Research Internship

Céline Comte Gersende Fort

October 31, 2025 (Updated on November 18, 2025)

Title Reinforcement Learning in Jackson Networks

Supervisors Céline Comte and Gersende Fort

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

Keywords Reinforcement learning, policy gradient, queueing theory, Jackson networks

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[1] is an area at the intersection of applied probability and operations research that studies queueing systems, i.e., 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. Examples of applications of queueing theory include resource management in hospitals, bandwidth sharing on the Internet, scheduling and load balancing in datacenters, and management of bike sharing systems. In this project, we will focus on Jackson networks, a model introduced in 1957 [4] 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; for instance, several algorithms have been proposed to predict the average queue length in Jackson networks [2, 12, 1]. Unfortunately, to date, there is no general-purpose algorithm to optimize performance in Jackson networks.

In this project, we will develop RL algorithms [15, 13] to optimize performance in Jackson networks. RL emerged as a research area in the 1980's and became popular in the 2010's after several noteworthy successes, such as the victory of AlphaGo against a professional Go player in 2015. 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: (i) The average-reward criterion is considered challenging in RL [16, 10]. (ii) The state space is infinite or extremely large (in typical applications, its cardinality grows exponentially with the "size" of the system). (iii) In some applications, rewards are extremely infrequent [8].

Objectives The goal of the project is 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, in particular SARSA [13, Section 6.4], Q-learning [13, Section 6.5], policy gradient [14, 6, 7, 3], and natural policy gradient [5, 9, 11]. We will run extensive numerical results to analyze the performance of these algorithms and, if time permits, we will prove theoretical bounds on their convergence speed.

[1] Spelled either *queueing theory* or *queuing theory*.

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 639,45€ per month. The research project will be carried out in the SARA team at LAAS, in Toulouse, France. The candidate will also be involved into the activities of the SOLACE team, which gathers researchers working on stochastic modeling at IRIT and LAAS. The SOLACE team 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 transcripts of grades in first and second year of Master's study (possibly incomplete for the second year), and a short covering letter (approximately 200–400 words).

References

[1] Ivo Adan and Jan van der Wal. "Mean Values Techniques". In: *Queueing Networks: A Fundamental Approach*. Ed. by Richard J. Boucherie and Nico M. van Dijk. International Series in Operations Research & Management Science. Boston, MA: Springer US, 2011, pp. 561–586. URL: https://doi.org/10.1007/978-1-4419-6472-4_13.

[2] Jeffrey P. Buzen. "Computational algorithms for closed queueing networks with exponential servers". In: *Communications of the ACM* 16.9 (Sept. 1, 1973), pp. 527–531. URL: https://doi.org/10.1145/362342.362345.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] M. Reiser and S. S. Lavenberg. "Mean-Value Analysis of Closed Multichain Queuing Networks". In: *Journal of the ACM* 27.2 (Apr. 1, 1980), pp. 313–322. URL: https://doi.org/10.1145/322186.322195.

[13] 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/.

[14] 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.

[15] 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.

[16] 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.