

Deep Reinforcement Learning for Adaptive Resource Allocation in Secure Wireless Networks with Strategic Eavesdroppers

Anton Fleming

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
a_fleming@ucf.edu

Brent Bush

Department of Computer Science, University of Houston, Houston, TX, USA.
brent838@uh.edu

Zixuanyichen Long

Department of Computer Science, University of North Texas, Denton, TX, USA.
zixuanyichen.long512@unt.edu

Abstract

Secure wireless networks face a fundamental tension between resource efficiency and confidentiality because the broadcast nature of radio propagation creates opportunities for strategic eavesdroppers. Traditional resource allocation methods often assume passive, stochastic, or worst-case adversaries, which limits their ability to cope with intelligent and adaptive interception strategies. Deep reinforcement learning offers a promising framework for learning resource allocation policies directly from high-dimensional and dynamic wireless environments. This paper presents a system-level examination of deep reinforcement learning for adaptive resource allocation in secure wireless networks where eavesdroppers behave strategically. Rather than focusing on algorithmic derivations, the analysis emphasizes architectural choices, adversarial modeling, structural trade-offs, robustness, fairness, infrastructure requirements, sustainability, and governance. The discussion highlights that deep reinforcement learning can capture nonstationary adversarial behavior and support continuous decision-making across power control, beamforming, cooperative jamming, channel selection, and user scheduling. However, practical deployment requires careful consideration of training stability, exploration safety, energy consumption, interpretability, and regulatory accountability. The paper further argues that secure resource allocation cannot be treated solely as an optimization problem because strategic eavesdroppers introduce multi-agent dynamics and incentive misalignment. A layered governance architecture with continuous monitoring, human oversight, and standardized certification is necessary to align adaptive learning systems with broader security and societal goals.

Keywords

deep reinforcement learning, adaptive resource allocation, secure wireless networks, strategic eavesdroppers, physical layer security, adversarial robustness, network governance.

1. Introduction

Wireless communication systems are increasingly expected to support heterogeneous services, dense device connectivity, and mission-critical applications under constrained spectrum and

energy budgets. The broadcast and superposition properties of wireless channels introduce fundamental security risks because unauthorized receivers can observe transmitted signals without necessarily revealing their presence. While cryptographic mechanisms protect higher-layer confidentiality, physical layer security exploits channel properties such as fading, interference, and spatial diversity to reduce the information that an eavesdropper can extract. The dynamic nature of mobility, traffic demand, and interference patterns requires resource allocation that continuously adapts power, bandwidth, beamforming, and cooperative jamming decisions. Sequential decision making under uncertainty provides a natural theoretical foundation for this adaptive behavior [1]. However, conventional optimization approaches often rely on simplified channel models and static threat assumptions, which may fail when adversaries learn and adapt.

Deep reinforcement learning has emerged as a powerful paradigm for sequential decision making in high-dimensional environments. The combination of reinforcement learning with deep neural network function approximation enables agents to learn policies from raw observations without requiring closed-form system models [2]. In wireless networking, deep reinforcement learning has been studied for resource management, spectrum access, mobility management, and edge computing because it can handle complex state spaces and delayed rewards [3]. At the same time, wireless security research has produced extensive analyses of authentication, confidentiality, jamming, and eavesdropping threats [4]. Physical layer security research further demonstrates that secrecy can be enhanced through cooperative transmission, artificial noise, and multi-antenna techniques [5]. These two lines of inquiry have begun to converge, giving rise to learning-based secure resource allocation that treats eavesdroppers as active decision-makers rather than passive channel impairments.

This paper provides a system-level perspective on deep reinforcement learning for adaptive resource allocation in secure wireless networks with strategic eavesdroppers. It does not propose a single algorithm but examines how architectural, operational, and governance decisions shape the effectiveness and trustworthiness of such systems. The analysis covers system architecture and threat modeling, learning architecture and adversarial adaptation, structural trade-offs, robustness and fairness, deployment and sustainability, and governance and policy implications. The central argument is that deep reinforcement learning can enable secure wireless networks to respond to intelligent eavesdropping behavior, but only if the learning system is embedded within a broader socio-technical framework that addresses safety, accountability, and long-term sustainability.

