan University (Japon)

Inria contact: Abdelkader Lahmadi

Summary: In DeceptIA, we address the global and local cybersecurity issues and define the following three research questions. What are the characteristics of new anomalous traffic observed in large-scale honeypots deployed across multiple geolocations and services? How can we make honeypots adapt on-the-fly to the attacker's behavior and also evolve interaction between them? How can we develop an effective phishing detection system that not only accurately identifies phishing attacks, but also educates and explains the risks to end-users in a way that increases their awareness and resilience to future phishing attempts? To solve these research questions, we deploy honeypots related to various locations (e.g., Japan, France,Germany) and various services (e.g., research institutes, cloud, home, edge), and conduct experiments and analysis of anomalous traffic.

9.1.2 Visits to international teams

Research stays abroad Jean-François Lalande participated to Dagstuhl Research Meeting : Methodological Advancements in Information Hiding, Oct 15 - Oct 17, 2025.

9.2 National initiatives

PEPR CyberSecurity project: DefMal (2022-2028)

Participants: Bassirou Badiane, Dorian Bachelot, Pierre-Francois Gimenez, Yufei Han, Sebastien Kilian, Jean-François Lalande, Jean-Marie Mineau, Vincent Raulin, Valerie Viet Triem Tong.

PEPR DefMal is a collaborative ANR project involving CentraleSupélec, Rennes University, Lorraine University, Sorbonne Paris Nord University, CEA, CNRS, Inria and Eurecom. Malware is affecting government systems, critical infrastructures, businesses, and citizens alike, and regularly makes headlines in the press. Malware extorts money (ransomware), steals data (banking, medical), destroys information systems, or disrupts the operation of industrial systems. The fight against malware is a national and European security issue that requires scientific advances to design new responses and anticipate future attack methods. The aim of the project DefMal is to study malicious programs, whether they are malware, ransomware, botnet, etc. The first objective is to develop new approaches to analyze malicious programs. This objective covers the three aspects of the fight against malware: (i) Understanding (ii) Detection and (iii) Forensics. The second objective of the project is the global understanding of the malware ecosystem (modes of organization, diffusion, etc.) in an interdisciplinary approach involving all the actors concerned.

PEPR CyberSecurity project: SuperViz (2022-2028)

Participants: Pierre-Francois Gimenez, Yufei Han, Frederic Majorczyk, Ludovic Mé, Yohann Morel.

PEPR SuperviZ is a collaborative ANR project involving CentraleSupélec, Eurecom, Institut Mines-Télécom, Institut Polytechnique de Grenoble, Rennes University, Lorraine University, CEA, CNRS and Inria. The digitalization of all infrastructures makes it almost impossible today to secure all systems a priori, as it is too complex and too expensive. Supervision seeks to reinforce preventive security mechanisms and to compensate for their inadequacies. Supervision is fundamental in the general context of enterprise systems and networks, and is just as important for the security of cyber-physical systems. Indeed, with the ever growing number of connections between objects, the attack surface of systems has become frighteningly wide. This makes security even more difficult to implement. The increase in the number of components to be monitored, as well as the growing heterogeneity of the capacity of these objects in terms of communication, storage and computation, makes security supervision more complex.

PEPR CyberSecurity project: Rev (2023-2028)

Participants: Pierre-Francois Gimenez.

PEPR REV is a project about vulnerability research and exploitation. A notable characteristic of complex targets is that they can generally no longer be attacked using a single technique or exploiting a single vulnerability, due to the deployment of numerous protections. For this reason, the REV project is tackling this problematic at multiple levels by addressing all layers, hardware, software and communication interfaces (web and IoT). In this purpose, one of the project's objectives is to combine several tools and approaches simultaneously: for example, memory analysis will benefit from advances in hardware attacks, and will be used to develop exploits. This broad-spectrum analysis is fundamental today: as an illustration, hardware attacks can be combined with software attacks, software attacks can be based on weaknesses in the micro-architecture or require advanced network interactions. Moreover, the impact of attacks and exploits

nowadays goes far beyond malicious use, allowing for instance to forensically investigate complex systems such as smartphones. The question also arises from an ethical and legal point of view, and this is a major societal issue: to which extent is it possible to use these techniques, in particular for law enforcement, from an ethical or legal point of view. What is the possible use of these attacks, when should they be corrected ("responsible disclosure") or used, and in what legal framework?

ANR project: CKRISP (2024-2027)

Participants: Yufei Han, Michel Hurfin, Frederic Majorczyk, Patrick Zounon
.

CKRISP is an collaborative project led by Yufei Han in PIRAT with Eurecom, CEA, Telecom Sud-Paris. The main contribution of CKRISP address the limited coverage of attack behaviour variety in the training data as well as the lack of interpretability of AI detection models. First, we will investigate the combination of AI systems such as, e.g., Large Language Models (LLM), and human-monitored cyber security knowledge graph (CSKG) for understanding, predicting and exploring new cyberattack behaviours via Human-AI interaction. The powerful LLMs can help identify entities and predict relations between entities from cyber threat reports and low-level run-time behaviour logs. CSKG can then be built automatically based on extracted knowledge about specific attack scenarios. The attack knowledge graph can substantially help human analysts verify the AI-based attack detection results and facilitate human analysts' inspection of new attacks. Second, we will elaborate further on the prediction of attack behaviours by organizing AI-assisted reasoning with inputs from security incidents collected from various sensors, like IDS, and manual inspection results of human analysts. It will help assess the vulnerability of a target IT system and reach an initial step of AI-assisted security response based on the detected incidents. Third, we will further propose data generation methods to produce synthetic normal/attack behaviour data to enrich training data and improve the robustness of AI-based detection methods based on the extracted knowledge representation and causalities of attacks. Finally, new visualisation and interaction interfaces will be developed in this project to simplify the human-AI interaction. These interfaces are expected to be intuitive and user-friendly so that both technical staff (analysts) and non-technical staff (managers) can effectively interact with the results, respond quickly, and make more efficient decisions.

