REFERENCES

[1] Kristopher De Asis, J. Hernandez-Garcia, G. Holland, and Richard S. Sutton. 2018. Multi-Step Reinforcement Learning: A Unifying Algorithm. In *AAAI Conference on Artificial Intelligence*.

[2] Pierre-Luc Bacon, Jean Harb, and Doina Precup. 2017. The Option-Critic Architecture. In *Proceedings of Association for the Advancement of Artificial Intelligence Conference (AAAI)*. 1726–1734.

[3] Kamil Andrzej Ciosek and Shimon Whiteson. 2017. OFFER: Off-Environment Reinforcement Learning. In *Proceedings of Association for the Advancement of Artificial Intelligence Conference (AAAI)*.

[4] Caroline Claus and Craig Boutilier. 1998. The dynamics of reinforcement learning in cooperative multiagent systems. *AAAI/IAAI* (1998), 746–752.

[5] Flaviu Cristian, Bob Dancey, and Jon Dehn. 1996. Fault-tolerance in air traffic control systems. *ACM Transactions on Computer Systems (TOCS)* 14, 3 (1996), 265–286.

[6] Aryeh Dvoretzky. 1956. On Stochastic Approximation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. University of California Press, 39–55.

[7] Jakob N Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[8] Javier Garcıa and Fernando Fernández. 2015. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research* 16, 1 (2015), 1437–1480.

[9] Chris Gaskett. 2003. Reinforcement learning under circumstances beyond its control. In *Proceedings of the International Conference on Computational Intelligence for Modelling Control and Automation*.

[10] Zhang-Wei Hong, Shih-Yang Su, Tzu-Yun Shann, Yi-Hsiang Chang, and Chun-Yi Lee. 2018. A deep policy inference q-network for multi-agent systems. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 1388–1396.

[11] Tommi Jaakkola, Michael I Jordan, and Satinder P Singh. 1994. Convergence of stochastic iterative dynamic programming algorithms. In *Advances in Neural Information Processing Systems*. 703–710.

[12] Richard Klima, Karl Tuyls, and Frans Oliehoek. 2016. Markov Security Games: Learning in Spatial Security Problems. *NIPS Workshop on Learning, Inference and Control of Multi-Agent Systems* (2016), 1–8.

[13] John C Knight. 2002. Safety critical systems: challenges and directions. In *Proceedings of the 24th International Conference on Software Engineering*. ACM, 547–550.

[14] Dmytro Korzhyk, Zhengyu Yin, Christopher Kiekintveld, Vincent Conitzer, and Milind Tambe. 2011. Stackelberg vs. Nash in Security Games: An Extended Investigation of Interchangeability, Equivalence, and Uniqueness. *Journal of Artificial Intelligence Research* 41 (2011), 297–327.

[15] Michael L. Littman. 1994. *Markov games as a framework for multi-agent reinforcement learning*. Technical Report. Brown University. 157–163 pages.

[16] Jing Liu, Yang Xiao, Shuhui Li, Wei Liang, and CL Philip Chen. 2012. Cyber security and privacy issues in smart grids. *IEEE Communications Surveys & Tutorials* 14, 4 (2012), 981–997.

[17] Jian Lou, Andrew M Smith, and Yevgeniy Vorobeychik. 2017. Multidefender security games. *IEEE Intelligent Systems* 32, 1 (2017), 50–60.

[18] Jun Morimoto and Kenji Doya. 2005. Robust reinforcement learning. *Neural computation* 17, 2 (2005), 335–359.

[19] Rémi Munos, Tom Stepleton, Anna Harutyunyan, and Marc Bellemare. 2016. Safe and efficient off-policy reinforcement learning. In *Advances in Neural Information Processing Systems (NIPS)*. 1054–1062.

[20] James Pita, Manish Jain, Janusz Marecki, Fernando Ordonez, Christopher Portway, Milind Tambe, Craig Western, Praveen Paruchuri, and Sarit Kraus. 2008. Deployed ARMOR Protection: The Application of a Game Theoretic Model for Security at the Los Angeles International Airport. In *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Vol. 3. 1805–1812.

[21] Herbert Robbins and Sutton Monro. 1951. A Stochastic Approximation Method. *The Annals of Mathematical Statistics* 22, 3 (1951), 400–407.

[22] Sui Ruan, Candra Meirina, Feili Yu, Krishna R Pattipati, and Robert L Popp. 2005. *Patrolling in a stochastic environment*. Technical Report. Electrical and Computer Engineering Department, University of Connecticut, Storrs.

[23] Eric Shieh, Bo An, Rong Yang, Milind Tambe, Craig Baldwin, Joseph DiRenzo, Ben Maule, and Garrett Meyer. 2012. PROTECT: A Deployed Game Theoretic System to Protect the Ports of the United States. *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* 1 (2012), 13–20.

[24] Martin L Shooman. 2003. *Reliability of computer systems and networks: fault tolerance, analysis, and design*. John Wiley & Sons.

[25] Satinder Singh, Tommi Jaakkola, Michael L Littman, and Csaba Szepesvári. 2000. Convergence results for single-step on-policy reinforcement-learning algorithms. *Machine learning* 38, 3 (2000), 287–308.

[26] Satinder P Singh, Andrew G Barto, Roderic Grupen, and Christopher Connolly. 1994. Robust reinforcement learning in motion planning. In *Advances in Neural Information Processing Systems (NIPS)*. 655–662.

[27] D. R. Smart. 1974. *Fixed point theorems*. Cambridge University Press, Cambridge.

[28] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, and Thore Grapel. 2017. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296* (2017).

[29] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction* (second ed.). The MIT Press, Cambridge, MA.

[30] Richard S Sutton, Doina Precup, and Satinder Singh. 1999. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence* 112, 1-2 (1999), 181–211.

[31] Csaba Szepesvari and Michael L Littman. 1997. *Generalized Markov Decision Processes: Dynamic-programming and Reinforcement-learning Algorithms*. Technical Report. Brown University.

[32] John N Tsitsiklis. 1994. Asynchronous stochastic approximation and Q-learning. *Machine Learning* 16, 3 (1994), 185–202.

[33] Harm van Seijen, Hado van Hasselt, Shimon Whiteson, and Marco Wiering. 2009. A theoretical and empirical analysis of Expected Sarsa. In *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning, ADPRL 2009*. 177–184.

[34] Ye Yan, Yi Qian, Hamid Sharif, and David Tipper. 2013. A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *IEEE Communications Surveys & Tutorials* 15, 1 (2013), 5–20.

[35] Kemin Zhou and John Comstock Doyle. 1998. *Essentials of robust control*. Vol. 104. Prentice hall, Upper Saddle River, NJ.