2. System Architecture and Threat Model Considerations

A secure wireless network that supports learning-based resource allocation consists of several interacting components, including base stations, authorized users, cooperative jammers, edge computing nodes, and monitoring infrastructure. The resource allocation problem involves continuous decision variables such as transmit power levels, beamforming weights, and time-frequency resource assignments. Because these decisions must be updated on short timescales, the learning system requires a control architecture that can process high-dimensional observations and produce continuous actions. Policy gradient and actor-critic methods have proven useful for continuous control tasks because they can learn stochastic or deterministic policies without discrete action enumeration [6]. In secure wireless settings, this supports fine-grained power adaptation and jamming coordination that would be difficult to achieve with rule-based mechanisms.

Resource management in dynamic networks has also been formulated as a learning problem in which a central controller observes traffic queues, channel states, and security indicators before selecting resource allocations [7]. The presence of strategic eavesdroppers complicates this formulation because the environment no longer evolves according to a stationary stochastic process. A strategic eavesdropper may observe the defender's allocation patterns, estimate the underlying policy, and adjust its interception strategy to maximize the amount of leaked information. This creates a form of adversarial nonstationarity that cannot be fully captured by standard Markov decision process models unless the adversary is included in the state representation or treated as a learning agent. Consequently, the system architecture must incorporate threat detection, adversary behavior estimation, and continuous policy updating as interconnected functions.

The system-level threat model should distinguish between passive eavesdroppers that only listen, adaptive eavesdroppers that modify their positions or reception strategies, and strategic adversaries that may use learning algorithms to infer defensive patterns. Cooperative jamming nodes can degrade the signal quality at potential eavesdropper locations, but the effectiveness of jamming depends on accurate channel state information and the ability to avoid harming authorized users. The infrastructure must therefore balance the information gained from channel estimation and adversary localization against the signaling overhead and potential exposure of defensive strategies. A centralized controller can achieve globally coordinated decisions, but it may become a bottleneck and a single point of failure. Distributed or hierarchical architectures can improve scalability and resilience, yet they introduce coordination overhead and partial observability. The design of the observation space, reward signals, and action interfaces must account for these architectural choices to ensure that learned policies align with system-level security objectives.

3. Deep Reinforcement Learning for Adaptive Decision-Making

Deep reinforcement learning combines sequential decision making with representation learning, allowing a control agent to discover useful features from high-dimensional sensory data. In secure wireless resource allocation, the state may include channel estimates, queue lengths, user mobility, detected eavesdropping activity, and historical transmission outcomes. The action space may consist of transmit power adjustments, beamforming configurations, cooperative jamming levels, and resource block assignments. The reward signal typically combines secrecy performance, throughput, energy consumption, and quality-of-service indicators. The learning objective is to adjust the policy so that cumulative reward improves over time despite the dynamic and partially observable nature of the network. Actor-critic architectures are particularly relevant because they maintain a separate estimate of the value function to reduce the variance of policy updates.

The strategic eavesdropper challenges the assumption that the environment is stationary. Cooperative jamming against an intelligent eavesdropper over parallel channels demonstrates that power allocation must account for the eavesdropper's ability to respond to defensive actions [8]. In such settings, an eavesdropper may selectively attack channels where the defender has allocated less protective power, thereby requiring the defender to learn a policy that anticipates adversarial adaptation. This can be understood as a leader-follower interaction, but the follower may not be perfectly rational or fully observable. Deep reinforcement learning is well suited to this challenge because it can learn defensive behavior through repeated interaction without needing an explicit model of the eavesdropper's objective, provided that the training process includes sufficient adversarial diversity.

Real-world deployment of deep reinforcement learning faces additional challenges related to sample efficiency, delayed rewards, safety constraints, and the gap between simulation and physical environments [9]. Training a secure resource allocation agent solely in simulation may produce policies that overfit to idealized channel models and predictable adversarial behaviors. Domain randomization, adversarial training, and periodic online fine-tuning can improve robustness, but they also increase computational cost and risk. Self-play mechanisms, in which the defender and an adversary model are trained against each other, can generate increasingly sophisticated strategies [10]. However, self-play may lead to policies that are specialized to a particular adversarial style and less effective against novel eavesdropping techniques. A resilient system therefore requires a portfolio of training strategies, including scripted adversaries, learned adversary models, and real-world monitoring data.