ANR PRCI project: SecLLM4SVD (2026-2029)

Participants: Yufei Han.

Nowadays, cyberattacks exploiting software vulnerabilities have become increasingly sophisticated and frequent, posing significant risks to individuals, organization, and critical infrastructure. Traditional vulnerability detection typically relies on rule-based [1] and signature-based [2] approaches, which are constrained by predefined patterns and rules to identify emerging security threats. These approaches struggle to cover various vulnerabilities and often miss complex security weaknesses. In recent years, artificial intelligence (AI) has demonstrated remarkable capabilities to automate the detection of software vulnerabilities. Particularly, large language models (LLMs) stand out due to their superior ability to process both natural language and programming languages seamlessly. This capability enables LLMs to interpret code in a way that closely resembles human reasoning, taking into account not only the structure of code but also its intent and broader context. Nevertheless, despite their potential in software vulnerability detection (SVD), there is a critical reliability concern laying under the state-of-the-art LLMs - they are susceptible to adversarial attacks [3,4]. Adversarial attacks against LLM-based SVD involve making subtle code modifications, such as renaming variables and/or manipulating control flows, to the input codes, which can significantly mislead the detection output of LLMs. This inherent susceptibility raises critical questions about the reliability of LLM-based code analysis to detect, categorize and fix software vulnerabilities. The adversarial risk in

LLM-based SVD solutions originates from the blackbox nature and complex model structures of LLM-based detection. The lack of mechanistic understanding about whether LLMs truly understand code semantics, or merely recognize superficial patterns in codes, e.g. statistical artefacts in training source code samples, obscure the decision-making process of LLMs over code analysis. It is thus difficult to diagnose misled detection output and verify reliability of LLM-based detection solutions. Furthermore, the adversarial risk exists widely in non-linear ML models, e.g. deep neural network models, due to curved classification boundaries [5-10]. Highly non-linear structures of LLMs bring intrinsic instability to minor modifications to input data [4]. These inherent limitations create considerable concerns regarding the reliability of deploying LLMs in the security-critical scenario of vulnerability detection. However, there have been little research efforts in understanding, mitigating and evaluating the stability of LLM-based SVD. Objective: In this context, SecLLM4SVD seeks to address the reliability concerns of LLM-based SVD through I) investigating mechanistic understanding of the LLM-based reasoning process in SVD problems, 2) unveiling the root causes of LLMs' susceptibility to adversarial attacks and 3) developing provably robust human-in-the-loop mitigation strategies. By doing so, SecLLM4SVD will enable the secure and effective deployment of LLMs in real-world software security applications, ultimately contributing to the development of more secure and trustworthy software systems.

BPI-France project: Cyberte (2025-2029)

Participants: Yufei Han.

The Cyberté project aims to transform the management, security, and energy efficiency of storage infrastructures by using artificial intelligence. The research conducted by Inria's project teams will allow Scality to integrate the following into its solutions:

- AI predictive models to anticipate hardware and software failures.
- Real-time anomaly detection algorithms to quickly identify early signs of attacks such as ransomware or data exfiltration, even in encrypted traffic.
- Frugal AI approaches and methodologies to reduce the energy footprint by using software-defined power meters (PowerAPI initiative).

ANR project: Byblos(2021-2025)

Participants: Emmanuelle Anceaume.

Byblos is a collaborative ANR project involving Rennes university and IRISA (CIDRE (now PIRAT) and WIDE research teams), Nantes university (GDD research team), and Insa Lyon, LIRIS (DRIM research team). This project aims at overcoming performance and scalability issues of blockchains, that are inherent to the total order that blockchain algorithms seek to achieve in their operations, which implies in turn a Byzantine-tolerant agreement. To overcome these limitations, this project aims at taking a step aside, and exploiting the fact that many applications – including cryptocurrencies – do not require full Byzantine agreement, and can be implemented with much lighter, and hence more scalable and efficient, guarantees. This project further argues that these novel Byzantine-tolerant applications have the potential to power large-scale multi-user online systems, and that in addition to Byzantine Fault Tolerance, these systems should also provide strong privacy protection mechanisms, that are designed from the ground up to exploit implicit synergies with Byzantine mechanisms.

ANR Project: BC4SSi (2023-2027)

Participants: Emmanuelle Anceaume.

BC4SSI is a JCJC ANR project led by Romaric Ludinard (SOTERN), involving the SOTERN and CIDRE (now PIRAT) research teams. Self-sovereign identities (SSI) are digital identities that are managed in a decentralized manner. This technology allows users to self-manage their digital identities without depending on third-party providers to store and centrally manage the data, including the creation of new identities. Implementing SSI requires a lot of care since identities are more than simple identifiers: they need to be checked by the service provider via, for instance, verifiable claims. Such requirements make blockchain technology a prime candidate for deploying SSI and storing verifiable claims. BC4SSI aims at studying the weakest synchrony assumptions enabling SSI deployment in a public Blockchain. Among the different existing challenges, BC4SSI will address the following scientific locks: alternatives to PoW security proofs, lightweight replication, scalability and energy consumption.

