Offer #2024-07147

PhD Position F/M Constructing Cyber Security Knowledge Encoding and Reasoning from Heterogeneous Sources with Large Language Models

**Contract type:** Fixed-term contract

**Level of qualifications required:** Graduate degree or equivalent

**Fonction:** PhD Position

**About the research centre or Inria department**

The Inria Centre at Rennes University is one of Inria's eight centres and has more than thirty research teams. The Inria Centre is a major and recognized player in the field of digital sciences. It is at the heart of a rich R&D and innovation ecosystem: highly innovative PMEs, large industrial groups, competitiveness clusters, research and higher education players, laboratories of excellence, technological research institute, etc.

**Context**

Within the framework of the ANR project "CKRISP".

**Assignment**

**Background**

AI-based cyber-attack detection models face several bottlenecks that need to be addressed. First, AI models require large amounts of training data to cover as many as possible attack behaviours. However, in the practice of cybersecurity, the deployed probes cannot guarantee to provide comprehensive coverage over different attack behaviours, especially emerging new attacks. The lack of coverage over different attack behaviours, i.e. the issue of out-of-distribution samples in attack detection [Yang21], makes it challenging to accurately categorise the threats that are being faced. Second, AI models may generate a large number of false positives that are difficult to distinguish from real intrusion. The origin of this issue stems from the lack of interpretability of AI models with complex architectures, e.g., Deep Neural Networks. The difficulty of comprehending the detection logic of AI models leads to mistrust and unaccountability of AI-based detection technologies. In light of the aforementioned limitations of AI-based attack detection, we believe representing cyber security knowledge, e.g. encoding and inferring attack behaviours with AI models is the key to deliver verifiable and accountable AI-based cyberattack prediction system.

**The objective of this thesis**

We will design automated attack knowledge fusion from heterogeneous attack information sources, such as dynamic analysis traces of malware samples and high-level threat reports of malware families written by human analysts, based on the combination of Large Language Model (LLM) and Knowledge Graph (KG) based AI models. First, we will focus on identifying entities of security incidents (Named Entity Recognition, NER) and relations between the entities (Extraction of Relation, ER) [Rastogi20] in logs and reports using LLMs. LLMs are deployed as a generative model to produce and match candidates of meaningful entities in security incidents [Rastogi20, Das22]. Furthermore, LLMs can be used to tokenise the low-level behavioural logs and perform reasoning between logs to explore possible causal relations via text embedding techniques. The extracted entities and relations between them will be used to build the security knowledge graph.

Over the constructed knowledge graphs, we propose to perform the causal-reasoning-based attack prediction via reinforcement learning (RL)-based exploration in cyber security knowledge graphs. We further integrate human analysts' feedback to refine the reasoning capability of the RL-based exploration models. To achieve this goal, we plan to develop few-shot learning methods [Hejna22], such as few-shot preference learning with human feedback to tailor the exploration policy learnt using knowledge graphs. Furthermore, we will focus on continual learning [Zhangxk22] to update dynamically the contents of the cyber security knowledge graphs. Similar as active learning, this algorithm also allows human experts to encode the verified and corrected causal relations as additional supervision to update links/entities in the knowledge graphs.

**Expectations**

The candidate for this thesis is expected to have accomplished courses on Machine Learning and/or have
experience of implementing Machine Learning algorithms using Python toolboxes, e.g. Pytorch, for practical data mining problems. Knowledge about cyber security, such as malware classification and/or network intrusion detection is required. Candidates with the experience of security data analysis will be preferred.

References


Main activities

This thesis will be conducted at INRIA Rennes and co-supervised with Eurocom researchers at Sophia-Antipolis. The Inria Rennes - Bretagne Atlantique Centre is one of Inria's eight centres and has more than thirty research teams. The Inria Center is a major and recognized player in the field of digital sciences. It is at the heart of a rich R&D and innovation ecosystem: highly innovative PMEs, large industrial groups, competitiveness clusters, research and higher education players, laboratories of excellence, technological research institute, etc. The monthly gross salary for the PhD candidate amounts around 2000 euros. For every applicant, please submit online your resume, cover letter and letters of recommendation.

Skills

Technical skills and level required : Machine Learning, Statistics, Information theory, Pytorch, Intrusion Detection

Languages : English

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Possibility of teleworking (90 days per year) and flexible organization of working hours
- Partial payment of insurance costs

Remuneration

Monthly gross salary amounting to 2100 euros for the first and second years and 2200 euros for the third year

General Information

- Theme/Domain: Data and Knowledge Representation and Processing
- Statistics (Big data) (BAP E)
- Town/city: Rennes
- Inria Center: Centre Inria de l'Université de Rennes
- Starting date: 2024-06-01
- Duration of contract: 3 years
- Deadline to apply: 2024-04-01

Contacts

- Inria Team: CIDRE
- PhD Supervisor: Han Yufei / yufei.han@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and
entrepreneurial projects that have a worldwide impact.

**The keys to success**

The candidate for this thesis is expected to have accomplished courses on Machine Learning and/or have experience of implementing Machine Learning algorithms using Python for practical data mining problems. Especially, expertise in using Pytorch will be required in the project. Theoretical developments are also expected based on statistics and theory of machine learning and approximation. Knowledge about intrusion detection systems and anomaly detection will be preferred.

**Warning**: you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

**Instruction to apply**

Please submit online: your resume, cover letter and letters of recommendation eventually

**Defence Security**: This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

**Recruitment Policy**: As part of its diversity policy, all Inria positions are accessible to people with disabilities.