Mobile Networks and Applications
https://doi.org/10.1007/s11036-023-02149-z

A new framework for Physical Layer Security in HetNets based on Radio Resource Allocation and Reinforcement Learning

Dania Marabissi[1] · **Andrea Abrardo**[2] · **Lorenzo Mucchi**[1]

Accepted: 3 December 2021

Abstract
Densification of networks through heterogeneous cells deployment is considered a key technology to satisfy the huge traffic growth in future wireless systems. In addition to achieving the required communication capacity and efficiency, another significant challenge arises from the broadcast nature of wireless channels: vulnerability to wiretapping. Physical-layer security is envisaged as an additional level of security to provide confidentiality of radio communications. Typical characteristics of the wireless channel (noise, interference) can be exploited to keep a message confidential from potential eavesdroppers. In particular, heterogeneous networks (HetNet) have inherent security features: while the legitimate user can benefit of the HetNet architecture, the eavesdropper is strongly affected by the inter-cell interference. This paper presents an overview of HetNets intrinsic security benefits, mainly focusing on users association and resource allocation policies. In particular, allocation of radio resources is a poorly investigated topic when related to information security. However, in systems with a large radio resource reuse like HetNets, co-channel interference can be suitably exploited to resist to the eavesdropper. This paper presents a new framework for radio resources allocation using reinforcement learning (Q-learning) to increase the security level in HetNets. A coordinated scheduling among different cells using the same radio resources is proposed based on the exploitation of the spatial information. The goal is to optimize the security at physical layer. The reinforcement learning approach represents a feasible and efficient solution to the proposed problem.

Keywords Physical-layer security · Resource allocation · Heterogeneous networks · Reinforcement learning

1 Introduction

Fifth generation (5G) wireless networks will accommodate network densification and different radio access technologies to satisfy the dramatic traffic growth and the huge connectivity demand triggered by the massive diffusion of Internet of Things (IoT), intelligent devices and multimedia contents. Heterogeneous networks (HetNets), consisting of a conventional macrocell coupled with dense low-power small cells, are considered a key driver to enable ultra-high data rates, high-reliability and ultra-low latency. Beside spectral and energy efficiency, also physical layer security can benefit from such heterogeneous dense architecture [1].

Due to the broadcast nature of wireless communications, confidential data exchanged between transmitter and receiver is vulnerable to eavesdropping, hence, one of driving aims of the 5G design is to have secure wireless connections. Toward this goal, *physical layer security* is a low-complexity approach that exploits the randomness of the wireless channel for providing secure transmissions. An unauthorized

This work was supported in part by the European Unions Horizon 2020 Research and Innovation Programme under Grant 872752. This work was also partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART")..

✉ Lorenzo Mucchi
lorenzo.mucchi@unifi.it

Dania Marabissi
dania.marabissi@unifi.it

Andrea Abrardo
andrea.abrardo@unisi.it

[1] Dept. of Information Engineering, University of Florence, via di S. Marta 3, 50139 Firenze, Italy

[2] Dept. of Information Engineering, University of Siena, via Roma 56, 53100 Siena, Italy

Published online: 03 August 2023

eavesdropper cannot acquire confidential information if its communication channel is a degraded version of the legitimate channel. In other words enhancing the legitimate channel quality and/or exacerbating the wiretap channel it is possible to have secure communications. In particular, the *secrecy capacity* is the maximum transmission rate of the legitimate User Equipment (UE) at which the eavesdropper cannot decode the message. Thanks to their random and changing environment HetNets have inherent security features. While the legitimate UE can benefit of the HetNet architecture and its close proximity to the serving base station (BS), by exploiting suitable beamforming and interference management schemes, the eavesdropper is strongly affected by the inter-cell interference. Consequently, physical layer security in HetNets depends on UE association policies as well as on resource allocation policies. Different types of resources have a significant impact on the quality of the legitimate and wiretap links. In most of the physical-layer security studies power, space and jamming/relay dimensions are exploited, while radio resources (RR) dimension is less investigated. This is mainly because previous studies on physical-layer security mainly focus on scenarios where the RR reuse is limited, while in HetNets there is a dense RR reuse that results in significant intra and inter-tier interference. Indeed, radio resource allocation (RRA) has been widely investigated in HetNets aiming at increasing spectrum/energy efficiency, fairness, throughput, but information security is rarely considered. However, in HetNets the co-channel interference arising from RR reuse can be exploited to degrade the wiretap channel thus becoming a positive effect. This brings new opportunities but also new challenges. Differently from the traditional RRA approaches where the goal is to eliminate interference as much as possible, in this context, it is needed to find a trade-off between the protection of the legitimate communications from interference and jamming the eavesdropper. This paper focuses on a RRA scheme for providing security in a HetNet. As better analysed in Sec.3.2.1, in the literature, RRA problem for providing security is investigated in a few papers focusing on underlying device-to-device (D2D) communications, exploiting the interference generated by the overlapping of primary and secondary transmissions. However, the complexity of a HetNet is not captured and analysed. The few papers that focus on HetNets are based on the knowledge of the eavesdropper position, while this hypothesis is completely removed here. Moreover, we propose a coordinated scheduling/coordinated beamforming approach that has not been presented before in this context.