CMA project : Train-Cyber-Expert (TCE) (2022-2026)

Participants: Yohann Rio, Alexandre Sanchez, Guillaume Vachez, Valerie Viet Triem Tong.

As part of the France 2030 recovery plan, we are involved in the Train-Cyber-Expert (TCE) project, funded by the CMA (Competences and Jobs of the Future) call for projects. TCE is a collaborative project involving several academic partners with the aim of developing educational resources in the form of digital content and technological platforms, organized into skill blocks, with a focus on modularity, reusability, and competency-based pedagogy leading to certifications. We are involved in this project together with the INRIA SUSHI team. Our goal is to propose pedagogical ressources in the field of system security.

BPI project : SECURITY TWIN (2024-2027)

Participants: Manuel Poisson, Valerie Viet Triem Tong.

The Security Twin project, developed in collaboration with Amossys, a company specializing in cybersecurity, aims to enhance the security of information systems through the creation of a digital twin. This digital twin will faithfully replicate a real information system (IS), incorporating its configurations and vulnerabilities, to simulate realistic attack scenarios and assess their impact.

This project involves several challenges, both technical and scientific, such as:

- Modeling a security digital twin.
- Automating its deployment.
- Developing an attack agent capable of testing the resilience of IS under real-world conditions.

BPI project : DECOR (2023-2025)

Participants: Jean-François Lalande, Omnia Mohamed.

In 2025, we handled the coordination of the DECOR project with Wallix and Malizen. During the second year of the project, we created an engine that computes remediations from the investigation of an analyst after an incident. We also helped Malizen to develop new feature in the Malizen tool and we added several datasets that can be used for training when an analyst learns how to use the tool in an investigation. Finally, at the end of 2025, Wallix has bought the Malizen company.

PTTC project : MIRAGE (2025-2027)

Participants: Guillaume Vachez, Valerie Viet Triem Tong.

In 2025, we handled the coordination of the MIRAGE project founded as a *projet de transfert du PTCC*. MIRAGE is a deceptive cybersecurity project between the company Ballpoint and the PIRAT team, a joint team from CentraleSupelec/CNRS/INRIA/Univ.Rennes of the IRISA Institute. In MIRAGE, we aim to deploy adaptive honeypots to gather intelligence on cyber threats faced by organizations. A honeypot is a deliberately vulnerable subnet designed to attract attackers. A honeypot can be deployed:

- To observe the behavior of opportunistic attackers: It is then deployed on the Internet without any obvious connection to a specific organization.
- To observe an attacker who may compromise the information system (IS). In this case, it is deployed at the edge of the IS, for instance, in a subdomain or subnet created for this purpose.
- To divert an attacker who has already compromised the IS. The honeypot is then deployed within the IS itself. Deploying a honeypot presents several challenges. Such an environment must be:
- Credible: A honeypot must be sufficiently credible so that an attacker believes it is real and spends time there.
- Attractive: A honeypot must include vulnerabilities that are indeed exploitable by an attacker. These vulnerabilities should be technically within reach of the attacker but difficult enough to engage them meaningfully.
- Monitored: A honeypot must be adequately supervised to allow for the collection of relevant information. In the MIRAGE project, we propose to build on the work conducted by the PIRAT team and on the Trapster solution developed and marketed by Ballpoint.

MIRAGE will allow the PIRAT team to support the development of URSID and the Poneypot platform. These tools will enable the team to carry out high-quality experiments in ongoing and future PhD thesis. This project will also allow Ballpoint to benefit from the PIRAT's experiences and from state-of-the-art research solutions in this field.

10 Dissemination

10.1 Promoting scientific activities

Ludovic Mé serves the steering committee of RESSI (Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information).

10.1.1 Scientific events: organisation

Jean-François Lalande organized EICC 2025 (European Interdisciplinary Cybersecurity Conference) the 18-19 of June 2025 in Rennes.

Pierre-François Gimenez organized ANUBIS 2025 (Assessment with New methodologies, Unified Benchmarks, and environments, of Intrusion detection and response Systems), a workshop colocalized with ESORICS 2025 on September 26th, 2025 in Toulouse: superviz.inria.fr/anubis25/.

Member of the conference program committees Jean-François Lalande was part of the technical program committee of:

- MOBILESoft 2025: International Conference on Mobile Software Engineering and Systems
- EICC 2025: European Interdisciplinary Cybersecurity Conference
- IWCC 2025: International Workshop on Cyber Crime
- CUING 2025: International Workshop on Criminal Use of Information Hiding

Pierre-François Gimenez was part of the technical program committee of:

- ERTS 2026: Embedded Real-Time Systems
- ARTMAN 2025: Workshop on Recent Advances in Resilient and Trustworthy MAchine learning-driveN systems
- THCon 2025: Toulouse Hacking Convention

Yufei Han was part of the program committee of:

- Usenix Security 2026
- ACM CCS 2026

Reviewer Jean-François Lalande served as reviewer for:

- ESORICS 2025

Pierre-Francois Gimenez served as reviewer for:

- ESORICS 2025

Yufei Han served as reviewer for:

- ICML 2025
- NeurIPS 2025
- AAAI 2026
- PAKDD 2026
- KDD 2025
- ICLR 2025
- ICLR 2026

10.1.2 Journal

Member of the editorial boards Yufei Han serves as associate editor for Computer & Security (Elsevier)

