AVIGNON UNIVERSITÉ

PROPOSITION SUJETS DE THESES

CONTRATS DOCTORAUX 2020-2023

Appel ciblé (merci de cocher la case correspondante):

- ✓ **Contrat doctoral ministériel ED 536**
- □ **Contrat doctoral ministériel ED 537**

Directeur de thèse : *Francesco DE PELLEGRINI,* Laboratoire d'Informatique d'Avignon (LIA)

Co-encadrant éventuel : *Rosa FIGUEIREDO,* Laboratoire d'Informatique d'Avignon (LIA), France
Lin CHEN, School of Computer Science and Engineering, Sun Yat-sen University, China.

Titre en français : Apprentissage en ligne et ordonnancement des coflows dans les réseaux de centres de données

Titre en anglais : Online Learning and Scheduling of Coflows in Datacenter Networks

Mots-clés : coflows, scheduling, rate control, online learning, datacenters

Co tutelle : Non **Pays :**

Opportunités de mobilité à l'international du doctorant dans le cadre de sa thèse / International mobility opportunity during the thesis: yes

Profil du candidat/Candidate profile : the position is intended for a highly motivated applicant holding a master degree in Computer Science or Applied Mathematics or Theoretical Physics. Key requirement is proficiency in mathematical programming and operations research; theoretical background in approximation algorithms and/or machine learning is considered a plus. Written communication skills (English) is a prerequisite. Applying candidates will prove and/or justify the requested knowledge and skills by providing:

- A full curriculum vitae;
- A motivation letter consistent with the proposed PhD project;
- All documents attesting the requested skills and knowledge;

- Academic records and marks of the two years of Master, or of the two last years of Engineer school;
- One or two recommendation letters are considered a plus.

Required skills: the ideal candidate should have solid background in applied computer science and applied mathematics. Her/his curriculum should prove excellent scholar records in relevant subjects such as, e.g., operations research, control theory, probability theory and applications to queueing theory, performance evaluation, statistical mechanics, mathematical programming and algorithmic design. Familiarity with SDN controller technology, datacenter routing and of Openflow is considered a plus but not mandatory.

Présentation détaillée du sujet / Detailed description : the Laboratoire Informatique d'Avignon (LIA) is seeking a highly motivated Ph.D. candidate in the field of traffic engineering for future datacenter networks. The candidate will work on the optimization of scheduling of coflows. The activity of the candidate will involve the modelling of the problem, its mathematical formulation for the resulting resources optimization problems and the development of efficient solutions by means of exact methods and approximation algorithms. The focus of this thesis requires a background in mathematical modeling, distributed optimization, performance evaluation and algorithmics applied to computer networks. Interested candidates who want to have further details on the thesis' subject and workplace laboratory, can contact directly the following members of the LIA Department

- Prof. Francesco de Pellegrini; email: francesco.de-pellegrini@univ-avignon.fr
- Prof. Rosa Figueiredo; email: rosa.figueiredo@univ-avignon.f

Domaine - Thématique / Research Domain : Operational Research and Communication Networks

Objectif / Objective. This Ph.D. position is aimed at the development of advanced algorithms and optimization frameworks to improve the completion time of coflows generated at runtime. The research project will start with the characterization of the coflow traffic sources in datacenter networks and, using suitable modelling and algorithmic tools, it will aim at improving the performance and fairness footprint of scheduling decisions operated by switches and network controllers. Overall, this PhD position is to address the following key objectives:

- A modelling framework for the optimization of coflows scheduling in datacenters;
- Mathematical modelling and Operation Research techniques to provide efficient offline solutions;
- Design of algorithms for online coflow scheduling, combining scheduling and online learning techniques.

Contexte et enjeux / Context and challenges: the focus of this thesis is the development of new mathematical models for coflows scheduling in datacenters, where the core issue is the presence of competing coflows, thus requiring optimizing routing and rate allocation decisions. Roughly speaking, at the network level a coflow is any series of flows that have common sets of sources and destinations. At the application level, coflows are generated under various frameworks for distributed computation, e.g., in Hadoop Map-Reduce, where consecutive computation stages are interleaved with communication, creating swarms of flows across server racks within the datacenter traffic. Coflows represent very specific traffic sources requiring to rethink the flow scheduling logic in order to be able to allocate flow routes over the datacenter fabric towards a global optimization objective. At the same time, several technological constraints are to be accounted for with respect to the datacenters' network fabric, e.g., with respect to the network topology and in terms of control channel requirements for SDN controllers and computing capabilities of the fabric switches. The classic problem of scheduling flows in datacenter networks has been described in several works appeared in the literature [1][2][3]. The focus of many works in this context are inter-coflow scheduling techniques, which have been proved to improve the performance figures of big data computing frameworks. Several heuristics have been proposed in order to optimize rate allocation for flows composing concurrent coflows [2][3]. Recently, solutions to minimize the weighted coflow completion time appeared in literature. They are based on scheduling techniques originally developed in the context of machine scheduling [3][4][5]. Both in the deterministic [3] and in the stochastic scheduling case [4][5], the coupling effects due to routing in the datacenter fabric render coflow scheduling a challenging problem. Approximation algorithms are typically proposed under strong assumptions on the apriori information available on the size and arrival time of composing flows. In this thesis, we propose to advance the state of the art by designing online solutions where stochastic scheduling techniques are combined with online learning in order to track and predict the size of the flows composing a coflow. Doing so, we can leverage online learning techniques such as stochastic approximations [7] and or reinforcement learning [8].

Références bibliographiques / Bibliography:

[1] M. Chowdhury, Y. Zhong, and I. Stoica, "Efficient Coflow Scheduling with Varys", ACM SIGCOMM, 2014.
[2] RAPIER: Integrating Routing and Scheduling for Coflow-aware Data Center Networks in Proc. of IEEE INFOCOM April 2015
[3] M. Chowdhury, Y. Zhong, and I. Stoica, "Efficient coflow scheduling with varys," in ACM SIGCOMM Computer Communication Review, vol. 44, no. 4. ACM, 2014, pp. 443–454.
[4] M. Shafiee and J. Ghaderi, "Scheduling coflows in datacenter networks: Improved bound for total weighted completion time," ACM SIGMETRICS 2017.

[5] R. Mao, V. Aggarwal, and M. Chiang, “Stochastic non-preemptive co-flow scheduling with time-indexed relaxation,” IEEE INFOCOM, DCPerf Workshop, 2018.
[6] M. Skutella, M. Sviridenko, and M. Uetz, *“Unrelated machine scheduling with stochastic processing times,”* Math. Oper. Res., vol. 41, no. 3, pp. 851–864, 2016.
[7 F. De Pellegrini, L. Maggi, A. Massaro, J. Leguey, and E. Altman. “Blind adaptive and robust flow segmentation in datacenters”, IEEE INFOCOM 2018.
[8] A. Massaro, F. De Pellegrini, and L. Maggi. “Optimal trunk reservation by policy learning,” IEEE INFOCOM 2019.

Les sujets devront être adressés à

gestion-ed@univ-avignon.fr

avant le 6 avril 2021