Reinforcement Learning for Scalable and Reliable Power Allocation in SDN-based Backscatter Heterogeneous Network

Furqan Jameel*, Wali Ullah Khan†, Muhammad Ali Jamshed‡, Haris Pervaiz§, Qammer Abbasi¶, Riku Jäntti*
* Department of Communications and Networking, Aalto University, FI-02150 Espoo, Finland.
† School of Information Science and Engineering, Shandong University, Qingdao, People's Republic of China.
‡ Institute of Communication Systems (ICS), Home of 5G Innovation Centre (5GIC), University of Surrey, UK.
§ School of Computing and Communications, Lancaster University, UK.
¶ School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK.

***Abstract*—Backscatter heterogeneous networks are expected to usher a new era of massive connectivity of low-powered devices. With the integration of software-defined networking (SDN), such networks hold the promise to be a key enabling technology for massive Internet-of-things (IoT) due to myriad applications in industrial automation, healthcare, and logistics management. However, there are many aspects of SDN-based backscatter heterogeneous networks that need further development before practical realization. One of the challenging aspects is the high level of interference due to the reuse of spectral resources for backscatter communications. To partly address this issue, this article provides a reinforcement learning-based solution for effective interference management when backscatter tags coexist with other legacy devices in a heterogeneous network. Specifically, using reinforcement learning, the agents are trained to minimize the interference for macro-cell (legacy users) and small-cell (backscatter tags). Novel reward functions for both macro- and small-cells have been designed that help in controlling the transmission power levels of users. The results show that the proposed framework not only improves the performance of macro-cell users but also fulfills the quality of service requirements of backscatter tags by optimizing the long-term rewards.**

***Index Terms*—Backscatter communications, Internet-of-things (IoT), Interference management, Reinforcement learning.**

I. Introduction

Backscatter communication is evolving as the cutting-edge technology to enable ultra-low-power and cost-efficient communications for future Internet-of-things (IoT) [1]. This is achieved by using ambient RF signals without the need for active RF transmission. Specifically, by modulating and reflecting the received RF signal, a backscatter tag transmits the data to nearby devices rather than generating RF signals using oscillators [2]. Due to such ease of use and efficiency, this emerging technology has been proposed to use in healthcare networks and for industrial automation. This communication method helps address the communication and energy efficiency-related issues of small devices with miniature power sources for remote tracking, and logistics management. Despite compelling advantages and applications of this technology, certain limitations restrict the adoption of worldwide backscatter communication systems [3].

In this regard, one of the promising solutions is to use backscatter communications with the existing legacy devices. This not only improves some of the limitations of the backscatter tags but also opens new avenues for improved spectral efficiency [4]. Due to recent advances in controllable switches, software-defined networking (SDN) has emerged as a flexible technology for configuring new network devices and deploying the different applications in cloud [5]. One of the main features of SDN is the decoupling of data and control planes [6]. In this manner, the control functions can be centralized which is viable for centrally controlled backscatter heterogeneous networks. An SDN-based backscatter heterogeneous network is considered one of the key enablers for the coexistence of legacy devices and backscatter tags. The low-power backscatter devices communicate with one another by taking advantage of the available signals in the surrounding environment [7]. The receiver receives the signals transmitted from the carrier emitter or an ambient RF source and decodes these signals as the useful information transmitted from the transmitting antenna. By segregating the backscatter receiver and the carrier emitter, the number of RF elements is reduced at backscatter devices, and they can operate more efficiently [8].

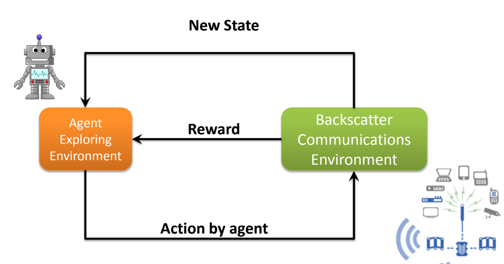

Fig. 1. A typical Q-learning model consisting of an agent that interacts with environment.