Reviewer - reviewing activities Jean-François Lalande served as reviewer for:

- Journal of Network and Systems Management, Springer
- Cluster Computing, Springer
- Journal of Information Security and Applications, Elsevier
- ACM Transactions on Software Engineering and Methodology
- International Journal of Information Security, Springer Nature

Pierre-François Gimenez served as reviewer for:

- Transactions on Information Forensics & Security
- Journal of Network and Systems Management

Yufei Han served as reviewer for:

- IEEE Transaction on Information Froensics and Security (TIFS)
- IEEE Transaction on Dependable and Secure Computing (TDSC)
- ACM Transactions on Software Engineering and Methodology (TOSEM)

10.1.3 Invited talks

Dorian Bachelot gave invited talks for:

- the Cyb'Air SUD 2025 conference (cybair-sud.fr) at the 701 airbase in Salon-de-Provence.
- the annual DefMal workshop at Sophia-Antipolis.

Pierre-Francois Gimenez gave invited talks for:

- the European Symposium on Security and Artificial Intelligence (ESSAI)
- the "AI-driven Cyber security" Summer School
- the DefMal webinar

Pierre-Francois Gimenez participated to a round table discussion at the European Cyber Week.
Yufei Han ave invited talks for:

- gave a tutorial talk at Artificial Intelligence for Cyber and Cyber/Physical Security worshop in Paris (ai4cyber-workshop.github.io).
- gave a flash presentation at DefMal workshop at Sophia-Antipolis.

Valerie Viet Triem Tong

- gave a talk at the *Rencontre France-Taiwan Cyber et IA* in October 2025.
- gave a talk at the *Rencontres Cybersécurité et Défense de l'Institut Polytechnique de Paris*, in October 2025.

10.1.4 Scientific expertise

Ludovic Mé acts as a co-director for the PEPR (Programme et Equipement Prioritaire de Recherche) dedicated to cybersecurity – 10 research projets, global budget of 65M€.

Ludovic Mé serves:

- the Scientific Council of the LIRIMA (Laboratoire International de Recherche en Informatique et Mathématiques Appliquées) ;
- the Expert Council of the DSTN (Digital Science and Technology Network) ;
- the technical committee of the PTCC (Programme de Transfert au Campus Cyber).

Pierre-Francois Gimenez, Yufei Han and Valerie Viet Triem Tong were reviewers for the ANR agency for AAPG2025.

Jean-François Lalande was a reviewer for phD funding of Normandie Université.

Valerie Viet Triem Tong was a reviewer for 2 projects of ESF (European Science Foundation).

Valerie Viet Triem Tong was member of the HCERES commitee for the examination of VERIMAG laboratory.

10.1.5 Research administration

Ludovic Mé is director of the cybersecurity program at the “algorithms, software, and applications” program agency, operated by Inria for the French academic community.

Michel Hurfin is in charge of the High Security Laboratory (LHS) of Rennes and, for 2025, of the FIRRST (Federation of Research Infrastructures in SecuriTy of Rennes) which federates equipment owned by 5 laboratories (Inria, INSA, CentraleSupelec, IMT Atlantique and University of Rennes). The LHS notably hosts platforms resulting from the activities of the Pirat team (ShareMal, PoneyPot/HopLab). These platforms are supervised and administered by Alexandr Sanchez.

Jean-François Lalande was member of the recruitment committees for

- two Assistants Professor positions at CentraleSupelec.
- a Professor position at IUT Charlemagne (LORIA).

Valerie Viet Triem Tong was member of the recruitment committees for

- two Assistant Professor positions at CentraleSupelec.
- An Assistant Professor position at Telecom Paris
- An Assistant Professor position at Telecom Nancy
- An Assistant Professor position at Université Paris Saclay

10.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

Valerie Viet Triem Tong, Jean-François Lalande, Pierre-Francois Gimenez, Samuel Pelissier, and Loic Miller teach at CentraleSupélec (Rennes campus) in the Cybersecurity track.

Jean-François Lalande is responsible of the major program dedicated to information systems security and the special track Infosec of CentraleSupélec engineering education; He is also involved in the organization committee of EUR CyberSchool and in the computer science master degree (SIF and Cyber tracks).

Valerie Viet Triem Tong is responsible of the mastère spécialisé (post-graduate specialization degree co-dellivered with IMT-Atlantique) in Cybersecurity for CentraleSupelec. This education was awarded best French master degree in the category "Master Cybersecurity masters and Security of systems" in the Eduniversal master ranking.

10.2.1 Supervision

PhD defended:

- Hélène Orsini, Weakly-Supervised Learning for Botnet Traffic Analysis and Ad- versarial Robustness Assessment, started October 2021, supervised by Yufei Han (50%) Valérie Viet Triem Tong (25%), David Lubicz (25%). The defense tooks place at CentraleSupelec on March 6th 2025.
- Vincent Raulin, Enhancing Malware Analysis with Machine Learning, started October 2021, supervised by Valérie Viet Triem Tong (25%), Yufei Han (25%), Pierre-François Gimenez (50%). The defense tooks place at CentraleSupelec on December 16th 2025.
- Natan Talon, Automatisation de tests d'intrusion d'applications web, started December 2021, supervised by Mathieu Jaume (25%), Gilles Guette (25%), Yufei Han (25%) and Valérie Viet Triem Tong (25%). The defense tooks place at CentraleSupelec on June 23th 2025.
- Jean-Marie Mineau, The Woes of Android Reverse Engineering: from Large Scale Analysis to Dynamic Deobfuscation, started November 2022, supervised by Jean-François Lalande (100%). The defense tooks place at CentraleSupelec on December 9th 2025.

