

2021-04118 - Engineer position on developing an open-source machine learning toolbox for network analytics

Contract type : Fixed-term contract
Level of qualifications required : Graduate degree or equivalent
Fonction : Temporary scientific engineer

Context

Team

This position is proposed by the RESIST team of the Inria Nancy Grand Est research lab, the French national public institute dedicated to research in digital Science and technology.

The team has strong research records in designing new methods and developing tools based on machine learning algorithms to manage networks. We have demonstrated the efficiency of our techniques in various scenarios and, most notably, on network traffic analysis: fingerprinting user actions on IoT devices, detection of anomalous behavior in encrypted TLS communications, analysis of large darknet,...

The team is actually one of the European research group in network management and is particularly focused on empowering scalability and security of networked systems through a strong coupling between monitoring, analytics and network orchestration. Its expertise is recognized and applied in large collaborative projects at an international scale.

About 30 members are in the team, that include permanent researchers, professors, PhD students and engineers working on various topics (artificial intelligence applied to network management, programmable dataplanes, virtualization of networks, security monitoring...)

The team is part of LORIA which is a joint lab between INRIA, University of Lorraine and CNRS. It provides a full ecosystem to support highly innovative research and development with more than 400 people in total within a larger scientific campus of Nancy.

Contacts

Jérôme François (jerome dot francois at inria dot fr) and Frederic Beck (frederic dot beck at inria dot fr)

Assignment

Project overview

During the last twenty years, there has been an increasing adoption of advanced analytics techniques, especially Machine Learning (ML), in all areas of networking developed to achieve a higher level of automation with the key objectives being to extract relevant information from observations in order to reach different goals such as enhancing performance or end-user experience, lowering the carbon footprint or improving network security.

With the exponential increase of the use and adoption of ML techniques in the last decade, tools to support ML have reached a high maturity level including scikit-learn, orange, keras, dask, etc. In particular those tools have been design to use of ML for non AI expert. Even further people are developing tools to auto-configure the ML algorithms like AutoML.

Historical communities in image or speech processing have been able to produce and standardize techniques and open-source tools available for all. Although network community is now both a user and a provider of techniques to support the use of ML, a very few techniques have been community-wide adopted or standardized and the main trend is to redefine and redevelop similar techniques for each use.

Therefore, our ambition is to support and lead a similar effort in our scientific community, networking and network management, with as a final goal the development of an extensible ML toolbox for networks.

Main activities

Activities

The objective is to create a first version of the toolbox which must be extensible and re-configurable. Indeed, as a starting point we will focus on the initial steps of ML pipeline that encompasses data ingestion, data pre-processing to represent data as graphs or vectors and feature extraction.

The toolbox will be open-source and must be interfaced with other existing tools as for example scikit-learn.

The initial version of the library will have the following expected functionalities:

- Extract features from network traffic data format (pcap and IPFIX) including temporal features and encrypted-specific features
- Extract meta behavioral feature from graph representation of the network activity
- Distance and similarity metrics over defined features
- Embeddings as fixed size vector of extracted feature to remove categorical dat

The engineer will have to directly interact with all team members to derive the requirement of such a library keeping in mind that the goal is to make this library accessible to everybody, even to non members (open-source project).

The tasks of the engineer will be:

- Specification of the software architecture
- Identification of features to be extracted through interaction with the research team
- Identification of existing tools to be reused
- Specification and developing modules to extract data from raw data files
- Specification and developing modules to extract knowledge from data and metrics or

General Information

- **Theme/Domain :** Networks and Telecommunications
Software engineering (BAP E)
- **Town/city :** Villers lès Nancy
- **Inria Center :** CRI Nancy - Grand Est
- **Starting date :** 2021-01-01
- **Duration of contract :** 2 years
- **Deadline to apply :** 2021-10-27

Contacts

- **Inria Team :** RESIST
- **Recruiter :**
François Jérôme / jerome.francois@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

How to apply:

Upload your file on jobs.inria.fr in a single pdf or zip file, and send it as well by email to jerome dot francois at inria dot fr and frederic dot beck at inria dot fr. Your file should contain the following documents:

- Your CV
- A cover/motivation letter describing your interest
- Your degree certificates and transcripts for Bachelor and Master (or the last 5 years).
- In addition, one recommendation letter from a person who supervises(d) your work and a letter from another colleague (with contact information: telephone and email)
- Applications are to be sent as soon as possible

Instruction to apply

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.

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.

- embedding from the constructed representation (vectors, graphs, time-series...)
- Integration of research work supporting auto-configuration of ML algorithms
- Maintaining the developer documentation and user guide
- Preparing and presenting tutorials, demos and hackathon in the RESIST team and for international venues (scientific conferences)
- Providing support to the beta tester (the team)

Skills

Required qualifications

- Required qualification: Diplôme d'ingénieur, Master degree in Computer Science or Computer engineering
- Required knowledge: networking, machine learning and their relative tools (wireshark, scikit learn, pandas, dask...)
- Languages: Shell, python and others are appreciated
- Software development: continuous integration and collaborative development using gitlab
- Fluent in english (writing and oral communication)
- Comfortable with meetings and webconference situations

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

According to profile