RRA problems are generally formulated as non-convex optimization problems, whose solution requires the use of optimization tools with unmanageable computational complexity. Sub-optimal solutions based on techniques such as Lagrangian relaxations, iterative distributed optimization, heuristic algorithms, and auction/game have been intensely researched in the last two decades, since the seminal work [2] published in 1999. Thereafter, despite the field of optimization for wireless communications is greatly evolved in all these years, the proposed solutions still suffer from high computational complexity. Compared to the optimization tools, machine learning (ML)-based RRA algorithms can be implemented online with reduced complexity and can optimize resource allocations in complicated networks [3]. In particular, reinforcement learning (RL) has been shown to be one effective solution for RRA in communication and computing systems that has attracted the attention of many researches in the last years, e.g., see among the others [4, 5]. In the RL framework, an RL agent can generate (near-)optimal control actions basing on the immediate reward feedback from interactions with the environment. Together with simply optimizing the current reward in a greedy manner, the RL agent can take a long-term goal into account, which is essentially important to time-variant dynamic systems.

This paper first introduces the HetNet intrinsic security features and a new security metric for providing a security measure independent on the eavesdropper position. Then a focus on the RRA is provided. A new framework based on a UE association and a RRA scheme specifically tailored to increase the physical security level in HetNet is presented. A possible solution is proposed using a machine learning (ML)-based RRA algorithm that can be implemented online with affordable complexity. More specifically, we consider a RL-based scheme, where each cell can autonomously evaluate its own rewards and deliver them to a central controller that is in charge of establishing the best resource allocation policy. To the best of our knowledge, this is the first case where a RL solution is proposed for RRA problems in the context of physical layer security. Finally, we want to underline that the proposed approach is suitable for an actual context where users need security in their communications. Indeed it does not require the knowledge of the position of the eavesdropper that is usually difficult to obtain, moreover, the proposed framework needs only a limited amount of information that is also usually available at the BS as detailed later.

The paper is organized as follow. First the physical-layer security, and in particular an area-related security metric is introduced. Then the inherent security features of HetNets are described mainly focusing on UE association and radio resource allocation. Finally, the proposed framework based on secure UE association and RL-based RRA is presented.

2 Physical Layer Security

The amount of sensitive and confidential data transmitted via wireless channels is continuously increasing, thus making privacy and secrecy vital issues for future wireless com-

munications. Traditional security is achieved by means of cryptography techniques that have a certain level of computational complexity, and may result in high latency and communication failure in fast changing networks, due to higher-layer key distribution and management especially in dense networks.

Physical layer security is a low-complexity approach that exploits the randomness of wireless channel along with efficient resource management schemes to keep messages confidential in the presence of malicious eavesdroppers. The basic idea is of exacerbating the wiretap channel and/or enhancing the legitimate link, so that the first one is a degraded version of the second.

To evaluate the effectiveness of security schemes, metrics commonly used in literature are mainly *secrecy capacity* and the *secrecy outage probability*. The former is the discrepancy of information quantity between the legitimate channel and wiretap channel, while the latter is the probability that the secrecy capacity goes below a previously set target capacity C. These two metrics are very effective to measure the secrecy of a communication link, but they require that both the position and the channel state of the eavesdropper are known. This assumption is unpractical in a real context. For this reason, a different metric that measures the level of security of an area is introduced to drop down this assumption.

2.1 Secure Area

Let us call the transmitting BS as Alice and the receiving UE as Bob. Positions of Alice and Bob in the area are known, differently the position of the eavesdropper (Eve) is not known. Alice wants to transmit a confidential message M to Bob, and Bob tries to recover the message M from the received vector Z_B (Fig. 1). Analogously, Eve tries to recover as much information as possible on the message M from the observed vector Z_E.

As known, the secrecy capacity is at least as large as the difference between two channel capacities: the legitimate and the eavesdropper [6]. Thus, the secrecy capacity depends on channel state and position of Bob and Eve compared to Alice, respectively. In other words, to compute the secrecy capacity of a link, the position and the channel of Eve must be known. To drop this unpractical assumption a different metric can be considered.

Given the positions of Alice and Bob, the secrecy capacity of each point $(x; y)$ of the area A can be calculated as Eve is located in that generic point. Thus, a *secrecy map* [7] showing the different levels of security of the entire area A can be obtained. From the secrecy map, a metric, called *secure area*, can be derived. First, we make the assumption that Eve could be located hypothetically in each point of the area A; then, a binary matrix is calculated, where each point (x, y) of the overall area A is marked as "1" if in that point the secrecy capacity is strictly positive (i.e., the legitimate BS-UE link capacity is higher than BS-Eve link capacity supposing that Eve is in (x, y)), and "0" otherwise. The secure area is calculated by integrating over all the points of the matrix. In other words, the secure area is the percentage of points in the total area A which can provide positive secrecy rate to the legitimate link. The secure area metric tells how much secure is an area managed by a BS, given the position of the legitimate user.