4. Structural Trade-offs in Secure Resource Allocation

Secure resource allocation cannot be reduced to a single objective. Increasing transmit power may improve authorized throughput while also improving the signal quality for an eavesdropper unless additional jamming or beamforming is used. Allocating more energy to cooperative jamming can reduce information leakage but may increase interference, shorten battery life, and violate regulatory emission limits. Learning-based resource management in ultra-reliable and low-latency vehicular networks illustrates how throughput, latency, and reliability interact under stringent operational constraints [11]. In secure settings, these trade-offs become even more pronounced because secrecy measures can consume resources that would otherwise support legitimate traffic. Deep reinforcement learning provides a mechanism for exploring these trade-offs through reward shaping, but the precise weighting of competing objectives remains a system design decision with ethical and operational consequences.

The stability of resource allocation policies is another important structural consideration. Scheduling and resource allocation in multihop radio networks have long been studied through the lens of throughput stability and queue management [12]. A learning-based controller that aggressively reallocates resources to improve short-term secrecy may destabilize queues, increase handover frequency, or create oscillations that degrade overall network performance. Information-theoretic perspectives on wireless security further show that secrecy depends on the relative advantage of the legitimate channel over the eavesdropper channel [13]. This relative advantage is itself affected by power allocation, spatial diversity, and cooperative transmission. A deep reinforcement learning agent must therefore learn to maintain an appropriate margin of secrecy without sacrificing the operating stability of the broader network.

Centralized and distributed architectures embody different structural trade-offs. A centralized agent can maintain a global view of the network and coordinate multiple base stations and jammers to achieve strong secrecy outcomes. However, centralized control increases the amount of state information that must be collected and processed, and it creates a high-value target for cyberattacks. Distributed agents can respond quickly to local conditions and reduce the burden on the core network, but they may make conflicting decisions that inadvertently create coverage gaps or interference hotspots. Hierarchical learning systems may combine global coordination with local adaptation, but they introduce additional training and communication complexity. The choice of architecture also affects fairness, because centralized controllers can enforce network-wide fairness policies more directly, whereas distributed learning may amplify local biases or resource competition.

5. Robustness, Fairness, and Adversarial Dynamics

Robustness is a central requirement for secure resource allocation because strategic eavesdroppers can intentionally exploit predictable patterns in the defender's policy. Multi-agent deep reinforcement learning has advanced the understanding of nonstationary environments in which multiple learning agents interact [14]. In secure wireless networks, the defender and eavesdropper can be modeled as competing learners with partially conflicting objectives. Cooperative jamming techniques in multiple-input multiple-output relay networks have demonstrated the importance of spatial coordination for secrecy [15]. When such coordination is learned through deep reinforcement learning, the resulting policy may be sensitive to small changes in channel conditions or adversary behavior if the training distribution is too narrow. Robustness therefore requires training across diverse channel realizations, mobility patterns, and eavesdropper strategies.

Fairness adds another dimension to the robustness challenge. Networks must provide acceptable service to users with different channel conditions, mobility profiles, and security requirements. A reward function that prioritizes aggregate secrecy may allocate excessive protective resources to users with already favorable channels while neglecting edge users. End-to-end learning approaches can produce policies that optimize a given objective but may hide unintended biases in the learned feature representations [16]. To prevent unfair resource allocation, system designers must monitor per-user secrecy performance, throughput, and energy consumption. Fairness constraints can be embedded into the learning objective or enforced through post hoc policy adjustments, but each approach has limitations. Embedding fairness into the reward function may require arbitrary weighting decisions, while post hoc adjustments may degrade the overall optimality of the learned policy.