PhD in progress:

- Jean Haurogne, Analyse de code malveillant, started November 2025, supervised by Valerie Viet Triem Tong (35%), Benjamin Marais and Tony Quertiers (CIFRE Orange).
- Lucas Giordani, Inventaire de caractéristiques de sécurité dans un systems d'information, started Frebruary 2025, supervised by Gilles Guette and Valerie Viet Triem Tong.
- Bassirou Badiane, Analyse forensic de machines compromises par des botnets, started November 2025, supervised by Valerie Viet Triem Tong (50%), Yufei Han (50%).
- Antoine Cellier, Génération par intelligence artificielle de données légitimes de 2 sortes, started October 2025, supervised by Pierre-François Gimenez (25%), Frédéric Majorczyk (25%), Gilles Guette (25%), and Ludovic Mé (25%).
- Samuel Hiron, Automatic discovery of vulnerabilities in binary executables through the use of hybrid approaches mixing static and dynamic analysis, started October 2025, supervised by Frédéric Tronel (33%), Yaëlle Vinçont (33%), and Ludovic Mé (33%).
- Pierre Lledo, On intrusion detection, started December 2023, supervised by Jean-François Lalande (50%) and Frederic Majorczyk (50%).
- Sebastien Kilian, From offensive data collection to automatic attack agent, started February 2024, supervised by Jean-François Lalande (50%), Valerie Viet Triem Tong (50%).
- Lucas Aubard, Ambiguïtés de recouvrement de données dans les protocoles d'Internet et supervision reseau, started October 2022, supervised by Pierre Chifflier (25%), Gilles Guette (25%), Johan Mazel (25%) and Ludovic Me (25%).
- Fanny Dijoud, Détection d'intrusions au niveau système d'informations : détection d'anomalies par traitement IA dans des graphes dynamiques hétérogènes représentant l'activité du système, started november 2023, supervised by Michel Hurfin (25%), Pierre-Francois Gimenez (25%), Frederic Majorczyk (25%) et Barbara Pilastre (25%, DGA).
- Chaoran Li, Secure Artificial Intelligence for the Smart Grid Energy Management, started November 2024, supervised by Anne Blavette (50%, IETR), Michel Hurfin (25%), and Yufei Han (25%).
- Yohann Morel, Adaptative Honeypot, started in October 2024, supervised by Gilles Guette (50%) and Valerie Viet Triem Tong (50%).

- Matthieu Mouzaoui, Adversarially Robust Machine Learning-based Network Intrusion Detection System, started February 2024, supervised by Yufei Han (50%), Michel Hurfin (25%), and Gabriel Rilling (25%, CEA).

- Dorian Pacaud, Towards blockchain frugality, started October 2024, supervised by Emmanuelle Anceaume.

- Manuel Poisson, Évaluation automatisée du niveau de sécurité d'un système d'information, started March 2023, supervised by Valerie Viet Triem Tong (25%), Gilles Guette (25%), Frédéric Guihéry (25%) and Damien Crémilleux (25%).

- Grégor Quetel, Détection d'anomalie et création d'une sonde d'inférence sémantique, started Octobre 2023, supervised by Pierre-Francois Gimenez (25%), Eric Alata (25%), Thomas Robert (25%) and Laurent Pautet (25%).

- Patrick Zounon, Constructing Cyber Security Knowledge Encoding and Reasoning from Heterogeneous Sources with Large Language Models, started October 2024, supervised by Yufei Han (50%), Michel Hurfin (25%), and Frederic Majorczyk (25%, DGA).

- Solene Delourme, IA Générative pour la sécurité SOC : détection et réponse automatisée, supervised by Jean-François Lalande (25%), Pierre-Francois Gimenez (25%), Colin Leverger (25%), Tony Quertier (25%).

- Aatif Altaf, Security in Energy Production Systems, supervised by Romain Bourdais (50%) and Jean-François Lalande (50%).

10.2.2 Juries

Ludovic Mé was member of the PhD committee for the following PhD thesis:

- Examiner: Jolahn Vaudey, *Reconfiguration des systèmes de contrôle industriels en réaction aux cyberattaques*, Université Grenoble Alpes.

- Examiner: Eddie Billoir, *Orchestration et application du principe du moindre privilège administratif dans les systèmes Linux*, Université de Toulouse.

- President of the jury: Hélène Orsini, *Apprentissage faiblement supervisé pour l'analyse du trafic des botnets et l'évaluation de la robustesse aux attaques adversariales*, CentraleSupélec.

Valerie Viet Triem Tong was member of the PhD committee for the following PhD thesis:

- Reviewer: Antonino Vitale, *Sur la diversité et la similarité des malwares : comprendre la variabilité entre les familles de malwares*, September 2025, Eurecom.

- Reviewer: Arthur Tran Van, *Improving Cryptographic Protocol Implementation using Grammatical Inference*, November 2025, Telecom Paris.

- Reviewer : Roxanne Cohen, *Analyzing binary programs and obfuscation with graph-based representations and machine learning*, November 2025, Université Paris-Dauphine.

- President of the jury: Leo Cosseron, *Simulation précise du réseau interconnectant des machines vir- tuelles basées sur de la virtualisation matérielle pour une analyse minimisant les perturbations*, December 2025, ENS Rennes.

