Offer #2024-07736

PhD Position F/M - Robust storage on DNA

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.

Assignment

Supervisors:

- Aline ROUMY (aline.roumy@inria.fr)
- Thomas MAUGEY (thomas.maugey@inria.fr)

Goal The goal of the project is to develop an algorithm to allow robust storage data on DNA.

Context Data volume growth has led to a projected data storage requirement of 175 ZB by 2025 [1]. However, the actual data storage capacity currently falls short of this forecast. One potential solution to address these challenges is DNA storage as it offers several advantages, including high data density, extended retention, and low energy cost [2]. Indeed, in terms of data density, DNA can store about bytes per , enabling the storage of all data generated throughout human history within a 30 cm-sided cube [3]. Regarding retention, DNA can endure for centuries, in contrast to contemporary storage mediums that typically last for decades [3]. Additionally, DNA storage is energy-efficient, since it can be stored at ambient temperatures, if it is kept away from light and humidity.

Challenges and envisaged approach Nonetheless, making DNA an efficient storage solution involves overcoming numerous challenges. These challenges encompass:

(i) Data Transformation: convert data into a quaternary code (ACGT).

(ii) DNA Synthesis: write data, essentially synthesizing DNA.

(iii) DNA Sequencing: extract the quaternary code from DNA, i.e., sequencing DNA.

(iv) Data Retrieval: transform back the read quaternary code into the original data.

The goal of the project is to address the first and fourth challenges by developing compression algorithms that are robust to sequencing errors that occur during step (iii). Indeed, efficient DNA storage heavily relies on rapid sequencing methods, which introduce errors. For instance, real time analysis has been achieved at the price of increased error rates with nanopore sequencing, developed by Oxford Nanopore Technologies (ONT). The main difficulty comes from the type of errors: nanopore introduces not only conventional substitution errors but also unconventional deletion and insertion errors [4-5]. Deletion differs from erasure errors, where it is known which part is missing (e.g., lost packets on the internet can be identified by packet headers). Such knowledge of the existence and position of the missing part is unavailable for deletions, and this complicates the correction of this type of error.

The goal of the project is to propose novel ways to structure the compressed DNA-stream in order to robustify nanopore sequencing. For instance, we will exploit the similarities between sequencing and network transmission, to develop robust compression, based on ideas from packet scheduling for noisy networks (for example Dynamic Adaptive Streaming over HTTP). However, there are also differences between network transmission and nanopore sequencing. One of the main differences between the two problems is the random position of the extracted DNA-segment. To address this issue, we will build upon ideas of random extraction in the compressed streams [7], but also on the compressive sensing framework [6].

Bibliography

[1] David Reinsel-John Gantz-John Rydning, John Reinsel, and John Gantz. The digitization of the world from edge to core.Framingham: International Data Corporation, 16:1–28, 2018.

[2] Luis Ceze, Jeff Nivala, and Karin Strauss. Molecular digital data storage using DNA. NatureReviews Genetics, 20(8):456–466, 2019.

[3] Victor Zhirnov, Reza M Zadegan, Gurtej S Sandhu, George M Church, and William L Hughes.Nucleic acid memory. Nature materials, 15(4):366–370, 2016.

[4] Delahaye, Clara, and Jacques Nicolas. "Nanopore MinION Long Read Sequencer: An Overview of Its Error Landscape," November 23, 2020. https://hal.inria.fr/hal-03123133.

[5] ———. "Sequencing DNA with Nanopores: Troubles and Biases." PLoS ONE, October 1, 2021, 1. https://doi.org/10.1371/journal.pone.0257521.

[6] Huo, Dongming, Xuehua Zhu, Guangzhen Dai, Huicheng Yang, Xin Zhou, and Minghui Feng. "Novel Image Compression–Encryption Hybrid Scheme Based on DNA Encoding and Compressive Sensing." Applied Physics B 126, no. 3 2020.

[7] T. Maugey, A. Roumy, E. Dupraz and M. Kieffer. ``Incremental coding for extractable compression in the context of Massive Random Access", IEEE Transactions on Signal and Information Processing over Networks, 2020

Skills

Candidate profile The candidate should have

- strong background in image/signal processing, optimization and programming,
- notions of source coding, information theory would be appreciated.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs

Remuneration

monthly gross salary amounting to 2100 euros

General Information

- **Theme/Domain :** Vision, perception and multimedia interpretation
  Information system (BAP E)
- **Town/city :** Rennes
- **Inria Center :** Centre Inria de l'Université de Rennes
- **Starting date :** 2024-11-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2024-07-01

Contacts

- **Inria Team :** SIROCCO
- **PhD Supervisor :**
  Roumy Aline / aline.roumy@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.

> **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.