Research Internship

Céline Comte Gersende Fort

October 16, 2026

Title Reinforcement Learning for Jackson Networks

Supervisors Céline Comte and Gersende Fort

Location LAAS–CNRS, 7 Avenue du Colonel Roche, 31400 Toulouse, France

Keywords Reinforcement learning, Markov decision processes, operations research, queueing theory, Jackson networks, stability, exponential families

Context The internship is at the interface of two research domains: we will design *reinforcement learning (RL)* algorithms to solve challenging problems in *queueing theory*.

Queueing theory is an area at the intersection of applied probability and operations research that studies queueing systems, that is, dynamical systems containing waiting lines. In a nutshell, the idea is to model queueing systems using Markov chains or Markov decision processes (MDPs), in order to predict or optimize their performance. Application areas include management and optimization of computer networks, public transports, road traffic, and supply chains.

In this internship, we will focus on Jackson networks, a model introduced in 1957 [2] that has played a leading role in the development of queueing theory. Jackson networks are celebrated because they capture realistic features while remaining simple enough to be analyzed exactly. Unfortunately, to date, there is no general-purpose algorithm to optimize performance in Jackson networks. This is particularly true for open Jackson networks, where stability (i.e., prevent the queues from growing unboundedly) is a critical issue.

We will develop RL algorithms [12, 10] to optimize performance in open and closed Jackson networks, with a focus on ensuring stability and scalability. Optimization in a Jackson network can be written as an infinite-horizon MDP with an average-reward criterion, which can in principle be optimized using the framework of RL. However, several features of Jackson networks make them challenging for classical RL algorithms: the average-reward criterion is knowingly difficult in RL [13, 8], the state space is infinite or prohibitively large, and in some applications, rewards are extremely infrequent [6].

Objectives We aim to design RL algorithms to optimize performance in Jackson networks, with low memory, time, and sample complexity. The key challenge is to exploit the (relatively) simple structure of Jackson networks to speed up the learning process.

We will consider and/or improve over several learning algorithms, such as SARSA [10, Section 6.4], Q-learning [10, Section 6.5], policy gradient [11, 4, 5, 1], and natural policy gradient [3, 7, 9]. We will run extensive numerical experiments to analyze the performance of these algorithms, study research papers proving theoretical bounds on their convergence speed, and possibly prove new convergence results.

Funding This internship is part of the **LIQUID** project (*Learn to Improve Queues Using Invariant Distributions*), funded by **ANR AAPG 2026 JCJC**, that aims at exploiting product-form invariant distributions to enable scalable control of MDPs with large or infinite state spaces. If the internship goes well, the candidate will have the opportunity to pursue a funded PhD thesis.

Prerequisites The candidate should be able to demonstrate a mathematical background, including one or more courses on probability theory and Markov chains, and proficiency in scientific computing software for prototyping algorithms, such as Python, Julia, or Matlab. Experience in stochastic optimization is also appreciated but not mandatory.

Work environment As an intern, you will receive a "gratification" of approximately 682.52€ per month (the precise value for 2027 hasn't been published yet). The research project will be carried out in the SARA team at LAAS–CNRS, in Toulouse, France. The candidate will also be involved in the activities of the SOLACE team, which gathers researchers working on stochastic modeling at IRIT and LAAS. The SOLACE team organizes a weekly seminar and has established international collaborations with leading academic institutions and industrial labs.

Application procedure Candidates are invited to send their application by email to Céline Comte ⟨celine.comte@laas.fr⟩ and Gersende Fort ⟨gersende.fort@laas.fr⟩, in PDF format. Applications should include a curriculum vitae, a transcript of grades in first and second year of Master's study (possibly incomplete for the second year), a transcript of grades of Bachelor's study if possible, and a short cover letter (approximately 200–400 words).

