Position Description

1. General Information

Name of the position	**Towards Explainable Intrusion Detection with the Human in the Loop**
Foreseen date of enrolment	January 2025
Position is funded by	• COFUND, Marie Skłodowska-Curie Actions (MSCA), Horizon Europe, European Union • Institut Mines Telecom (IMT) Atlantique • The University of Adelaide (UoA)
Research Host	IMT Atlantique
PhD awarding institutions	IMT Atlantique & The University of Adelaide
Locations	Primary: Brest, France Secondary: Adelaide, Australia
Supervisors	Pr Fancoise SAILHAN (IMT Atlantique) Kaie Maennel and Olaf Maennel (University of Adelaide)
Group of discipline	Computer Science

2. Research topics (only one of these projects will be funded)

Project 1: *Enhanced Anomaly/intrusion detection with the Human in the loop*

The systems used to **detect anomalies or intrusions** have evolved considerably in recent years following the advances in **artificial intelligence (AI)** techniques and sensors integrated into modern devices and widely deployed (including in Security Operation Centres). Despite many cyber security tasks being automated, currently and in the future, **decision-making still needs a human** who can critically assess and spot mistakes in AI agents' tasks and find the most effective ways of working together as a team. In particular, the risks associated with human (such as fatigue, difficulty in grasping/understanding/explaining the indicators provided by the detection system) and the lack of explicability of the results provided by the detection system can have detrimental consequences. Explainable AI (XAI) or making AI models interpretable to human users, is critical in the cyber security domain in that XAI may allow security operators, who are overwhelmed with tens of thousands of security alerts per day (most of which are false positives), to assess the potential threats better and reduce alert fatigue (Charmet et al. 2022). Therefore, an unbiased and empirical understanding of the human-AI cyber defence teams' collaboration is critical.

This research project aims to:
- Study and propose enhancements to an AI-based intrusion detection system that provides indicators of compromise to explain the results provided by the AI model(s);

This project has received funding from the European Union’s Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 101081465

- Study and propose different types of explanations (e.g., semantic or causal, including counterfactual examples) of the results (classification), rather than just features as in current solutions;
- Measure the adequacy of these indicators and explanations, taking into account the influence of human/analyst and environmental factors;
- Identify the biases, barriers and vulnerabilities in information sharing, risk assessment and effective decision-making introduced by human-AI teamwork.

The research will be performed with the support of the Defence Science and Technology Group (DSTG) in South Australia.

Supervisors: Francoise SAILHAN (IMTA), Kaie Maennel (UoA), Olaf Maennel (UoA)

Research Fields: Computer science, cyber security, human aspects

Project 2: *Adverserial data poisoning attacks*

The systems used to **detect anomalies or intrusions** have evolved considerably in recent years following the advances in **artificial intelligence (AI)** techniques and sensors integrated into modern devices and widely deployed (including in Security Operation Centres). Despite significant advances, the detection process is error-prone due to its dynamic and evolving nature of threat landscape. The cybercriminals and threat actors are racing to use AI to find innovative new hacks (e.g., encompassing techniques for image, audio and video generation) and adversarial AI- and generative AI-enabled attacks are already taking place. The attack techniques include poisoning a benign training dataset, evading (i.e., using a malicious input to get an unexpected output), oracling (i.e., stealing information by probing the model) and adversarial reprogramming. Thus, effective log-driven analysis with high-level of explainability is critical, as many of these attacks can have an impact on society and human life considering our digitalised world.

The research project aims to:
- identify and develop efficient AI-based algorithms to parse heterogenous log formats accurately to identify adversarial attacks, including evasion attacks designed to evade detection by manipulating log data, and extract relevant features from log data to improve detection accuracy and reduce false positives;
- design real-time log analysis frameworks capable of promptly detecting intrusions as they occur and support detection across diverse domains, such as cloud, IoT networks, and industrial control systems;
- investigate and propose scalable architectures and optimization methods to handle large volumes of log data without sacrificing detection accuracy or performance;
- Develop standardized benchmarks and evaluation metrics for assessing performance of log-driven intrusion detection systems, facilitating fair comparisons between different approaches and enabling reproducible research.

The research will be performed with the support of the Defence Science and Technology Group (DSTG) in South Australia.