A. Related Work

In recent years, the performance of backscatter communications has been extensively studied for different communication conditions. In this regard, the authors of [9] presented an optimization framework to improve the performance of backscatter networks. More specifically, their considered network setup consists of a single source and destination along with several wireless-powered backscatter tags. To maximize the throughput, they solved the problem using the interior point method. Similarly, the authors of [10] considered a scenario with multiple backscatter tags and a single reader. Their considered tags used monostatic formation and reflected the RF carrier to the RF source. They optimized the successive decoding rate of the system and showed that their scheme improves the network performance. The authors of [1] explored the rate-energy tradeoff of the backscatter communication system. Their system model was composed of wireless-powered backscatter tag communicating under the Rayleigh fading. The authors of [11] provided measurement results for backscatter communication in healthcare networks. They used monostatic backscatter tags communicating at low frequencies for both indoor communication conditions.

The reinforcement learning-based artificial intelligence algorithms have been applied to wide areas of wireless communication such as D2D communication [12]. Q-learning is among the most explored and successful reinforcement learning techniques [13]. In this case, the problem consists of an environment and either single or multiple agents. As shown in Fig. 1, by observing the current state of the system, an agent takes action according to a stochastic policy. The authors in [14] used a reinforcement learning-based power control algorithm in underlay D2D communication and compared a centralized Q-learning based algorithm with distributed Q-learning. It was shown that distributed Q-learning users are enabled to self-organize by learning independently, thus, reducing the overall complexity of the system. In [15], the problem of vehicle-to-vehicle (V2V) transmission of the message was considered. However, there exist a few studies on backscatter communication that employ machine learning techniques [16], [17]. For instance, the authors of [18] used a supervised machine learning technique (support vector machine) to detect the signal from a backscatter tag by transforming the tag detection into a classification task. In [19] the authors discussed the ambient backscatter communication that enables wireless devices to communicate without utilizing radio resources. The system is modeled by the Markov decision process and the optimal channel is obtained by the iterative algorithm.

B. Motivation and Contribution

The aforementioned research efforts have significantly advanced the state-of-the-art on backscatter communications. However, the optimization aspect of backscatter communications has received little attention. The feasible adoption of backscatter communication largely depends on the optimization of existing solutions. Motivated by this objective, we aim to provide a novel Q-learning solution for SDN-based backscatter heterogeneous networks. The control plane in SDN-based backscatter heterogeneous networks provide the much-need flexibility for deploying Q-learning solutions. In this way, the performance and behavior of the network can be monitored and relevant information can be easily accessed by different applications. According to the best of authors' knowledge, the interference mitigation via Q-learning has not been performed for SDN-based backscatter heterogeneous networks. The main contribution of our work is twofold as detailed below:

1) We develop an SDN-based backscatter heterogeneous network to improve the spectral efficiency of the legacy networks. Specifically, the monostatic backscatter tags are associated with the small-cells that share the resources with macro-cell. The quality of service requirements and interference constraints of both the monostatic backscatter and legacy users have been taken into account.
2) A Q-learning optimization framework has been proposed to mitigate inter-cell interference. The Q-learning model considers different rewards for both macro-cell and small-cell. The results show that the proposed optimization framework improves the performance of the SDN-based backscatter heterogeneous network.

The remainder of the paper is organized as follows. Section II provides the details of the considered system model. In Section III, the proposed Q-learning framework for interference minimization is provided. Section IV presents the simulation results and provides a relevant discussion. Section V, finally, presents some concluding remarks and future research directions.

II. System Model

We consider an SDN-based uplink backscatter heterogeneous network having single macro-cell BS and multiple small-cell BSs operating at sub-6 GHz, as illustrated in Fig. 2. The control plane manages the communication, computation, and storage control functions and runs different machine learning applications. The control functions are centralized in the SDN controller and all the BSs are connected to the controller via optical fiber links, Based on the quality of service requirements of different users [20], the SDN controller can control the power levels of macro-cell BS and small-cell BS. The macro-cell BS and small-cell BS are assumed to be operating on the same channels using the same number of resource blocks. Without loss of generality, we consider that a user (i.e., monostatic backscatter or legacy user) can connect to only one BS at a time, whereas, the users are considered to be already associated with either the macro-cell BS or small-cell BS. The monostatic backscatter users are assumed to be communicating to small-cell BS while the legacy users are considered to be communicating with the macro-cell BS.