Valerie Viet Triem Tong was member of the HDR committee for

- Reviewer: Abdelkader Lahmadi, *Contributions to the Monitoring and Security of Networked Systems*, March 2025, Université de Lorraine. .-

10.2.3 Educational and pedagogical outreach

Jean-François Lalande has

- participated the program 1 scientifique, 1 classe : Chiche ! for the high school of Verrières-en-Anjou.
- animated the round table "Cybersecurity careers" in CentraleSupélec.

Valerie Viet Triem Tong has

- participated the program 1 scientifique, 1 classe : Chiche ! for the high school of Saumur, Lorient and Dinan.
- participated at the event "Vive la recherche" for first year students at CentraleSupelec.
- gave a talk *Une cyber attaque ce n'est pas de la magie* at the conference "Carrefour des Humanités" at Lorient in November 2025.

Dorian Bachelot has

- Managed and monitored a 6-month student project at the Rennes Cyberschool (ISTIC) on the development of an informatic worm within the LHS of Rennes.

10.3 Popularization

On the Youtube page of the PIRAT team, many scientific talks are published. Most of them are recordings from the biweekly PIRAT seminars organized by Pierre-François Gimenez. In 2025, the channel reached 240 subscribers, with 66 published videos, about 14,911 views and more than 1000 hours of cumulated watch time.

10.3.1 Participation in Live events

The PIRAT team has participed to the BreizhCTF 2025 by elaborating a sponsor challenge Pirhack, funded by Inria. The challenge has been procoposed to 120 teams of players for 600 people. It has been deployed during 24h in OVH cloud and entirely solved by 7 teams. The challenge was also replayed at RESSI 2025, for 30 people. The challenge is available on Pirat public gitlab.

11 Scientific production

11.1 Major publications

[1] X. Lyu, Y. Han, W. Wang, J. Liu, Y. Zhu, G. Xu, J. Liu and X. Zhang. 'Lurking in the shadows: Unveiling Stealthy Backdoor Attacks against Personalized Federated Learning'. In: Usenix Security 2024 - 33rd USENIX Security Symposium. Philadelphia,, United States, 2024, pp. 1–19. URL: https://inria.hal.science/hal-04827820.

11.2 Publications of the year

International journals

[2] E. Alata and P.-F. Gimenez. 'A theory of injection-based vulnerabilities in formal grammars'. In: *Theoretical Computer Science* 1057 (6th Dec. 2025), p. 115554. DOI: 10.1016/j.tcs.2025.115554. URL: https://hal.science/hal-05379247 (cit. on p. 16).

[3] G. Ishmaev, E. Anceaume, D. Frey and F. Taïani. 'Ethical Risk Analysis of L2 Rollups'. In: *ACM Transactions on the Web* (25th Dec. 2025). DOI: 10.1145/3786147. URL: https://cnrs.hal.science/hal-05446710 (cit. on p. 12).

[4] J.-M. Mineau and J.-F. Lalande. 'Class loaders in the middle: confusing Android static analyzers'. In: *Digital Threats: Research and Practice* 6.3 (24th July 2025), pp. 1–19. DOI: 10.1145/3754457. URL: https://centralesupelec.hal.science/hal-05186153 (cit. on pp. 11, 13).

[5] X. Xu, Y. Zhao, Y. Han, Y. Zhu, Z. Han, G. Xu, B. Wang, S. Ji and W. Wang. 'VFLMonitor: Defending One-Party Hijacking Attacks in Vertical Federated Learning'. In: *IEEE Transactions on Information Forensics and Security* 20 (6th May 2025), pp. 4828–4843. DOI: 10.1109/TIFS.2025.3564879. URL: https://hal.science/hal-05426773 (cit. on p. 12).

International peer-reviewed conferences

[6] E. Alata and P.-F. Gimenez. 'Towards programming languages free of injection-based vulnerabilities by design'. In: *2025 IEEE Security and Privacy Workshops (SPW)*. LangSec 2025 - Eleventh Language-theoretic Security Workshop at IEEE Security & Privacy Symposium. San Francisco CA, United States, May 2025, pp. 1–17. DOI: 10.1109/SPW67851.2025.00008. URL: https://hal.science/hal-05079251 (cit. on p. 16).

[7] L. Aubard, J. Mazel, G. Guette and P. Chifflier. 'Overlapping data in network protocols: bridging OS and NIDS reassembly gap'. In: *7*. DIMVA 2025 - Conference on Detection of Intrusions and Malware & Vulnerability Assessment. Graz, Austria, 9th July 2025. URL: https://hal.science/hal-05282871 (cit. on p. 14).

[8] L. Aubard, J. Mazel, G. Guette and P. Chifflier. 'Overlapping IPv4, IPv6, and TCP data: exploring errors, test case context, and multiple overlaps inside network stacks and NIDSes with PYROLYSE'. In: RAID 2025 - 28th International Symposium on Research in Attacks, Intrusions and Defenses. Gold Coast, Australia, Australia, 2025, pp. 1–19. URL: https://hal.science/hal-05283756 (cit. on p. 14).

[9] B. Badiane, V. Viet Triem Tong and Y. Han. 'Towards Automated Botnet Threat Intelligence with Knowledge-Guided Large Language Models'. In: *Proceedings of 18th International Symposium on Foundations & Practice of Security*. FPS 2025 - 18th International Symposium on Foundations & Practice of Security. Brest, France, 31st Dec. 2025. URL: https://inria.hal.science/hal-05411121 (cit. on p. 13).