3 Physical Layer Security in HetNet

The basic idea of HetNets is deploying nodes to create a multi-tier hierarchical architecture: high-power macro cells are overlaid by a large number of heterogeneous small cells characterized by different levels of transmit power, antenna configurations and bandwidth. These characteristics, together with the random deployment of cells, make the wireless channels more diverse and random than ever before, and

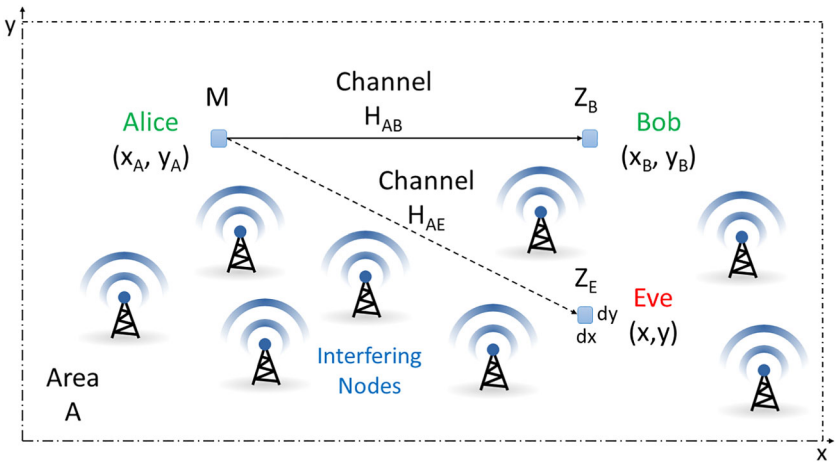

Fig. 1 Scheme of the transmission of the confidential message M from Alice to Bob

strongly affected by inter-cell interference. This propagation environment offers new opportunities for achieving security at physical layer. In fact, HetNets have an inherent high secrecy capacity [8, 9]. While for the legitimate user multiple antenna and interference management systems enable densification benefits, the wiretap channel is degraded by the very strong interference, thus resulting in an improved secrecy. In particular, network densification allows to reduce communication distances and, hence, the path-loss. However, only the authorized UE can achieve this benefit, while eavesdroppers suffer of more interference generated from surrounding cells. Moreover, usually in HetNets suitable interference management schemes are used to avoid that severe interference affects the intended UE communications, while the eavesdropper can only increase its receiving antenna gain. Finally, differently from the eavesdropper, UE benefits of transmit antenna gain, that also allows to reduce the transmitting power, thus reducing the capacity of the eavesdropper to intercept the communications.

3.1 User Association

UE is usually under the coverage of multiple cells in a HetNet, and the dense and random deployment of the network infrastructure requires of rethinking the user association strategies. The main problem is the different power level of cells belonging to different tiers, that leads to inefficiency of the conventional UE association policy based on maximum received signal to noise plus interference ratio (SINR) (*max-SINR association*). For this reason, several studies in the literature focus on UE association in HetNets proposing strategies suitably designed to optimize specific metrics that typically are load balancing, throughput, spectrum/energy efficiency. However, also physical layer security can be improved. Even if, most of the paper focusing on physical-layer security in HetNets assume a max-SINR UE association, in [10, 11] it is shown that selecting a "non-best" UE association criterion or the maximum secrecy capacity criterion, the secrecy can improve. In fact, UE is associated with the best cell, depending on specific association rules and optimization metrics, and makes handover when conditions change, conversely eavesdroppers have no choice but to undergo arbitrarily varying channel conditions when intercepting a specific UE.

In what follows we present a new UE association policy whose goal is maximizing the secure area that has been proposed in [12] and is used here only to distribute the users among the BSs before performing the proposed RRA algorithm. We want to underline that the proposed RRA algorithm could work also using a different association policy. UE-BS association is used only for the system setting, to detect which is the serving BS of each UE, but it does not impact on the functioning of RRA algorithm proposed below.

