



Offer #2021-03539

PhD Position F/M Artificial intelligence-based data analysis in heterogeneous and volatile environments: application to cooperative autonomous vehicles

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Level of experience : Recently graduated

About the research centre or Inria department

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.

Context

Financing Project

This PhD will be hosted by Inria (KerData team, Rennes Bretagne Atlantique) and will be co-funded by Inria and by the Ministry of the Armies.

PhD advisors

- [Gabriel Antoniu](#) (Inria, KerData team)
- [Alexandru Costan](#) (Inria, KerData team)
- [François Tessier](#) (Inria, KerData team)
- Loïc Cudennec (DGA MI)

Context

During the last fifteen years, artificial intelligence has been experiencing a new boom, made possible by significant theoretical, algorithmic and technological advances. These technological advances concern hardware gas pedals such as GPGPUs, which allow machine learning from a large volume of data in a reasonable time, and embedded systems in general, which allow inference tasks to be deployed as close as possible to the operational context. One of the particularities of AI is to interfere – positively – in all domains: computer-aided design, the search for optimized solutions and of course all business applications with tedious, repetitive and sufficiently complex tasks that cannot be performed by a conventional computer program.

One of the application domains benefiting greatly from these advances is the design of cooperative and autonomous vehicle fleets. In these complex systems, decision making occurs at multiple levels:

- at the vehicle level to perform tasks related to driving or piloting, in direct contact with mechanical systems and which must be effected in a short time loop, sometimes in constrained real time. AI is for example used in this framework to perform inference tasks on data from sensors for object recognition.
- at the level of a sub-system composed of several vehicles, where choices are decided in a cooperative way, in the form of consensus, such as calculating trajectories to avoid collisions. AI can be used for inference to predict and propose trajectories.
- at the level of a generally offline system with substantial computational capabilities, allowing machine learning, exploration, simulation, and evaluation of scenarios using operations research algorithms. One application is solving the constrained cooperative traveler problem.

One of the major challenges of these distributed heterogeneous systems lies in the ability to have relevant data at a given location and at a given time. For this, three mechanisms must be finely studied.

- Data locality: data sharing between devices, including the ability to locate, transfer and ensure data integrity and consistency.
- Task scheduling: scheduling computational tasks on equipment, which includes the possibility of knowing which tasks can be executed on each piece of equipment, depending on the computational resources (which hardware, which software stack) and on the state of the system at

a given time (what rate of use of the hardware, what state of charge in the case of a system running on batteries).

- **Orchestration:** a mechanism having global knowledge or multiple local knowledge of the system, and able to make or propose decisions on both of the above mechanisms. The orchestration mechanism includes the evaluation of different scenarios to satisfy a single setpoint, each scenario involving potentially different costs for communications and computations on each piece of equipment. It further includes knowledge or prediction of task computation and data transfer times.

These issues call upon many well-studied research themes in homogeneous distributed systems such as supercomputers or large-scale file-sharing systems. But these themes run into new problems once the targeted system is made up of several heterogeneous systems, each bringing different paradigms for task and data management. The research context of this thesis is related to the management of this heterogeneity.

Use cases

This thesis may focus on the analysis of the needs of one of the following applications:

Application 1: SCAF Project

The SCAF project [6, 5], the air combat system of the future, features all the issues related to the deployment of AI in embedded, volatile and heterogeneous operational environments. This European project initiated in 2017 by France and Germany, then joined by Spain in 2019, must offer an operational system by 2040. It is being conducted in collaboration with major aerospace and defense companies. The vision of the system includes ground-based computing centers and on-board computing units in satellites, aircraft and drones of all sizes. These carriers form a "combat cloud", i.e. a distributed, heterogeneous and volatile system, on which complex processing tasks and shared data management between agents are planned to be deployed. The "combat cloud" is a hierarchical logical system in which some subsystems are themselves distributed, heterogeneous and volatile. For example, a first system (RC for "remote carrier") consists of a swarm of drones, whose communications are based on the peer-to-peer model. A second system (NGF for "next generation fighter") consists of a fighter aircraft, in which several computing units are specialized for various tasks, based on the model of the calculator with massively parallel gas pedals. Finally, a third system (NGWS for "next generation weapon system") includes the two previous systems, on the hierarchical model (the NGF playing the role of "leader" for the RC). Each of these systems thus exhibits various paradigms for task and data management, making them a characteristic case study for this thesis topic.