For backscatter communications with small-cell BS, we consider a monostatic backscatter configuration and that all the backscatter tags[1] are equipped with single antennas. Each monostatic backscatter user is assumed to use a reflection amplifier that is characterized by the negative load impedance [3]. The channel coefficients between the small-cell BS and

[1]The phrases 'backscatter tag' and 'backscatter user' are used interchangeably throughout this paper.

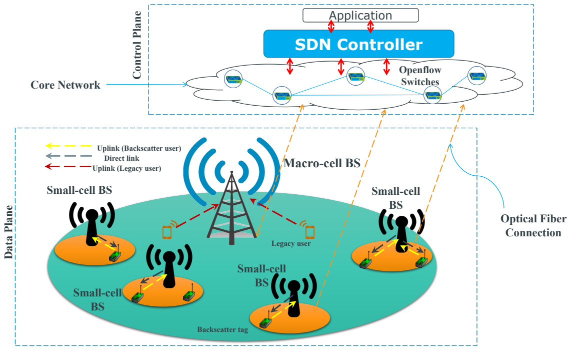

Fig. 2. Illustration of SDN-based backscatter heterogenous network consisting of macro-cell and multiple small-cells.

backscatter transmitter (i.e., direct link), and between the backscatter transmitter and small-cell BS (i.e., backscatter link) are denoted as g_{st} and g_{tr}. During each time slot, the backscatter transmitter has to decide whether to operate in energy harvesting mode or backscatter mode. All the monostatic backscatter tags are considered to be equipped with a rechargeable battery such that the backscatter transmitter can improve its life cycle by harvesting the wireless-power from the RF carrier of small-cell BS.

Specifically, the backscatter transmitter uses the RF signal to harvest energy by converting the RF signal into a direct current. The collected energy can be used for charging the battery or transferring the data back to the small-cell BS. Thus, the harvested energy can be denoted as $E_h^i = \eta\rho^E|g_{st,i}^{k,f}|^2 P_i^f$, where η denotes the energy harvesting efficiency, ρ^E is the power-splitting ratio, and P_i^f is the transmit power of the small-cell BS. Generally, the transmission signal is known at the small-cell BS and, thus the small-cell BS can apply interference cancellation techniques to obtain the signal from monostatic backscatter tag [21].

The instantaneous SINR at the k-th small-cell BS can be written as

$$\Omega_i^{k,f} = \frac{\mu\rho^I P_i^f |g_{st,i}^{k,f}|^2 |g_{tr,i}^{k,f}|^2}{I_1 + I_2 + N_0} \tag{1}$$

where $0 < \mu < 1$ is the reflection coefficient, ρ^I is the information processing power-splitting ratio, $g_{st,i}^{k,f}$ is the channel gain between the i-th backscatter user and the k-th small-cell BS, N_0 represents the noise variance of additive white Gaussian noise (AWGN), P_i^f the transmission power of small-cell BS, I_1 and I_2 are the interferences, respectively, given as

$$I_1 = \sum_{j \in I_m} P_j^m |g_{tr,j}^m|^2, \tag{2}$$

and

$$I_2 = \sum_{r \in I_f} \mu\rho^I P_r^f |g_{st,r}^f|^2 |g_{tr,r}^f|^2. \tag{3}$$

In (2), I_m is the set of interfering macro-cell users on the same sub-channels. Similarly, I_f in (3) represents the interference from monostatic backscatter users in the other small-cell BSs. Now, the received SINR at the macro-cell BS for decoding the l-th legacy user's message can be given as

$$\Omega_l^m = \frac{P_l^m |g_{tr,l}^m|^2}{I_1 + I_2 + N_0}, \tag{4}$$

where P_l^m denotes the transmission power of the legacy user to the macro-cell BS, and $g_{tr,l}^m$ is the channel gain between l-th legacy user and macro-cell BS.

