© Springer International Publishing AG, part of Springer Nature 2018
Xuemin (Sherman) Shen
,
Xiaodong Lin
and
Kuan Zhang
Encyclopedia of Wireless Networks
10.1007/978-3-319-32903-1_71-1

Reinforcement Learning-Based Wireless Communications Against Jamming and Interference

Liang Xiao[1]
(1)Department Communication Engineering, Xiamen University, Xiamen, China

Liang Xiao
Email: lxiao@xmu.edu.cn

Without Abstract

Definition

Learning-based anti-jamming communication strategy applies reinforcement learning algorithms for mobile users in wireless networks to achieve the optimal transmission policy against jamming and interference, without knowing the network model, radio channel model, and jamming model.

Historical Background

Due to the broadcast nature of radio propagation, wireless networks are vulnerable to jamming attacks, as jammers purposefully inject replayed or faked signals into wireless media to interrupt the ongoing radio transmissions between legitimate users (Xu et al., *2005*; Xiao, *2015*). With the pervasion of smart and programmable radio devices such as universal software radio peripherals (USRPs) (Rahbari et al., *2016*), smart jammers choose to launch multiple types of attacks, such as eavesdropping and spoofing attacks, and select the jamming power, frequency, and time against the ongoing wireless transmissions (Trappe, *2015*; Xiao et al., *2018a*). Smart jammers can even analyze the ongoing anti-jamming transmission policy and induce the mobile devices to use a specific communication mode and then block them accordingly (Akyildiz et al., *2008*; Xiao et al., *2015*). Jamming attackers aim to degrade the communication efficiency of the ongoing transmission, increase power consumption of the radio nodes, and even lead to denial of services (DoS) attacks (Xiao et al., *2012a*).
Traditional anti-jamming wireless communication solutions, spread spectrum techniques, such as frequency hopping and direct-sequence spread spectrum, have been used for decades to address jamming attacks in wireless networks (Xu et al., *2005*; Wang et al., *2012*; Xiao et al., *2013*). However, significant challenges have to be addressed in 5G systems, e.g., spread spectrum technique requires both the transmitter and the receiver to share the physical-layer secrets such as the spreading codes or

frequency hopping pattern in advance. However, jammers can derive the spreading codes of the transmitter by eavesdropping the public control channels and compromising cognitive radio nodes in large-scale dynamic wireless network (Attar et al., 2012).

Anti-jamming techniques such as uncoordinated frequency hopping (Liu et al., 2010; Xiao et al., 2012b) have been proposed to address these problems. However, many of these still suffer from low communication efficiency and high-energy cost in dynamic wireless networks (Li et al., 2010, 2012; Xiao et al., 2012c). In wireless communications, a mobile device has difficulty obtaining the network and jamming model, such as the channel condition, jamming, and interference strength in dynamic environments.

The transmission policy such as the transmit power and channel selection in a dynamic game among the mobile devices, jammers, and interference sources can be viewed as a Markov decision process (MDP) (Xiao et al., 2017, 2018b, c). Reinforcement learning (RL) techniques, such as Q-learning and DQN, can be used to find the optimal strategy in MDP via trial and error (Sutton and Barto, 1998). Therefore, the channel selection based on Q-learning can improve the resistance against jamming (Conley and Miller, 2013), and the channel accessing with Q-learning can address smart jammers in hostile environments (Gwon et al., 2013). Deep reinforcement learning technique, such as DQN, can accelerate the learning speed of mobile device to achieve the optimal anti-jamming strategy, in the environment with a large number of frequency channels (Han et al., 2017).

Foundations

We assume a wireless communication system using frequency hopping over N frequency channels. At time slot k, the mobile device chooses its transmission policy denoted by $x^{(k)} \in \{0, \cdots, N\}$ and sends a message with power $P\ s$ on channel $x^{(k)}$. If $x^{(k)} = 0$, the mobile device leaves the area for better transmission condition. The mobile device observes the state at time k denoted by $\mathbf{s}^{(k)}$, which consists of the presence of the PUs, and the previous SINR, i.e., $\mathbf{s}^{(k)} = [\lambda^{(k-1)}, \mathrm{SINR}^{(k-1)}]$.

The CNN is used to estimate the Q-value, denoted by $Q(\mathbf{s}^{(k)}, x)_{0 \le x \le N}$, for each action. The CNN consists of two convolutional (Conv) layers and two fully connected (FC) layers. The first Conv layer includes 20 filters each with size 3×3 and stride 1, and the second Conv layer has 40 filters each with size 2×2 and stride 1. Both Conv layers use the rectified linear unit (ReLU) as the activation function. The first FC layer involves 180 rectified linear units, while the second FC layer has $N + 1$ units for the action set. The filter weights of the four layers in the CNN at time k are denoted by $\vartheta^{(k)}$.