[10] S. Chennoufi, Y. Han, G. Blanc, E. de Cristofaro and C. Kiennert. 'PROTEAN: Federated Intrusion Detection in Non-IID Environments Through Prototype-Based Knowledge Sharing'. In: *Computer Security – ESORICS 2025. ESORICS 2025. Lecture Notes in Computer Science*. ESORICS 2025 - 30th European Symposium on Research in Computer Security. Vol. 16053. Lecture Notes in Computer Science. Toulouse, France: Springer Nature Switzerland, 13th Oct. 2026, pp. 103–125. DOI: 10.1007/978-3-032-07884-1_6. URL: https://hal.science/hal-05379203 (cit. on p. 13).

[11] L. Cornanguer and P.-F. Gimenez. 'TADAM: Learning Timed Automata from Noisy Observations'. In: *Proceedings of the 2025 SIAM International Conference on Data Mining (SDM)*. SDM 2025 - SIAM International Conference on Data Mining. Alexandria Virginia, United States: SIAM, 2025, pp. 1–14. DOI: 10.1137/1.9781611978520.10. URL: https://hal.science/hal-04886774 (cit. on p. 13).

[12] R. Garcia, A. Lahmadi, P.-F. Gimenez and C. Sala. 'ROSCA: Robust and Scalable Security Alert Correlation and Prioritisation using the MITRE ATT&CK Framework'. In: WATCH 2025 - First International Workshop on Analytics, Telemetry, and Cybersecurity for HPCC (High Performance Computing and Communications). Taipei, Taiwan, 2025. DOI: 10.1145/3733826.3762680. URL: https://inria.hal.science/hal-05351162 (cit. on p. 14).

[13] P.-F. Gimenez. 'Synthetic Network Traffic Generation for Intrusion Detection Systems: a Systematic Literature Review'. In: ANUBIS 2025 - 1st International Workshop on Assessment with New methodologies, Unified Benchmarks, and environments, of Intrusion detection and response Systems. Vol. Computer Security. Esorics 2025 International Workshops: Anubis 2025, Secai 2025, Secassure 2025, Stmus 2025, Toulouse, France, September 22-24, 2025, (Lecture Notes in Computer Science #1623). Toulouse, France, 2025, pp. 1–18. URL: https://hal.science/hal-05379236 (cit. on p. 17).

[14] P.-F. Gimenez, S. Sivaprasad and M. Fritz. 'Certifiably Robust Malware Detectors by Design'. In: *IFIP Advances in Information and Communication Technology*. SEC 2025 - 40th IFIP TC-11 International Information Security and Privacy Conference. Vol. 746. IFIP Advances in Information and Communication Technology. Maribor, Slovenia: Springer Nature Switzerland, 16th May 2025, pp. 125–139. DOI: 10.1007/978-3-031-92886-4_9. URL: https://hal.science/hal-05079220 (cit. on p. 16).

[15] A. Khalin, J.-F. Lalande and R. Bourdais. 'Detecting Energy Theft Attacks on an Off-Grid Charging Station'. In: *Proceedings of ACM eEnergy '25*. EnergySP 2025 - ACM SIGEnergy Workshop on Cybersecurity and Privacy of Energy Systems. Rotterdam, Netherlands: ACM, 2025, pp. 1–7. DOI: 10.1145/3679240.3734650. URL: https://centralesupelec.hal.science/hal-05059599 (cit. on pp. 11, 14).

[16] S. Kilian, V. Viet Triem Tong, J.-F. Lalande, F. Majorczyk, A. Sanchez, N. Talon, P.-V. Besson, H. Orsini, P. Lledo and P.-F. Gimenez. 'CasinoLimit: An Offensive Dataset Labeled with MITRE ATT&CK Techniques'. In: *Proceedings of the 28th International Symposium on Research in Attacks, Intrusions and Defenses*. RAID 2025 - 28th International Symposium on Research in Attacks, Intrusions and Defenses. Gold Coast, Australia, 2025. URL: https://hal.science/hal-05224264 (cit. on pp. 11, 16).

[17] M. Lanvin and F. Majorczyk. 'Get out of DEDALE with RESCOUSSE: a New Dataset and Testbed for Evaluating the Detection of APT attacks among Network and System Logs'. In: *Computer Security. Esorics 2025 International Workshops: Anubis 2025, Secai 2025, Secassure 2025, Stmus 2025, Toulouse, France, September 22-24, 2025, (Lecture Notes in Computer Science #1623).* ANUBIS 2025 - Assessment with New methodologies, Unified Benchmarks, and environments, of Intrusion detection and response Systems. Toulouse, France, 2025, pp. 1–21. URL: https://hal.science/hal-05329482 (cit. on pp. 11, 17).

[18] F. Majorczyk, B. Pilastre and F. Dijoud. 'A New Hope for DARPA OpTC'. In: CSET 2025 - Cyber Security Experimentation and Test at ACSAC 2025 International Workshops. Honolulu, United States, 2025. URL: https://inria.hal.science/hal-05474126 (cit. on p. 15).

[19] L. Miller, D. Pacaud, N. Derousseaux-Lebert, E. Anceaume and R. Ludinard. 'Mining in Logarithmic Space with Variable Difficulty'. In: *ACM SIGSAC Conference on computer and communications security (CCS)*. CCS 2025 - Conference on computer and communications security. Tapei, Taiwan: ACM, 2025, pp. 1–15. DOI: 10.1145/3719027.3744874. URL: https://cnrs.hal.science/hal-05294549 (cit. on p. 15).