III. Scalable, and Reliable Power Allocation via Reinforcement Learning

Before solving the interference problem in SDN-based backscatter heterogeneous networks, it is important to understand the dynamics of different entities in the network. We consider that a dynamic resource allocation application running in the control plane performs power allocation via reinforcement learning. In this case, we resort to using Q-learning which allows us to find the optimal policy over the long-time interaction of agents. In general, Q-learning has three main components, i.e., state, reward, and action. In that, the agent is rewarded based on the action it takes for pre-defined states. Specifically, the agent uses the Q-learning table to maximize the reward by interacting with the environment. In our considered network setup, the agents are the members of the macro and small-cells that are competing for the constrained resources. As a result of this interaction, a significant amount of interference is introduced in the network known as co-channel interferences. We anticipate this interference can be mitigated after finding the optimal policy for each agent that learns independently about the environment and do not cooperate.

To avoid the unjust distribution of resources and make sure that members of the macro-cell do not fall in the low SINR region, we assume that the small-cell BS knows the channel of the macro-cell. This assumption is incorporated in the reward function for developing the Q-learning model. We now provide details of the state-reward function for the proposed Q-learning model. Here, for the sake of simplicity, we consider that we have one macro-cell BS and two small-cell BSs. This does not undermine the significance of the proposed solution since this optimization framework can be easily extended for a larger number of small-cells in the network. Moreover, due to the different dynamics of macro-cell and small-cell, we define the reward function separately for both cells.

A. State

The state of any small-cell $s^{SM} = \{I_{\Omega_m}, I_{\Omega_i}, I_r\}$ is represented, respectively, as a tuple of following indicators:

$$I_{\Omega_m} = \begin{cases} 1 & \Omega_m \geq \Omega_T \\ 0 & \Omega_m < \Omega_T \end{cases} \tag{5}$$

$$I_{\Omega_i} = \begin{cases} 1 & \Omega_i \geq \Omega_T \\ 0 & \Omega_i < \Omega_T \end{cases} \tag{6}$$

TABLE I
STATES OF MACRO-CELL AND SMALL-CELL.

S No.	Macro-cell States	Small-cell States
1.	$s_0(0,0)$	$s_0(0,0,0)$
2.	$s_1(1,0)$	$s_1(0,1,0)$
3.	$s_2(0,1)$	$s_2(0,0,1)$
4.	$s_3(1,1)$	$s_3(0,1,1)$
5.	$s_4(0,2)$	$s_4(0,0,2)$
6.	$s_5(1,2)$	$s_5(0,1,2)$
7.		$s_6(1,0,0)$
8.		$s_7(1,1,0)$
9.		$s_8(1,0,1)$
10.		$s_9(1,1,1)$
11.		$s_{10}(1,0,2)$
12.		$s_{11}(1,1,2)$

and

$$I_r = \begin{cases} 0 & \frac{\Omega_i}{P_i} \leq T_a \\ 1 & T_a < \frac{\Omega_i}{P_i} < T_z \\ 2 & \frac{\Omega_i}{P_i} \geq T_z \end{cases} \tag{7}$$

where Ω_T is the level of required SINR for reliable communication, Ω_m represents the instantaneous SINR of the macro-cell, whereas, Ω_i denotes the instantaneous SINR of the small-cell. Furthermore, T_a and T_z are the thresholds for energy efficiency ratio. More specifically, if energy efficiency is below T_a, then the small-cell is in experiencing low efficiency. On the other hand, if the energy efficiency is above T_z, then it is in the desirable region of high efficiency. The small-cell not only considers the SINR of itself but also the macro-cell's SINR. This results in small-cell to achieve a high efficiency without compromising the quality of service requirements of the macro-cell. On the other hand, the macro-cell only cares about itself and tries to achieve a high efficiency. Due to this reason, the states of macro-cell is a tuple of two indicators $s^{MA} = \{I_{\Omega_m}, I_{\Omega_i}\}$. Thus, for the considered setup, there are 12 possible states for both macro-cell and small-cell members as given in Table I.

B. Action

The actions of the agents are predefined transmission power levels. Specifically, the power levels are increased from a specific level with a constant step size.