Artificial intelligence is involved in many aspects of the SCAF project : in inference to ensure carrier autonomy, to decrease the mental load on pilots and operators, and for optimization and decision support in the event of mission reconfiguration. Machine learning tasks are also likely to be conducted in operation in order to refine a pre-existing model or continue learning a neural network as information is fed back from carrier sensors. Distributed, embedded, online machine learning is currently a fast-developing subject in the academic world and in R&D structures. Whether in inference or in learning, these applications require the implementation of adapted mechanisms so that raw and semantic information is routed between carriers, from sensors to processing units, from processing units to decision systems, while avoiding heterogeneous subsystems that may be crossed. These mechanisms must also accommodate a volatile environment, opportunistic communications and reconfiguration of objectives during the mission. In this respect, SCAF requires a continuity of digital services between systems.

Application 2: Early warning systems for disaster risk reduction

Earthquakes cause substantial loss of life and environmental damage in areas hundreds of kilometers from their origin. These large ground movements often lead to hazards such as tsunamis, fires, and landslides. To mitigate the disastrous effects number of earthquake early warning systems have been proposed. These critical systems, operating 24 hours a day, 7 days a week, are supposed to automatically detect and characterize earthquakes as they occur, and issue alerts before ground motion reaches sensitive areas so that protective measures can be taken. It is essential for such a system to detect all large earthquakes with 100% accuracy because the decisions following a large earthquake warning involve important measures for the potentially affected population.

This type of detection can be likened to a classification problem, where the input is sensor data and the output is a class (normal activity / medium / large earthquake). Recent machine learning approaches designed to combine large volumes of data from multiple data sources can be applied. The challenge remains the integration and real-time processing of data streams high-speed data from multiple sensors scattered over a large area. A traditional centralized approach that transfers all data to a single point may be impractical. Thus, detection solutions based on distributed machine learning and relying on high-performance computing techniques and equipment are needed to enable real-time alerts.

Assignment

Objectives

The "computing continuum", i.e., all the computing resources from the periphery (Edge computing) to Cloud or High-Performance Computing (HPC) infrastructures, is a recent research focus [4, 3, 1, 2]. The

deluge of data generated by the multiplication of geo-distributed systems encourages the execution of calculations and data processing locally in order to minimize the cost of data movements between infrastructure components. In addition, there are constraints on the reactivity of systems: for example, an autonomous vehicle must make a decision within a time window that is incompatible with a request in a cloud. An additional constraint is data security, which may require that data does not leave a subsystem. From then on, processing can be done on various systems (IoT, Edge, Fog, Cloud, HPC) depending on operational constraints. One approach is to describe data-driven workflows, deployed dynamically and opportunistically on available resources. However, to our knowledge, there is no "computing continuum" approach that takes into account the high volatility of equipment as well as online modifications of mission objectives. The application of the "computing continuum" in the defense domain is in itself an innovative subject. The objective of this thesis is to identify and adapt emerging approaches from the "computing continuum" in order to address the issues of distribution of calculations and processing, particularly in the case of workflows involving AI. This exploratory thesis has concrete applications such as the SCAF project or on civil warning systems and is situated upstream of these to help guide future technical choices. The steps foreseen for the development of this thesis are:

1. Investigate and characterize different middleware for managing shared computations and data on HPC, Cloud, Fog, Edge systems, in centralized (client-server), decentralized (peer-to-peer), or hybrid configurations.
2. Identify the communicating systems present in the use cases presented in Section 2 and characterize them according to the reading grid established previously.
3. Survey emerging scientific work on "continuum computing", select and propose adaptations for relevant contributions in the case of data intensive workflows.
4. Design or select realistic and relevant operational scenarios involving remote AI processing tasks (learning and inference), in the presence of volatility and dynamic reconfiguration. Formalize the objectives to be achieved and the constraints to be respected for each scenario.
5. Design and implement a high-level prototyping tool to build, in a simplified way, a logical system for task orchestration and shared data management based on several subsystems from peer-to-peer, cloud, fog, edge and HPC contexts [7]. This tool should:
 - Functionally demonstrate interoperability and continuum between heterogeneous computing systems;
 - allow to replay the scenarios established in the previous point;
 - be able to interface with existing simulation tools (those of DGA MI within the framework of SCAF for example);
 - allow to give some tracks or preliminary evaluations on the technical solutions proposed by the industrialists in case of uses presented;
6. Conducting experiments, disseminate results in international publications and writing a thesis manuscript for the successful completion of the defense and graduation.