[20] S. Pélissier, N. Mehanna, S. Roux, Q. Perez, W. Rudametkin, J. Bourcier and P. Laperdrix. 'Users Pay Twice: The Hidden Energy Cost of Web Advertising'. In: WWW 2026 - ACM Web Conference. Dubai, United Arab Emirates: ACM, 2026. DOI: 10.1145/3774904.3792414. URL: https://hal.science/hal-05479340.

[21] G. Quetel, E. Alata, P.-F. Gimenez, T. Robert and L. Pautet. 'Superviz25-SQL: High-Quality Dataset to Empower Unsupervised SQL Injection Detection Systems'. In: ESORICS - ANUBIS 2025 - 1st International Workshop on Assessment with New methodologies, Unified Benchmarks, and environments, of Intrusion detection and response Systems. Vol. Computer Security. Esorics 2025 International Workshops: Anubis 2025, Secai 2025, Secassure 2025, Stmus 2025, Toulouse, France, September 22-24, 2025, (Lecture Notes in Computer Science #1623). Toulouse, France, Sept. 2025, pp. 1–20. URL: https://hal.science/hal-05314211 (cit. on pp. 11, 17).

[22] X. Xu, Z. Li, Y. Han, B. Wang, J. Liu and W. Wang. 'From Risk to Resilience: Towards Assessing and Mitigating the Risk of Data Reconstruction Attacks in Federated Learning'. In: SEC 2025 - 34th USENIX Conference on Security Symposium. Seattle, United States, 13th Aug. 2025, pp. 1–20. DOI: 10.5555/3766078.3766240. URL: https://hal.science/hal-05426791 (cit. on p. 15).

National peer-reviewed Conferences

[23] P. Zounon, Y. Han, M. Hurfin and F. Majorczyk. 'From Text to Insight: Encoding Cyber Attack Patterns in Threat Intelligence Reports Using Knowledge Graphs and LLMs'. In: RESSI 2025 - Rendez-Vous

de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information. Lanniron, France, 2025, pp. 1–4. URL: https://hal.science/hal-05471829 (cit. on p. 15).

Conferences without proceedings

[24] D. Pacaud, L. Miller, E. Anceaume and R. Ludinard. 'Preuves non-interactives : la nouvelle ère des chaînes compressées'. In: *ALGOTEL 2025 – 27èmes Rencontres Francophones sur les AspectsAlgorithmiques des Télécommunications*. ALGOTEL 2025 : 27èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications. Saint Valery-sur-Somme, France, 2nd June 2025. URL: https://hal.science/hal-05034002.

Reports & preprints

[25] T. Albouy, E. Anceaume, D. Frey, M. Gestin, A. Rauch, M. Raynal and F. Taïani. *Asynchronous BFT Asset Transfer: Quasi-Anonymous, Light, and Consensus-Free*. 18th Feb. 2025. URL: https://inria.hal.science/hal-04578985.

[26] E. Anceaume, R. Frédérique and B. Sericola. *Proof-of-Eligibility and Ephemeral participation to Solve Consensus in a Permissionless Environment*. 1st Aug. 2025. URL: https://cnrs.hal.science/hal-05423728.

[27] O. Barais, R. D. Cosmo, L. Mé, S. Zacchiroli and O. Zendra. *Software Identification for Cybersecurity: Survey and Recommendations for Regulators*. Software Heritage Security project (SWHSec), 28th Mar. 2025. URL: https://hal.science/hal-05009757.

[28] H. Debar, L. Mé, J. Leneutre, V. Nicomette, J. François, C. Gouy-Pailler, G. Blanc and S. Mocanu. *SuperviZ supervision et orchestration de la sécurité Rapport d'avancement à mi-projet*. Télécom SudParis (Institut Mines-Télécom); Inria, Oct. 2025, pp. 1–56. URL: https://hal.science/hal-05315778.

[29] G. Ishmaev, E. Anceaume, D. Frey and F. Taïani. *Ethical Risk Analysis of L2 Rollups*. 5th Dec. 2025. URL: https://inria.hal.science/hal-05404607 (cit. on p. 12).

[30] L. Miller, D. Pacaud, N. Derousseaux, E. Anceaume and R. Ludinard. *Technical Report: Mining in Logarithmic Space with Variable Difficulty*. 1st July 2025. URL: https://cnrs.hal.science/hal-05138795 (cit. on p. 15).

Scientific popularization

[31] M. Poisson and F. Guihéry. *La défense par les chemins d'attaque: En s'appuyant sur des modèles partagés, les organisations peuvent analyser les techniques, cartographier les chemins d'attaque et bâtir une défense véritablement proactive.* 15th Nov. 2025. URL: https://inria.hal.science/hal-05387433.

11.3 Cited publications

[32] A. Crowder, A. Lu, K. Childs, C. Stillman, P. Traynor and K. R. Butler. ' Data to Infinity and Beyond: Examining Data Sharing and Reuse Practices in the Computer Security Community '. In: *2025 IEEE Symposium on Security and Privacy (SP)*. Los Alamitos, CA, USA: IEEE Computer Society, May 2025, pp. 2678–2696. DOI: 10.1109/SP61157.2025.00180. URL: https://doi.ieeecomputersociety.org/10.1109/SP61157.2025.00180 (cit. on p. 11).