C. Reward

Again, the reward for both macro-cell and small-cell members differ since macro-cell members must be kept above a certain quality of service and capacity limit. However, we also intend to maximize the capacity of the small-cell members. We define the reward functions of macro- and small-cells based on their corresponding SINRs. More specifically, the reward functions of macro-cell and small-cell members can, respectively, be given as

$$r_i^m = \begin{cases} 100 & \Omega_m \geq \Omega_T \\ -1 & \text{otherwise} \end{cases} \tag{8}$$

and

$$r_i^f = \begin{cases} 100 & \frac{\Omega_m}{\Omega_T} \geq 1, \frac{\Omega_i}{\Omega_T} \geq 1 \\ -1 & \frac{\Omega_m}{\Omega_T} \geq 1, \frac{\Omega_i}{\Omega_T} < 1 \\ -1 & \frac{\Omega_m}{\Omega_T} < 1, \frac{\Omega_i}{\Omega_T} < 1 \end{cases} \tag{9}$$

The macro-cell is rewarded a large value if the SINR is higher than the threshold, while it receives a small punishment if it is below. Similarly, the small-cell members receive the reward and punishment for their corresponding SINR values. In this regard, it is worth highlighting that a negative reward may result in low Q-value. By iteratively updating the Q-values, the gap between the rewards of macro-cell and small-cell members can be mitigated. Ultimately, the agent selects the action with the highest Q-value for every state.

Fig. 3 illustrates the working of the proposed Q-learning technique. Let us consider the small-cell agent is at a random state $s_0(0,0,0)$ and selects a random action for transmission power level 10 dB. The small-cell agent can now estimate the SINR since the other agents also choose their actions randomly. This information can be shared among agents through the BSs. Now, the instantaneous SINR can be easily estimated. Let us consider that based on the predefined values, the agent now moves to the state $s_{11}(1,1,2)$ and receives a reward. Again, the agent selects a random action and supposedly moves to the next state $s_7(1,1,0)$. Here, the agent again selects an action and learns that the calculated SINR has been reduced and that the SINR of the macro-cell is also below the threshold. These sorts of dynamics are modeled by our proposed Q-learning technique for interference management in SDN-based backscatter heterogeneous networks.

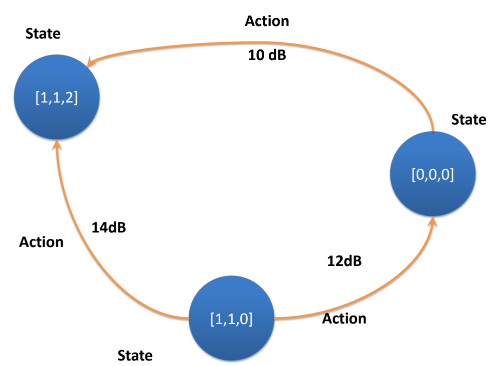

Fig. 3. Working example of the Markov Decision Process.

IV. PERFORMANCE EVALUATION

In this section, we provide the simulation results and their relavent disucssion. To validate the performance of the proposed Q-learning optimization framework, we perform Monte-Carlo simulations. In that, the channel gains between BS and the backscatter tag/legacy user depends on their distance $d^{-\chi}$, where, $\chi = 4$ is the pathloss exponent. Unless stated otherwise, the simulation parameters and their corresponding values are given as follows: learning rate = 0.5, $N_0 = 0.1$,