3.2 Resource Allocation

Resource allocation in HetNet has been widely investigated in the literature in order to improve energy/spectrum efficiency, throughput, fairness, and it can be exploited also to increase the physical layer security. In fact, there are different types of resources that have a significant impact on the quality of legitimate and wiretap links. Thus, the communication secrecy can be effectively achieved by suitably managing resources with the goal of enhancing physical layer security [13]. Different resource allocation strategies in different domains have been investigated

- *Power control* - the co-channel interference can be used as a friendly-jamming. The power must be suitably managed and distributed among different jammers, in order to jam the eavesdropper but without generating harmful interference on the intended UE.
- *Jammer/relay selection* - the network can rely on some nodes that act as relay or jammer to improve the communication secrecy. These nodes have a different impact on the eavesdropper and legitimate UE due to their location and channel conditions. The UE has to select proper relays or jamming nodes finding a trade-off between enhancing its own communication link and degrading the eavesdropper link. Moreover, the time-optimization between the two slots in a relaying system, can be beneficial due to asymmetry of the channels of transmitter-eavesdropper and relay-eavesdropper links.
- *Space* - generally BSs are equipped with multiple antenna systems, hence, suitably selecting the antenna pre-coding vectors it is possible to control the direction of transmitted and jamming signals so that the former is directed toward the legitimate UE while the latter falls in the null space of the legitimate channel. As a result, the wiretap channel is deteriorated while the legitimated channel is improved.

3.2.1 Radio Resources

Another resource dimension that indeed is less discussed and investigated compared to the previous ones in the context of the physical layer security, is the *Radio Resource* (i.e., time-frequency resources) dimension. This is the focus of our paper. RRA has attracted less interest for improving communication secrecy, because scenarios considered for the physical layer security usually have no/limited resource reuse. Most of the works in the literature are based on simple scenarios with a source-destination link and an eavesdropper whose position and channel state information is usually assumed to be known even if it is unpractical in many contexts. Conversely, the study in actual and more complex networks as HetNets is quite immature, and efforts are needed to understand potentialities and limitations.

In [14–16] a single cell scenario is considered, differently from our case. In particular, in [14] RRA for an Orthogonal Frequency Division Multiple Access system is considered, where subcarriers are suitably assigned to secure and non-secure users to maximize the aggregate information rate while maintaining a secrecy rate for the secure users, but interference arising form RRs reuse is not present. Differently [15, 16] consider underlying D2D communications, hence a certain level of RRs reuse is present: concurrent transmissions of primary and underlying users are used as a friendly jammers to increase the resistance to eavesdropping. Differently form our approach the focus is on D2D communications where each D2D pair selects the channel that maximizes the secrecy rate, while the complexity of interference coordination in a dense HetNet is not captured, as well as the spatial information is not considered. Moreover, these approaches assume to known the eavesdropper position, while this assumption is completely removed in our approach.

If HetNets are considered, a high-level of RR reuse is present for increasing spectral efficiency and managing the high number of links with limited radio resources. This results in intra-layer and inter-layer interference that can be beneficial to resist to eavesdroppers without the need of generating additional jamming signals. Consequently, RRA is a crucial point in HetNet as it has a strong impact on the interference management. It would be optimum to have enough interference to completely jam the eavesdropper link while avoiding a severe degradation of the legitimate link. This is different from the conventional approach where the goal is only to eliminate as much as possible the interference. A few papers in the literature focus on physical-layer security in HetNet and in general these assume to know the eavesdropper position. In [17] a location-based frequency allocation schemes in a two-layer HetNet is proposed to enhance only the macro-layer security exploiting the small-cell interference. A potential-game is used to suitably select the sub-band used by the small-cells. In order to relax the constraint on the eavesdropper position knowledge, in [18] the RRA problem in a HetNet is considered when the position of Eve is unknown to the considered BS but it can be localized by surrounding BSs. Then joint power and subcarrier allocation is performed taking into account security and fairness. The solution is obtained by using the Lagrangian method.

The assumption of a specific position of the eavesdropper means that previous approaches mainly focus on generating inter-layer interference in Eve direction. This assumption is removed in our paper, where the concept of secure area is considered, hence, there is not a specific direction to protect, but RRA must follow different criteria.

Hence, differently from previous approaches, we propose a RRA scheme aiming at increasing the communication security of the users in an area where different cells layers are present and the position of Eve is unknown. The idea is to exploit a coordinated scheduling/coordinated beamforming approach. Transmissions of users served by neighbour cells, can be accommodate on suitable RRs and spatial directions, so that the interference generated by the allocation of different users on the same resource block finds a good trade-off between reducing the interference for the intended user and increasing the resistance to the eavesdropping in the whole area.

4 Proposed Framework

We focus on the physical layer security of the downlink of a HetNet where several heterogeneous small-cells are densely deployed in the macrocell area, A. The signal transmitted by the BSs can be received by legitimate UEs but also by eavesdroppers whose position is not known. Consequently, we want to map the BS-UE link secrecy all over the area A.

A reuse factor equal to 1 is considered[1], hence all cells share the same frequency band and use all available physical resources to communicate with their associated UEs. Consequently, the signal received by the legitimate UE and the eavesdropper is affected by inter-cell interference in addition to fading and path-loss.