References

1. The edge: A powerful platform within the compute continuum. <https://www.arm.com/blogs/blueprint/forrester-edge-compute-continuum>. Accessed: 2021-03-30.
2. European commission. building an ecosystem where iot, edge and cloud converge towards a computing continuum. Shaping europe's digital future, event report, Oct.2020. <https://ec.europa.eu/digital-single-market/en/news/building-ecosystem-where-iot-edge-and-cloud-converge-towards-computing-continuum>. Accessed: 2021-03-30.
3. Moustafa AbdelBaky, Mengsong Zou, Ali Reza Zamani, Eduard Renart, Javier Diaz-Montes, and Manish Parashar. Computing in the continuum: Combining pervasive devices and services to support data-driven applications. In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pages 1815-1824. IEEE, 2017.
4. Daniel Balouek-Thomert, Eduard Gibert Renart, Ali Reza Zamani, Anthony Simonet, and Manish Parashar. Towards a computing continuum: Enabling edge-to-cloud integration for data-driven workflows. The International Journal of High Performance Computing Applications, 33(6):1159-1174, 2019.
5. Ministry of the Armed Forces. Press kit scaf, what is it about? https://www.defense.gouv.fr/salle-de-presse/dossiers-de-presse/dossier-de-presse_scaf-de-quoi-s-agit-il. Accessed: 2021-03-30.
6. Ronan Le Gleut and Hélène Conway-Mouret. 2040, l'odyssée du scaf - le système de combat aérien du futur. travaux parlementaires, rapport d'information, juillet 2020. <http://www.senat.fr/rap/r19-642/r19-642.html>. Accessed: 2021-03-30.

7. Daniel Rosendo, Pedro Silva, Matthieu Simonin, Alexandru Costan, and Gabriel Antoniu. E2clab: Exploring the computing continuum through re-peatable, replicable and reproducible edge-to-cloud experiments. In 2020 IEEE International Conference on Cluster Computing (CLUSTER), pages 176-186. IEEE, 2020.

Skills

- An excellent master's degree in computer science or equivalent
- Knowledge of distributed systems and machine learning
- Knowledge of storage and (distributed) file systems
- Ability and motivation to conduct high quality research, including publication of results in high impact conferences and journals
- Strong programming skills (C/C++, Python)
- Professional experience in Big Data, Cloud Computing, AI and/or HPC is an advantage
- Very good communication skills in English (oral and written)
- Open-mindedness, strong integration capacity and team spirit

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport

Remuneration

monthly gross salary amounting to 1982 euros for the first and second years and 2085 euros for the third year

General Information

- **Theme/Domain** : Distributed and High Performance Computing System & Networks (BAP E)
- **Town/city** : Rennes
- **Inria Center** : [Centre Inria de l'Université de Rennes](#)
- **Starting date** : 2021-10-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2021-06-09

Contacts

- **Inria Team** : [KERDATA](#)
- **PhD Supervisor** :
Antoniou Gabriel / gabriel.antoniu@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 will have to show motivation, autonomy and an ability to initiate links between the research activities carried out at INRIA and the business activities carried out at DGA MI.
- Citenzip condition to apply: EU, UK or Switzerland.

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

For more information, please contact francois.tessier@inria.fr or gabriel.antoniou@inria.fr or alexandru.costan@inria.fr

In parallel to the online submission on the Inria web site, please send an email with a cover letter, CV, contact address of at least two references (work collaborator, internship advisor, teacher in a related field, ...) and copies of degree certificates of the latest two years of study to Gabriel Antoniu (gabriel.antoniou@inria.fr), Alexandru Costan (alexandru.costan@inria.fr) and François Tessier (francois.tessier@inria.fr).

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.