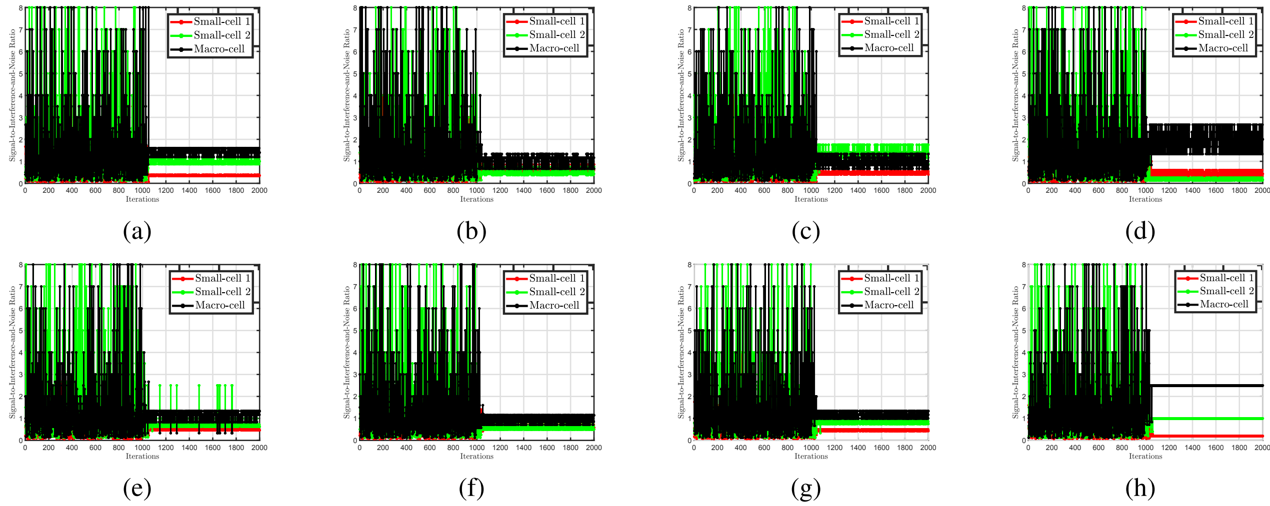

Fig. 4. Illustration of SINR for small-cells and macro-cell, where, (a) $T_z = 3$, (b) $T_z = 4$, (c) $T_z = 5$, (d) $T_z = 6$, (e) $T_z = 7$, (f) $T_z = 8$, (g) $T_z = 9$, (h) $T_z = 10$.

$\eta = 0.9$, $\rho^I = 0.8$, discount factor=0.9, transmit power range = 2 - 16 dB, T_z=10, $T_a = 2$, $\Omega_T = 5$.

Fig 4 the values of SINR for the increasing number of iterations. In general, it can be noted that the model converges around 1100 iterations for small-cells and macro-cell. Due to the high priority of the macro-cell (legacy) users, it can be seen that the SINR value of macro-cell users is higher than the backscatter (small-cell) users. Moreover, the SINR value of small-cell backscatter users generally fluctuates. As the value of T_z increases, given Fig 4 (a) to Fig 4 (h), the value of SINR of macro-cell users gradually improves. Specifically, the gap between small-cell and macro-cell SINR increases and becomes stable. Finally, when $T_z = 10$, the gap between the SINRs of macro-cell and small-cell users is generally greater.

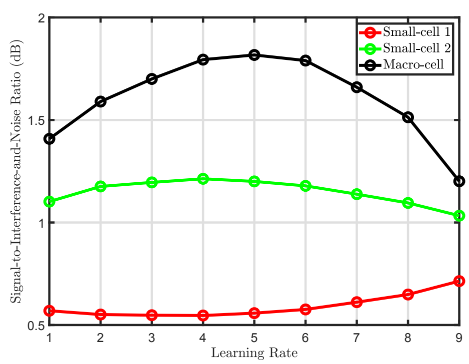

Fig. 5. SINR against learning rate.

Fig 5 shows the impact of different learning rates on the performance of the proposed Q-learning based optimization framework. The learning rate has an impact on the performance of the Q-learning agent since it decides the tradeoff between exploration and exploitation of the agent. It can be seen from the figure that for macro-cell and small-cell 2, the increase in learning rate first improves that SINR and then results in decreasing the value of the SINR. However, the same is not true for small-cell 1. For the case of small-cell 1, the increase in learning rate from 0.1 to 0.9 first reduces the SINR and then increases the SINR as the learning rate approaches 0.9. Moreover, the peak and lowest value of SINR differ for both macro-cell, small-cell 1 and small-cell 2. This shows that based on the priority of macro-cell and small-cell users, the learning rate could be different to improve the learning capabilities of the agent.