Let $\varphi^{(k)}$ denote the state sequence at time k, which consists of the current system state and the previous W system state-action pairs, i.e., $\varphi^{(k)} = [\mathbf{s}^{(k-W)}, x^{(k-W)}, \cdots; x^{(k-1)}, \mathbf{s}^{(k)}]$. The state sequence is then reshaped into a 6×6 matrix as the input to the CNN to estimate $Q(\varphi^{(k)}, x | \vartheta^{(k)}), \forall 0 \le x \le N$. The CNN parameters $\vartheta^{(k)}$ are updated at each time slot based on experience replay.

The experience observed by the mobile device is denoted by $\mathbf{e}^{(k)} = [\varphi^{(k)}$, \(x^{(k)},u_s^{(k)}\), $\varphi^{(k+1)}]$, and the memory pool at time k is given by \(\mathscr {D}=\{{\mathbf {e}}^{(1)},\cdots ,{\mathbf {e}}^{(k)}\}\). The experience replay chooses an experience $\mathbf{e}^{(d)}$ from memory pool \(\mathscr {D}\) at random, with $1 \le d \le k$ to update $\vartheta^{(k)}$ according to a stochastic gradient descent (SGD) algorithm. The mean-squared error of the target optimal Q-function value is minimized with minibatch updates, and the loss function is chosen by Mnih et al. (2015) as
$$\displaystyle \begin{aligned} &L(\theta^{(k)})\\ &=\mathbb{E}_{ {\varphi}^{(k)},x,u_s, {\varphi}^{(k+1)}}\left[\Big(R-Q\big({\varphi}^{(k)},x;\theta^{(k)} \big) \Big)^2\right], \end{aligned} $$

(1)

where R is the target optimal Q-function, which is given by $$\displaystyle \begin{aligned} R = u_s+\gamma \max_{x'}Q\left({\varphi}^{(k+1)},x';\theta^{(k-1)}\right). \end{aligned} $$

(2)

The gradient of the loss function with respect to the weights $\vartheta^{(k)}$ is given by $$\displaystyle \begin{aligned} \begin{aligned} &\nabla_{\theta^{(k)}}L(\theta^{(k)})\\ &=\mathbb{E}_{ {\varphi}^{(k)},x,u_s, {\varphi}^{(k+1)}}\Big[R\nabla_{\theta^{(k)}}Q\left({\varphi}^{(k)},x;\theta^{(k)} \right) \Big] \\ &\quad -\mathbb{E}_{ {\varphi}^{(k)},x}\Big[Q\left({\varphi}^{(k)},x;\theta^{(k)} \right)\nabla_{\theta^{(k)}}Q\\ &\qquad \left({\varphi}^{(k)},x;\theta^{(k)} \right) \Big]. \end{aligned} \end{aligned} $$

(3)

This process repeats B times at each time slot, and $\vartheta^{(k)}$ is updated according to the B randomly selected experiences.

By using the ε-greedy algorithm, the mobile device chooses the action $x^{(k)}$ according to the Q-function and $\mathbf{s}^{(k)}$. If $x^{(k)}$ is 0, the mobile device leaves the area. Otherwise, the mobile device transmits the message on channel $x^{(k)}$. The BS estimates the SINR of the device signal and feedbacks it to the mobile device. The mobile device evaluates the utility \(u_s^{(k)}\) based on the SINR and the transmission cost. According to the next state sequence $\varphi^{(k+1)}$, the mobile device stores the new experience \([{\varphi }^{(k)},x^{(k)},u_s^{(k)}, {\varphi }^{(k+1)}]\) in the memory pool \(\mathscr {D}\).

Key Application

The RL-based anti-jamming wireless communication scheme can be applied in mobile communications and Internet of things.

Cross-References

Game Theory in Wireless Network
Interference Characterization and Mitigation
Neural Networks and Artifical Intelligence for Wireless Networks
Wireless Security

Acknowledgements
This work is supported by the National Natural Science Foundation of China under Grant 61671396.