4.1 Max-SA Association

The UE association policy we assume to use here is based on secure area metric, and is named maximum secure area (*max-SA*) [12]. It is used as starting point for the following RRA scheme we propose here. A UE that has to connect to the network listens the broadcast channel of surrounding cells (i.e., those received with a SINR value over a given threshold). Then the measured SINR is reported to these cells. Each BS independently calculates the secure area for the UE, that is the area in which a potential eavesdropper cannot decode the signal. The secure area is then forwarded by each BS to the UE that selects as serving cell the one that provides the highest value. As an example Fig. 2 shows the area where a potential eavesdropper cannot decode the signal (i.e., secure area - green) and the insecure area (red) that a UE achieves selecting as serving cell BS1 or BS3 following the *max-SINR* and the *max-SA* criterium, respectively.

[1] A reuse factor equal to 1 is challenging especially when BSs densification increases, but it enables key benefits of HetNet deployment, even if requires suitable strategies to manage interference as the one we propose here.

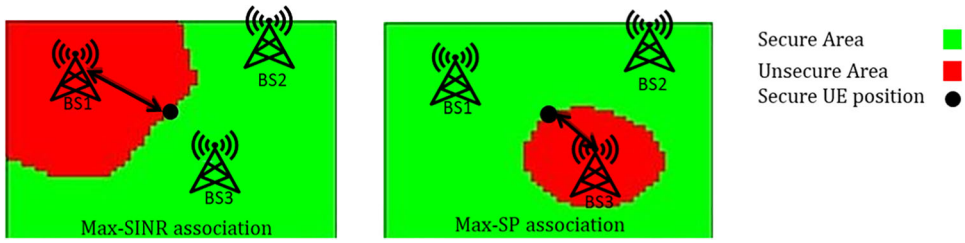

Fig. 2 Example of secure area achieved with max-SINR and max-SA association policies [12]

4.2 Radio Resources Allocation Scheme

As stated before by suitably assigning the physical resource blocks (PRBs) to UEs belonging to different neighbour cells it is possible to jam the eavesdropper thus reducing its capacity of decoding the signal. Toward this goal spatial dimension can be exploited together with time-frequency dimensions. We assume that each BS transmits data toward its associated UEs using a Maximum Gain beamforming, thus maximizing the SINR in the direction of the intended UE. Being the eavesdropper position unknown the selection of the PRBs is based on the maximization of the secure area. In particular, BSs perform a coordinated scheduling (CS) selecting the most suitable group of UEs (i.e., each UE in the group is associated with a different BS) that sharing the same PRBs increase the secure area. However, calculating the secure area for each possible group of UEs in a HetNet is very complex, moreover it requires a lot of information that cannot be available: for all the possible PRBs allocations the SINR of the eavesdropper should be known in every point of the considered area. For this reason we propose to exploit only the knowledge of the Direction of Arrival (DoA) of the signals. In particular, the secure area can be related to the DoA as shown in Fig. 3. The figure represents two BSs and the associated UEs that are paired, that means they use the same time-frequency resources for communicating. There are three possible cases: *(i) irrelevant* - the transmissions toward UEs do not interfere each other, hence there is not influence on the determination of the secure area, *(ii) damaging* - the transmission of the paired UE interferes with the reception of the intended UE, thus worsening its SINR while the eavesdropper (Eve) SINR is not affected, thus the secure area is reduced, *(iii) beneficial* - the transmission of the neighbour BS toward its UE (i.e., the paired UE) creates interference in the critical area (near to the serving BS) where Eve can intercept the signal directed to the intended UE, this is a beneficial interference that increases the secure area.

This is shown as an example in Figure 4 where the secure area is represented as a function of the difference between the DoA of the intended UE and the paired UE respect to the neighbour BS, ϕ. In this figure we can see the three cases described before.

Consequently, instead of maximizing the secure area that can be unpractical, a *pairing gain* is defined, that is simply based on DoA knowledge: the maximum gain is achieved when the paired UEs satisfy the *beneficial* condition, while no gain is achieved in the *irrelevant* case, and finally a negative gain (i.e., a cost) is obtained in the *damaging* case. The goal is to maximize the *sum of the pairing gains* of all the UEs. Fig. 5 shows the *secure area* as a function of the active users per cell, achieved in a scenario with two cells using an optimal matching approach to pair UEs belonging to the two cells. In particular, the optimal matching is made using as function to be optimized either the sum of the proposed pairing gains and the sum of the effective secure area (this is derived assuming to have the knowledge of the SINR of the eavesdropper in any point of the area in any configuration). Moreover, a greedy allocation is considered where each cell selects the PRB where to allocate the UE based only on its own pairing gain. It is possible to see that the proposed pair-