The historical understanding of physical layer security has evolved from the wiretap channel model to complex multi-user and multi-antenna scenarios [17]. Strategic eavesdropping introduces a qualitatively different challenge because the adversary may use learning, game-theoretic reasoning, or data-driven inference. The defender's policy must avoid becoming predictable while still maintaining stable service. This may require deliberate randomization, periodic policy perturbation, or ensemble methods that select among multiple trained policies. At the same time, excessive randomization can harm legitimate users and complicate network management. Balancing predictability for legitimate service with unpredictability for adversaries is a structural trade-off that should be explicitly designed rather than left as an emergent property of the learning process.

6. Infrastructure, Deployment, and Sustainability

Deploying deep reinforcement learning for secure resource allocation requires significant infrastructure support. Edge computing nodes, software-defined radio platforms, network monitoring tools, and secure data pipelines must operate together to provide timely observations and policy updates. Machine learning techniques have already been applied to security tasks in Internet of Things environments, demonstrating that resource-constrained devices can participate in learning-based security architectures when computational tasks are offloaded appropriately [18]. However, secure wireless networks often involve strict latency requirements and limited backhaul capacity, which constrain the frequency and complexity of central policy updates. A sustainable deployment strategy must therefore consider where inference occurs, how data are collected, and how models are retrained without disrupting active services.

Energy consumption is a critical sustainability concern. Deep reinforcement learning training can be computationally intensive, especially when large state and action spaces require extensive simulation. Although online inference is typically less expensive than training, continuous policy updates and exploration may still increase the energy footprint of base stations and edge nodes. Cooperative jamming consumes additional power and contributes to electromagnetic interference. Sustainable secure networks must balance the energy cost of secrecy measures against their security benefits. Techniques such as periodic retraining, model compression, asynchronous updates, and edge inference can reduce energy demands, but they may also reduce policy responsiveness or robustness.

The operational life cycle of a learning-based secure resource allocation system includes data collection, simulation, training, validation, deployment, monitoring, and decommissioning. Each stage has distinct risks. During data collection, privacy and security concerns arise because channel state information and user mobility data may be sensitive. During simulation and training, the gap between simulated and physical environments can produce policies that fail in deployment. During operation, concept drift and adversarial adaptation require ongoing monitoring and intervention. Sustainable deployment therefore depends on organizational capacity, technical documentation, reproducible evaluation protocols, and clear ownership of model updates. Systems that cannot be operated and maintained by existing network engineering teams may become brittle despite their theoretical advantages.

7. Governance and Policy Implications

The deployment of autonomous resource allocation in secure wireless networks raises governance questions that extend beyond technical optimization. Deep reinforcement learning policies can make decisions that affect user privacy, service quality, spectrum use, and even national security. Ethical frameworks for artificial intelligence emphasize the need for beneficence, non-maleficence, autonomy, justice, and explicability [19]. These principles suggest that secure resource allocation systems should be designed with clear accountability, transparent evaluation criteria, and mechanisms for contesting automated decisions. Regulatory approaches in different jurisdictions vary, and comparative analyses have shown that the governance of artificial intelligence remains fragmented across the United States, the European Union, and the United Kingdom [20]. Wireless network operators must navigate this landscape while ensuring that security automation does not undermine legal obligations or public trust.

A layered governance architecture is appropriate for learning-based secure resource allocation. At the technical layer, policies should be tested against adversarial scenarios, stress tests, and fairness audits before deployment. At the organizational layer, network operators should assign responsibility for model updates, incident response, and stakeholder communication. At the regulatory layer, standards bodies and spectrum authorities can define minimum requirements for robustness, interference management, and security assurance. Human oversight is necessary because deep reinforcement learning systems can fail in ways that are difficult to predict. Oversight may take the form of approval gates for major policy changes, anomaly detection on learned behavior, and periodic review of reward function weights.

Policy implications also extend to international coordination and spectrum governance. Cooperative jamming can affect neighboring networks and raise cross-border interference concerns if not carefully managed. Learning-based power control may produce behaviors that are difficult to characterize using traditional compliance testing. Therefore, certification processes may need to evolve from static rule checking to continuous behavioral monitoring

and simulation-based auditing. Secure wireless networks are part of critical infrastructure, and their automation must be subject to resilience requirements, cybersecurity safeguards, and supply chain transparency. A governance framework that combines technical robustness with institutional accountability can help ensure that deep reinforcement learning contributes to secure and equitable wireless connectivity.