References

Akyildiz IF, Lee WY, Vuran MC, Mohanty S (2008) A survey on spectrum management in cognitive radio networks. IEEE Commun Mag 46(4):40–49
CrossRef

Attar A, Tang H, Vasilakos AV, Yu FR, Leung VCM (2012) A survey of security challenges in cognitive radio networks: solutions and future research directions. Proc IEEE 100(12):3172–3186
CrossRef

Conley WG, Miller AJ (2013) Cognitive jamming game for dynamically countering ad hoc cognitive radio networks. In: Military communications conference, MILCOM 2013-2013 IEEE. IEEE, San Diego, CA, pp 1176–1182

Gwon Y, Dastangoo S, Fossa C, Kung H (2013) Competing mobile network game: embracing antijamming and jamming strategies with reinforcement learning. In: 2013 IEEE conference on communications and network security (CNS). IEEE, National Harbor, MD, pp 28–36

Han G, Xiao L, Poor HV (2017) Two-dimensional anti-jamming communication based on deep reinforcement learning. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing (ICASSP), New Orleans, LA, pp 1–5

Li C, Dai H, Xiao L, Ning P (2010) Analysis and optimization on jamming-resistant collaborative broadcast in large-scale networks. In: 2010 conference record of the forty fourth asilomar conference on signals, systems and computers (ASILOMAR). IEEE, Pacific Grove, CA, pp 1859–1863
CrossRef

Li C, Dai H, Xiao L, Ning P (2012) Communication efficiency of anti-jamming broadcast in large-scale multi-channel wireless networks. IEEE Trans Signal Process 60(10):5281–5292
MathSciNet CrossRef

Liu A, Ning P, Dai H, Liu Y (2010) USD-FH: jamming-resistant wireless communication using frequency hopping with uncoordinated seed disclosure. In: 2010 IEEE 7th international conference on mobile adhoc and sensor systems (MASS). IEEE, San Francisco, CA, pp 41–50

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G (2015) Human-level control through deep reinforcement learning. Nature 518(7540): 529–533
CrossRef

Rahbari H, Krunz M, Lazos L (2016) Swift jamming attack on frequency offset estimation: the achilles heel of OFDM systems. IEEE Trans Mob Comput 15(5): 1264–1278

CrossRef

Sutton R, Barto AG (1998) Reinforcement learning: an introduction. MIT Press, Cambridge

Trappe W (2015) The challenges facing physical layer security. IEEE Commun Mag 53(6):16–20
CrossRef

Wang Q, Xu P, Ren K, Li XY (2012) Towards optimal adaptive UFH-based anti-jamming wireless communication. IEEE J Sel Areas Commun 30(1):16–30
CrossRef

Xiao L (2015) Anti-jamming transmissions in cognitive radio networks. Springer, Cham
CrossRef

Xiao L, Chen Y, Lin WS, Liu KR (2012a) Indirect reciprocity security game for large-scale wireless networks. IEEE Trans Inf Forensics Secur 7(4): 1368–1380
CrossRef

Xiao L, Dai H, Ning P (2012b) Jamming-resistant collaborative broadcast using uncoordinated frequency hopping. IEEE Trans Inf Forensics Secur 7(1):297–309
CrossRef

Xiao L, Dai H, Ning P (2012c) MAC design of uncoordinated FH-based collaborative broadcast. IEEE Wirel Commun Lett 1(3):261–264
CrossRef

Xiao L, Yan Q, Lou W, Chen G, Hou YT (2013) Proximity-based security techniques for mobile users in wireless networks. IEEE Trans Inf Forensics Secur 8(12):2089–2100
CrossRef

Xiao L, Chen T, Liu J, Dai H (2015) Anti-jamming transmission Stackelberg game with observation errors. IEEE Commun Lett 19(6):949–952
CrossRef

Xiao L, Li Y, Dai C, Dai H, Poor HV (2017) Reinforcement learning-based NOMA power allocation in the presence of smart jamming. IEEE Trans Veh Technol 67(4):3377–3389

CrossRef

Xiao L, Jiang D, Wan X, Su W, Tang Y (2018a) Anti-jamming underwater transmission with mobility and learning. IEEE Commun Lett 22:542–545
CrossRef

Xiao L, Lu X, Xu D, Tang Y, Wang L, Zhuang W (2018b) UAV relay in VANETs against smart jamming with reinforcement learning. IEEE Trans Veh Technol 67:4087–4097
CrossRef

Xiao L, Wan X, Dai C, Du X, Chen X, Guizani M (2018c) Security in mobile edge caching with reinforcement learning. IEEE Wirel Commun Mag 25(3):116–122
CrossRef

Xu W, Trappe W, Zhang Y, Wood T (2005) The feasibility of launching and detecting jamming attacks in wireless networks. In: Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing. ACM, New York, NY, pp 46–57