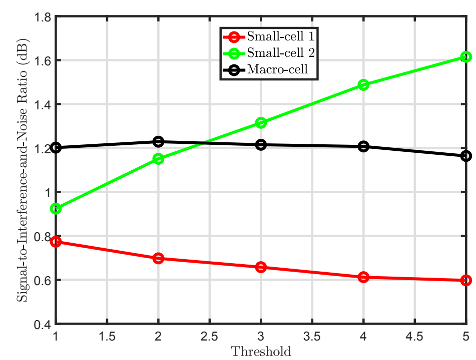

Fig. 6. Illustration of SINR for increasing values of threshold.

Fig 6 shows the change in values of SINR for increasing threshold values (Ω_T). It can be seen that the value of SINR for both macro-cell and small-cell users generally decreases with an increase in the threshold value. However, due to the reduced information gap among the small-cell users, the value of small-cell 2 increases, whereas, the SINR value of small-cell 1 decreases. Moreover, it can also be observed that the gap between SINR of small-cell 1 and small-cell 2 increases with an increase in the value of the threshold. Thus, one can deduce that the improvement in SINR of small-cell 2 users comes at the cost of small-cell 1 users.

In Fig 7, we illustrate the impact of T_a on the SINR of both

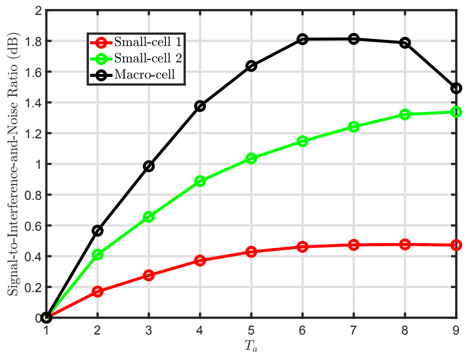

Fig. 7. SINR versus different values of T_a.

macro-cell and small-cell users. In general, it can be seen that an increase in the value of T_a improves the SINR of both the macro-cell and small-cell users. However, as the value of T_a increases further, the SINR values approaches a ceiling. This shows that despite a large increase in the value of T_a no significant increase in the value of SINRs of macro-cell and small-cell users is observed.

V. Conclusion and Future Work

SDN-based backscatter heterogeneous networks are going to play a critical role in enabling low-powered massive IoT applications. However, some challenges need to be addressed before the practical realization of such networks. In this regard, this work has provided a Q-learning based optimization framework for mitigating the impact of interference in SDN-based backscatter heterogeneous networks. To do so, we have proposed a novel reward function for both macro-cell (legacy) and small-cell (backscatter) users. The proposed Q-learning model is formulated in a way where macro-cell users are given priority over small-cell users. The simulation results indicate that the Q-learning framework increases the performance of the macro-cell users while maintaining a recommended level of SINR for the backscatter users in the small-cell.

Although the result provided here show considerable promise, they can be extended in many ways. For instance, future studies can consider multi-antenna backscatter tags to improve the overall throughput of the system. Our proposed framework can be used to improve the performance of the SDN-based backscatter heterogeneous network through an efficient beamforming mechanism. Furthermore, an effective scheduling mechanism can further reduce the impact of interference in SDN-based backscatter heterogeneous networks. Our proposed framework can be combined with the scheduling technique which may improve the performance manifolds. These challenging yet interesting works are left for future studies.

Acknowledgment

The work of F. Jameel and R. Jäntti was supported in part by Business Finland under the project 5G Finnish Open Research Collaboration Ecosystem (5G-FORCE) and the EPSRC Global Challenges Research Fund (GCRF) DARE project grant no. EP/P028764/1.

References

[1] F. Jameel, T. Ristaniemi, I. Khan, and B. M. Lee, “Simultaneous harvest-and-transmit ambient backscatter communications under Rayleigh fading,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 166, 2019.

[2] R. Long, H. Guo, L. Zhang, and Y.-C. Liang, “Full-duplex backscatter communications in symbiotic radio systems,” *IEEE Access*, vol. 7, pp. 21 597–21 608, 2019.