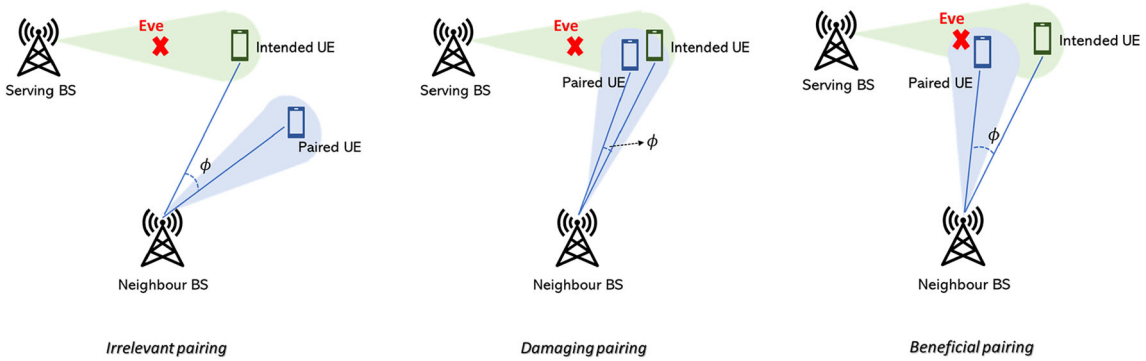

Fig. 3 UEs pairing effects for different DoA configurations

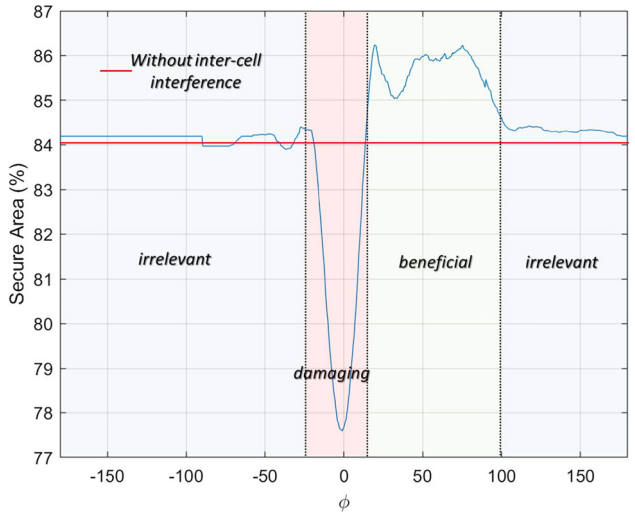

Fig. 4 Secure area (%) as a function of the angle ϕ (deg) depicted in Fig. 3. Angle ϕ is defined as the difference between the DoA of the intended UE and the paired UE respect to the neighbour BS

ing gain allows to follow the optimal allocation behaviour with a significant gain respect to the greedy approach. Users are randomly distributed in the considered area, and we can see that there is an increase of the secure area when the number of users per cell increases, this is because there are more pairs of users with different positions in the area that can be selected by the algorithm.

4.3 Rationale to Use RL in RRA

RL techniques can be used to solve a Markov Decision Problem (MDP). An MDP is defined via a dynamic environment, a state space $\mathcal{S}$, an action space $\mathcal{A}$, and a reward function $R(a, s)$ with $a \in \mathcal{A}$, $s \in \mathcal{S}$. In its basic setting, the decision maker gets a reward from the environment upon taking an action, and the environment changes its internal state. To take its actions, the decision maker follows a policy $\pi(s)$, i.e., a deterministic or stochastic action for each possible state. The objective is to find a policy such that the expected discounted reward is maximized. The discount factor $\gamma \in [0, 1)$ determines the length of the time horizon, where $\gamma \Longrightarrow 1$ means infinite horizon while $\gamma = 0$ means optimizing the current reward in a greedy manner. Hence, the goal of RL is to obtain an optimal policy to maximized the long-term weighted cumulative reward $T = \sum_{t=0}^{\infty} \gamma^t R(\pi(s), s)$, where t in general represents the time instant. Give the above, RL appears to be particularly suited for RRA problems when compared with supervised learning or model-bases solutions mainly in the following two situations:

- the optimum is unknown or very difficult to know, and only a reward associated to a given policy is available.
- The reward/loss function cannot be expressed in a closed-form as a differentiable function of the allocation decisions, i.e., the actions. Indeed, in this case the gradient descent method to train a neural network in supervised deep learning cannot be activated.

In the considered scenario, both the two conditions hold. Indeed, as discussed in the next Section, finding an optimal solution for the problem at hand is a very complex task which becomes unfeasible for large networks. Moreover, the reward, consisting in the sum of the pairing gains, cannot be definitely expressed as a differentiable function of the allocation decisions.