8. Conclusion

Deep reinforcement learning offers a powerful paradigm for adaptive resource allocation in secure wireless networks where strategic eavesdroppers learn and adapt. The ability to learn policies from high-dimensional channel, traffic, and security observations enables networks to manage power, jamming, beamforming, and scheduling in ways that static optimization cannot easily achieve. However, the presence of strategic adversaries transforms the resource allocation problem into a dynamic and partially observable interaction that demands robustness, exploration safety, and ongoing policy adaptation. This paper has argued that system-level analysis is essential for understanding the structural trade-offs among secrecy, throughput, energy, latency, fairness, and stability.

The deployment of such systems depends on more than algorithmic performance. Infrastructure choices, data governance, energy sustainability, regulatory compliance, and human oversight shape whether learning-based security measures can be trusted in operational networks. Fairness must be monitored across users and services, while robustness must be evaluated against diverse and adaptive eavesdropper strategies. The required reference to cooperative jamming against a strategic eavesdropper over parallel channels illustrates the importance of adversarial adaptation in secure resource allocation [8]. Future research should continue to integrate learning architectures with system engineering, governance, and policy perspectives. A layered approach that combines technical innovation with institutional accountability is most likely to produce secure wireless networks that are both resilient and socially responsible.

References

1. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
2. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.
3. Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y.-C., & Kim, D. I. (2019). Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Communications Surveys & Tutorials*, 21(4), 3133–3174.
4. Zou, Y., Zhu, J., Wang, X., & Hanzo, L. (2016). A survey on wireless security: Technical challenges, recent advances, and future trends. *Proceedings of the IEEE*, 104(9), 1727–1765.
5. Mukherjee, A., Fakoorian, S. A. A., Huang, J., & Swindlehurst, A. L. (2014). Principles of physical layer security in multiuser wireless networks: A survey. *IEEE Communications Surveys & Tutorials*, 16(3), 1550–1573.

6. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. In *Proceedings of the International Conference on Learning Representations*.
7. Mao, H., Alizadeh, M., Menache, I., & Kandula, S. (2016). Resource management with deep reinforcement learning. In *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, 50–56.
8. Xu, Z., & Baykal-Gürsoy, M. (2022). Power allocation for cooperative jamming against a strategic eavesdropper over parallel channels. *IEEE Transactions on Information Forensics and Security*, 18, 846–858.
9. Dulac-Arnold, G., Mankowitz, D., & Hester, T. (2019). Challenges of real-world reinforcement learning. *arXiv preprint arXiv:1904.12901*.
10. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354–359.
11. Yang, H., Xie, X., & Kadoch, M. (2019). Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks. *IEEE Transactions on Vehicular Technology*, 68(5), 4157–4169.
12. Tassiulas, L., & Ephremides, A. (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37(12), 1936–1948.
13. Bloch, M., Barros, J., Rodrigues, M. R. D., & McLaughlin, S. W. (2008). Wireless information-theoretic security. *IEEE Transactions on Information Theory*, 54(6), 2515–2534.
14. Nguyen, T. T., Nguyen, N. D., & Nahavandi, S. (2020). Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Transactions on Cybernetics*, 50(9), 3826–3839.
15. Huang, J., & Swindlehurst, A. L. (2012). Cooperative jamming for secure communications in MIMO relay networks. *IEEE Transactions on Signal Processing*, 60(9), 4871–4884.
16. Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(1), 1334–1373.
17. Wyner, A. D. (1975). The wire-tap channel. *Bell System Technical Journal*, 54(8), 1355–1387.
18. Xiao, L., Wan, X., Lu, X., Zhang, Y., & Wu, D. (2018). IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? *IEEE Signal Processing Magazine*, 35(5), 41–49.
19. Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.

20. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the Good Society: The US, EU, and UK approach. *Science and Engineering Ethics*, 24(2), 505–528.