[3] F. Amato, C. W. Peterson, B. P. Degnan, and G. D. Durgin, “Tunneling RFID tags for long-range and low-power microwave applications,” *IEEE J. Radio Freq. Identif.*, vol. 2, no. 2, pp. 93–103, 2018.

[4] G. Yang, X. Xu, and Y. Liang, “Resource allocation in NOMA-enhanced backscatter communication networks for wireless powered IoT,” *IEEE Wireless Commun. Lett.*, pp. 1–1, 2019.

[5] K. Zheng, L. Hou, H. Meng, Q. Zheng, N. Lu, and L. Lei, “Soft-defined heterogeneous vehicular network: Architecture and challenges,” *IEEE Network*, vol. 30, no. 4, pp. 72–80, 2016.

[6] S. Bera, S. Misra, and A. V. Vasilakos, “Software-defined networking for internet of things: A survey,” *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1994–2008, 2017.

[7] D. Li and Y.-C. Liang, “Adaptive ambient backscatter communication systems with MRC,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12 352–12 357, 2018.

[8] Q. Zhang, L. Zhang, Y. Liang, and P. Y. Kam, “Backscatter-NOMA: An integrated system of cellular and internet-of-things networks,” in *IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[9] B. Lyu, C. You, Z. Yang, and G. Gui, “The optimal control policy for RF-powered backscatter communication networks,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2804–2808, 2017.

[10] J. Guo, X. Zhou, S. Durrani, and H. Yanikomeroglu, “Backscatter communications with NOMA (Invited Paper),” in *Int. Symp. Wireless Commun. Sys. (ISWCS)*, Aug. 2018, pp. 1–5.

[11] F. Jameel, R. Duan, Z. Chang, A. Liljemark, T. Ristaniemi, and R. Jantti, “Applications of Backscatter Communications for Healthcare Networks,” *arXiv preprint arXiv:1906.09209*, 2019.

[12] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, “Machine learning paradigms for next-generation wireless networks,” *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, 2016.

[13] A. Syed, K. Yau, H. Mohamad, N. Ramli, and W. Hashim, “Channel selection in multi-hop cognitive radio network using reinforcement learning: An experimental study,” 2014.

[14] S. Nie, Z. Fan, M. Zhao, X. Gu, and L. Zhang, “Q-learning based power control algorithm for D2D communication,” in *PIMRC*. IEEE, 2016, pp. 1–6.

[15] S. Sharma and B. Singh, “Cooperative Reinforcement Learning Based Adaptive Resource Allocation in V2V Communication,” in *SPIN*. IEEE, 2019, pp. 489–494.

[16] T. T. Anh, N. C. Luong, D. Niyato, Y.-C. Liang, and D. I. Kim, “Deep Reinforcement Learning for Time Scheduling in RF-Powered Backscatter Cognitive Radio Networks,” *arXiv preprint arXiv:1810.04520*, 2018.

[17] A. Rahmati and H. Dai, “Reinforcement Learning for Interference Avoidance Game in RF-Powered Backscatter Communications,” *arXiv preprint arXiv:1903.03600*, 2019.

[18] Y. Hu, P. Wang, Z. Lin, M. Ding, and Y.-C. Liang, “Machine Learning Based Signal Detection for Ambient Backscatter Communications,” in *ICC*. IEEE, 2019, pp. 1–6.

[19] X. Wen, S. Bi, X. Lin, L. Yuan, and J. Wang, “Throughput Maximization for Ambient Backscatter Communication: A Reinforcement Learning Approach,” in *ITNEC*. IEEE, 2019, pp. 997–1003.

[20] T. Han, Y. Han, X. Ge, Q. Li, J. Zhang, Z. Bai, and L. Wang, “Small cell offloading through cooperative communication in software-defined heterogeneous networks,” *IEEE sensors journal*, vol. 16, no. 20, pp. 7381–7392, 2016.

[21] R. Duan, E. Menta, H. Yiğitler, and R. Jäntti, “Hybrid Beamformer Design for High Dynamic Range Ambient Backscatter Receivers,” *arXiv preprint arXiv:1901.05323*, 2019.