Supervisors: Francoise SAILHAN (IMTA), Kaie Maennel (UoA), Olaf Maennel (UoA)

Research Fields: Computer science, cyber security, artificial intelligence, machine learning

Project 3: *Combining Metagraphs and AI-Enhanced Anomaly Detection*

The systems used to **detect anomalies or intrusions** have evolved considerably in recent years following the advances in **artificial intelligence (AI)** techniques and sensors integrated into modern devices and widely deployed (including in Security Operation Centres). Despite significant advances, the detection process is error-prone due to its dynamic and evolving nature of threat landscape. The vast amount of log data presents a significant challenge, requiring efficient processing and analysis to detect and respond to threats effectively. **Visualization** is critical, offering analysts intuitive insights into complex cybersecurity data for rapid threat detection and decision-making. This research aims to **integrate metagraphs and AI-enhanced anomaly detection** for intrusion detection context develop a versatile framework applicable across diverse cyber domains.

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 101081465

The research project aims to:
- develop an innovative framework that integrates metagraphs with AI techniques for enhanced anomaly detection;
- utilize metagraphs to model complex systems, capturing hierarchical relationships and interdependencies to provide a holistic view of system dynamics, facilitating visual anomaly detection;
- employ AI algorithms, such as deep learning and reinforcement learning, to enhance anomaly detection capabilities by training AI models on metagraph-structured data to learn complex patterns and detect anomalies indicative of cyber security threats;
- evaluate the performance of the integrated system using real-world datasets, assessing its effectiveness in detecting and mitigating cyber security threats.

The research will be performed with the support of the Defence Science and Technology Group (DSTG) in South Australia.

Supervisors: Francoise SAILHAN (IMTA), Kaie Maennel (UoA), Olaf Maennel (UoA)

Research Fields: Computer science, cyber security, mathematics

3. Employment Benefits and Conditions

L'École Nationale Supérieure Mines-Télécom Atlantique Bretagne - Pays de la Loire (IMT Atlantique) offers a 36-months full-time work contract (with the option to extend up to a maximum of 42 months).

The remuneration, in line with the European Commission rules for Marie Skłodowska-Curie grant holders, will consist of an annual gross salary of 30,300 EUR. Of this amount, the estimated salary to be perceived by the Researcher is 1,975 EUR net per month. However, the definite amount to be received by the Researcher is subject to national tax legislation (3.5 %).

Benefits include

- Becoming a Marie Skłodowska-Curie fellow and be invited to join the Marie Curie Alumni Association
- Access to all the necessary facilities and laboratories at IMT Atlantique, Lab-STICC and the University of Adelaide.
- Tuition fee waiver at both PhD enrolling institutions.
- Yearly travel allowance to cover flights and accommodation for participating in AUFRANDE events.
- 10,000 EUR allowance to cover flights and living expenses for up to 12 months in Australia.
- 49 days paid holiday leave
  - Contractual doctoral students benefit from the paid leaves, under the same conditions as all employees of the Institut Mines-Télécom
- Sick leave
  - Serious illness leave, granted after three years' service and examination by the medical committee - provision which can only be applied in cases of contract extension beyond three years, as it is granted after three years' service (article 61-1),
  - Leave for occupational injury or illness (article 63)
  - Exceptional leave of absence for family events (article 54)
  - Ordinary sick leave with compensation based on seniority (article 60-1)
- Parental leave

4. PhD enrolment

Successful candidates for this position will be enrolled by the following institutions and must comply with their specific entry requirements, in addition to AUFRANDE's conditions.

Applicants must demonstrate sufficient background and experience in independent supervised research to successfully complete a PhD. This includes holding Master of Science or Master of Engineering or another similar world-class master's degree containing a minimum of 12 credit points by research, with an overall Grade Point Average (GPA) of 5.0 or higher and a GPA of 6.0 or higher in the Research Component.

Applicants must demonstrate English language proficiency in the form of an English test that has been taken within the two years preceding the date of commencement. The following test types are accepted:

- IELTS (International English Language Testing System) Academic
- TOEFL (Test of English as a Foreign Language) Internet Based Test
- PTE (Pearson Test of English) Academic
- C1 Advanced (formerly CAE - Cambridge English: Advanced)

Note that a security background check might be required.

IMT Atlantique

Applicants from foreign countries may have to be evaluated by French Authorities before being allowed to be hosted by IMT Atlantique. In case of denial, the employment will not be carried out.

More information: https://www.imt.fr/en/education/our-degrees/phd/

The University of Adelaide

Short-listed applicants will need to demonstrate their suitability for entry to the program. More information: https://www.adelaide.edu.au/graduate-research/future-students/how-to-apply#step-4-apply-online

More information: https://www.adelaide.edu.au/degree-finder/2023/hrdoctor_philosophy.html

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 101081465