Thème/Domaine : Réseaux et télécommunications
Ingénierie logicielle (BAP E)
Ville : Villers-lès-Nancy
Centre Inria : CRI Nancy - Grand Est
Date de prise de fonction souhaitée : 2021-04-01
Dureté de contrat : 2 ans
Date limite pour postuler : 2022-02-25

Contexte et atouts du poste

An open-source ML library for network researchers and practitioners

In collaboration with network experts used to apply ML to their problems, you'll be the main developer of a new toolbox to support their activities, improve the reusability and the scalability of their tools. If you have a good background in networking and some knowledge in ML and you want to develop a tool whose the ambition is to be adopted by a worldwide community, you should consider this offer.

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 darknets,

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 (jérôme.francois@inria.fr) and Frederic Beck (frederic.beck@inria.fr)

Mission confiée

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 and standardized and the 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.

Principales activités

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 data

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).

Informations générales

- Thème/Domaine : Réseaux et télécommunications
- Ingénierie logicielle (BAP E)
- Ville : Villers-lès-Nancy
- Centre Inria : CRI Nancy - Grand Est
- Date de prise de fonction souhaitée : 2021-04-01
- Durée de contrat : 2 ans
- Date limite pour postuler : 2022-02-25

Contacts

- Équipe Inria : RESIST
- Recruteur :
  François Jérôme / jérôme.francois@inria.fr

A propos d'Inria

Inria est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnels d’appui à la recherche et à l’innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde.

Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L’institut s’efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l’économie.

L’essentiel pour réussir

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 jérôme.francois@inria.fr and frederic.beck@inria.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 (or the last 5 years).

Applications are to be sent as soon as possible.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

Attention : Les candidatures doivent être déposées en ligne sur le site Inria.

Le traitement des candidatures adressées par d'autres canaux n'est pas possible.
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 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)

Compétences

Required qualifications

- Required qualification: Diplôme d’ingénieur; Master degree in Computer Science or Computer engineering
- Required knowledge: networking (you must know how machines communicate over Internet, TCP/IP, DNS,..., some knowledge in machine learning (no need to be an expert on neural architectures for deep learning but you should have a good understanding of what is a data analysis pipeline, data processing, feature engineering, etc.) and their relative tools (wireshark, sci-kit learn, pandas, dask...)
- Languages: Shell, python and others are appreciated
- Software development: design, implementation, testing, continuous integration and collaborative development using gitlab, writing technical and user documentation
- Fluent in english (writing and oral communication)
- Comfortable with meetings and webconference situations
- Project management: ability to define achievable milestones, anticipate/mitigate risks

Avantages

- 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

Rémunération

According to profile