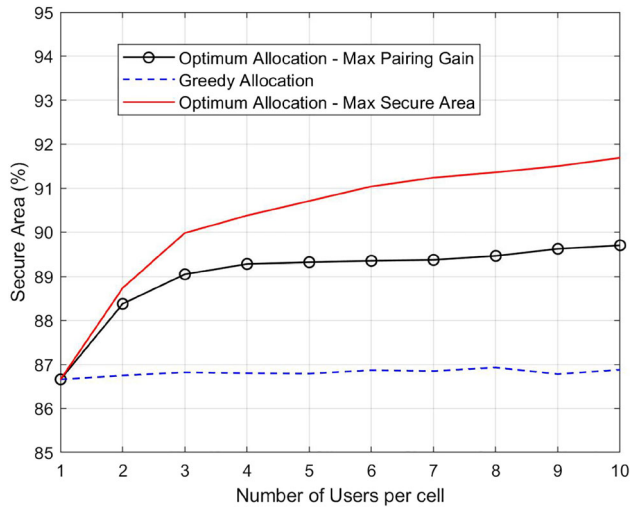

Fig. 5 Secure area vs number of users per cell

4.4 A Q-learning Framework for Optimizing Physical Security in HetNets

In the considered scenario, the environment is represented by an HetNet composed of C heterogeneous small-cells and a given number of UEs associated to the BSs, according to the max-SA association policy previously described [12]. In this setting, the reward $r_t = R(a_t, s_t)$ is the *sum of the pairing gains* defined before. It is worth noting that allocations among cells are intrinsically intertwined, i.e, a PRB allocation in cell t affects the rewards of each other cell. In this case, a possible solution could be to cyclically re-allocate the PRBs in each cell given the allocation of the other cells, mimicking a case of (eventually) infinite time horizon. However, the PRB allocation problem cannot be solved through distributed approaches, since a cell cannot evaluate the effect that its PRB allocations will have on the other cells. Accordingly, we make the reasonable assumption that each cell is able to evaluate its own reward simply knowing the DoA of its

UEs, and that such an information is delivered to a central controller that is in charge of PRB allocations of all cells. Note that in the considered setting, the central controller will perform PRB allocations without any other information apart from the cells local reward. This limits also the exchange of sensitive data (e.g. UEs positions) with the controller that could arise privacy issues. Moreover, even assuming that such an information is timely available, the centralized optimal PRB allocation problem is hard, and becomes unfeasible for large networks.

To sum up, in the considered scenario, the central controller carries out the PRB allocation in an iterative way, where the action a_t is the selection of the PRB to be allocated to each UE in cell t, and the state s_t is represented by the PRBs allocation of all users in the c-th cell, for all $c \neq t$. The procedure is cyclically repeated with a given pre-defined time-horizon (episode length). Owing to inter-dependencies among cells, a good policy should aim at maximizing a long-term reward, where long-term reward in this case must be intended as the weighted cumulative rewards of all cells in the system. Hence, the above formulation of the resource allocation problem for optimizing physical security naturally fits the general framework of RL. The estimate of the weighted cumulative reward for every state-action pair $\{s, a\}$ is obtained following a Q-learning approach where, at the end, the agent is able to establish the optimal policy for each state.

In particular, in Q-learning the Q-function is used to estimate the weighted cumulative reward for every state-action pair $\{s, a\}$. The Q function corresponding to the optimal policy can be computed from the Bellmans equation or by means of iterative approaches. In any case, the agent is able to establish the optimal policy by simply taking the action corresponding to the maximum Q for each possible state. For a better insight into Q-learning, the reader can refer to the broad literature on this topic.

In order to prove the effectiveness of the proposed approach, we consider a simple scenario with $C = 3$, $N = 4$ and $F = 4$, where C and N are the number of cells and UEs per cell,respectively, and F is the number of available PRBs. In this scenario, the number of possible actions is $N! = 24$ and the number of all possible states is $(N!)^{C-1} \times C = 1728$, corresponding to all the possible actions of the other cells multiplied by the number of cells. Such a number can be easily manageable by an iterative Q-learning approach where the Q-function is a table of size 1728×24. To prove the effectiveness of the Q learning solution we compare it with an optimum exhaustive approach where the optimum is evaluated computing the cumulative rewards for all the possible allocations. Note that the optimum allocation requires the knowledge of the DoAs of all users in a cell respect to the other cells. Moreover, we consider a greedy scheme where each cell optimize its own reward without considering the effects on the other cells, corresponding to a Q-learning scheme with $\gamma = 0$. In this case, the iterative algorithm always failed to converge and, hence, we consider the outcome obtained after a given number of iterations. The comparisons are shown on Table 1, where we report the final cumulative rewards (CR) averaged over 100 instances. Negative/positive values are obtained when the allocations of the adjacent cells reduce/improve the pairing gain of a cell. It is worth noting that the greedy approach on average lead to a reduction of the pairing gain, while the Q-learning scheme allows to approach the optimum where the PRB allocations allow to increase the pairing gain.

Table 1 Comparison in terms of average reward among Optimum, Q-learning and Greedy PRB allocations

Algorithm	Average Reward
Optimum	0.6422
Q-learning	0.5343
Greedy	−0.6550

Of course, in typical HetNet scenarios the number of cells and the number of PRBs can be much higher than 3 and 4, respectively. In this case, the Q-function could be properly estimated by deep neural networks (DNNs) through the establishment of a mapping between each state and the corresponding Q-values of all actions. The application of deep Q-learning (DQN) to the considered scenario is a subject that is currently under investigation by the same authors.