References

[1] Céline Comte et al. "Score-Aware Policy-Gradient and Performance Guarantees using Local Lyapunov Stability". In: *Journal of Machine Learning Research* 26.132 (2025), pp. 1–74. URL: http://jmlr.org/papers/v26/24-1009.html.

[2] James R. Jackson. "Networks of Waiting Lines". In: *Operations Research* 5.4 (Aug. 1, 1957). Publisher: INFORMS, pp. 518–521. URL: https://doi.org/10.1287/opre.5.4.518.

[3] Sham M Kakade. "A Natural Policy Gradient". In: *Advances in Neural Information Processing Systems*. Vol. 14. MIT Press, 2001. URL: https://proceedings.neurips.cc/paper_files/paper/2001/hash/4b86abe48d358ecf194c56c69108433e-Abstract.html.

[4] Vijay Konda and John Tsitsiklis. "Actor-Critic Algorithms". In: *Advances in Neural Information Processing Systems*. Vol. 12. MIT Press, 1999. URL: https://proceedings.neurips.cc/paper/1999/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

[5] Vijay R. Konda and John N. Tsitsiklis. "On Actor-Critic Algorithms". In: *SIAM Journal on Control and Optimization* 42.4 (Jan. 2003). Publisher: Society for Industrial and Applied Mathematics, pp. 1143–1166. URL: https://epubs.siam.org/doi/abs/10.1137/S0363012901385691.

[6] D. Mastropietro et al. "Fast-exploring reinforcement learning with applications to stochastic networks". In: *Queueing Systems* 109.3 (Sept. 9, 2025), p. 23. URL: https://doi.org/10.1007/s11134-025-09950-5.

[7] Tetsuro Morimura et al. "A New Natural Policy Gradient by Stationary Distribution Metric". In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Walter Daelemans, Bart Goethals, and Katharina Morik. Berlin, Heidelberg: Springer, 2008, pp. 82–97.

[8] Yashaswini Murthy, Mehrdad Moharrami, and R. Srikant. "Performance Bounds for Policy-Based Average Reward Reinforcement Learning Algorithms". In: *Advances in Neural Information Processing Systems* 36 (Dec. 15, 2023), pp. 19386–19396. URL: https://proceedings.neurips.cc/paper_files/paper/2023/hash/3da8e709fa1a7d9e23bee89d3c25b5b4-Abstract-Conference.html.

[9] Johannes Müller and Guido Montúfar. "Geometry and convergence of natural policy gradient methods". In: *Information Geometry* 7.1 (Jan. 1, 2024), pp. 485–523. URL: https://doi.org/10.1007/s41884-023-00106-z.

[10] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Red. by Francis Bach. 2nd ed. Adaptive Computation and Machine Learning series. Cambridge, MA, USA: A Bradford Book, Nov. 13, 2018. 552 pp. URL: https://mitpress.mit.edu/9780262039246/reinforcement-learning/.

[11] Richard S. Sutton et al. "Policy gradient methods for reinforcement learning with function approximation". In: *Proceedings of the 12th International Conference on Neural Information Processing Systems*. NIPS'99. Cambridge, MA, USA: MIT Press, Nov. 29, 1999, pp. 1057–1063. URL: https://proceedings.neurips.cc/paper_files/paper/1999/file/464d828b85b0bed98e80ade0a5c43b0f-Paper.pdf.

[12] Csaba Szepesvári. "Algorithms for Reinforcement Learning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 4.1 (Jan. 2010). Publisher: Morgan & Claypool Publishers, pp. 1–103. URL: https://www.morganclaypool.com/doi/abs/10.2200/S00268ED1V01Y201005AIM009.

[13] Yiming Zhang and Keith W. Ross. "On-Policy Deep Reinforcement Learning for the Average-Reward Criterion". In: *Proceedings of the 38th International Conference on Machine Learning*. International Conference on Machine Learning. ISSN: 2640-3498. PMLR, July 1, 2021, pp. 12535–12545. URL: https://proceedings.mlr.press/v139/zhang21q.html.