5 Conclusions

This paper presented an overview of the inherent physical-layer security features of heterogeneous networks. The paper mainly focused on exploiting the inter-cell interference to provide additional secrecy by optimizing the users associations and the resource allocation. In HetNets, the inter-cell interference arising from a strong radio resource reuse can be limited toward the legitimate user by means of suitable strategies, while it can be used to jam the eavesdropper link.

The user association and the resource allocation policies have been formulated so that the secure area is maximized. The framework is based only on the knowledge of the direction of arrival of the users signals at the base station and the optimization problem is efficiently solved resorting to a reinforcement Q-learning approach. Our scheme provides

near-optimum results with limited complexity and with limited knowledge of the network.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Wang L, Wong K, Jin S, Zheng G, Heath RW (2018) A new look at physical layer security, caching, and wireless energy harvesting for heterogeneous ultra-dense networks. IEEE Communications Magazine 56(6):49-55
2. Wong CY, Cheng RS, Lataief KB, Murch RD (1999) Multiuser OFDM with adaptive subcarrier, bit, and power allocation. IEEE J Sel Areas Commun 17(10):1747-1758
3.] K. I. Ahmed, H. Tabassum, and E. Hossain, Deep Learning for Radio Resource Allocation in Multi-Cell Networks, IEEE Network, November/December 2019
4. Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang, Deep Reinforcement Learning Based Resource Allocation for V2V Communications, IEEE Transactions in Vehicular Technology, Vol. 68, No. 4, 2019
5. Faris B. Mismar , Brian L. Evans , and Ahmed Alkhateeb , Deep Reinforcement Learning for 5G Networks: Joint Beamforming, Power Control, and Interference Coordination, IEEE Transactions on Communications, VOL. 68, NO. 3, 2020
6. Bloch M, Barros J (2011) Physical-Layer Security: From Information Theory to Security Engineering. Cambridge University Press
7. L. Mucchi, L. Ronga, X. Zhou, K. Huang, Y. Chen, and R. Wang, A new metric for measuring the security of an environment: The secrecy pressure, IEEE Trans Wireless Commun, vol. 16, no. 5, pp. 3416-3430, 2017
8. M. Kamel, W. Hamouda, and A. Youssef, Physical layer security in ultra-dense networks, IEEE Wireless Communications Letters, vol. 6, no. 5, pp. 690-693, 2017
9. S. Wang, Y. Gao, N. Sha, G. Zhang, and G. Zang, Physical layer security in k -tier heterogeneous cellular networks over nakagami-m channel during uplink and downlink phases, IEEE Access, vol. 7, pp. 14581-14592, 2019
10. M. Zhou, M. Xie, Y. Zhang, X. Jia, and L. Yang, Safeguarding non-best user association aided 5g k-tier hetnets using physical layer security, in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1-5
11. H. Wang, T. Zheng, J. Yuan, D. Towsley, and M. H. Lee, Physical layer security in heterogeneous cellular networks, IEEE Transactions on Communications, vol. 64, no. 3, pp. 1204-1219, 2016
12. D. Marabissi, L. Mucchi, and S. Casini, Physical-layer security metric for user association in ultra-dense networks, in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 487-491
13. Y. Wang, Z. Miao, and L. Jiao, Safeguarding the ultra-dense networks with the aid of physical layer security: A review and a case study, IEEE Access, vol. 4, pp. 9082-9092, 2016
14. X. Wang, M. Tao, J. Mo, and Y. Xu, Power and Subcarrier Allocation for Physical-Layer Security in OFDMA-Based Broadband Wireless Networks, IEEE Transactions on Information Forensics and Security, vol. 6, no. 3 pp. 693-702, 2011
15. K. Zhang, M. Peng, P. Zhang and X. Li, Secrecy-Optimized Resource Allocation for Device-to-Device Communication Underlaying Heterogeneous Networks, IEEE Transactions on Vehicular Technology, vol. 66, no. 2, pp. 1822-1834, 2017
16. M. Ahmed, Y. Li, Z. Yinxiao, M. Sheraz, D. Xu and D. Jin, Secrecy Ensured Socially Aware Resource Allocation in Device-to-Device Communications Underlaying HetNet, IEEE Transactions on Vehicular Technology, vol. 68, no. 5, pp. 4933-4948, 2019
17. , Z. Miao, and Y. Wang, Physical-Layer-Security-Oriented Frequency Allocation in Ultra-Dense-Networks Based on Location Informations, IEEE Access, vol. 7, pp. 90190-90205, 2019
18. G. Shiqi, X. Chengwen, F. Zesong, and K. Jingming, Resource allocation for physical layer security in heterogeneous network with hidden eavesdropper, China Communications, vol. 13, no. 3, pp. 82-95, 2016