A Hybrid Route Selection Scheme for 5G Network Scenarios: An Experimental Approach

Mohammad Kazem Chamran 1, Kok-Lim Alvin Yau 2,* Mee Hong Ling 1 and Yung-Wey Chong 3,*

1 Department of Computing and Information Systems, School of Engineering and Technology, Sunway University, Petaling Jaya 47500, Malaysia
2 Lee Kong Chian Faculty of Engineering & Science, Universiti Tunku Abdul Rahman (UTAR), Kajang 43200, Malaysia
3 National Advanced IPv6 Centre, Universiti Sains Malaysia (USM), Penang 11800, Malaysia
* Correspondence: yaukl@utar.edu.my (K.-L.A.Y.); chong@usm.my (Y.-W.C.)

Abstract: With the significant rise in demand for network utilization, such as data transmission and device-to-device (D2D) communication, fifth-generation (5G) networks have been proposed to fill the demand. Deploying 5G enhances the utilization of network channels and allows users to exploit licensed channels in the absence of primary users (PUs). In this paper, a hybrid route selection mechanism is proposed, and it allows the central controller (CC) to evaluate the route map proactively in a centralized manner for source nodes. In contrast, source nodes are enabled to make their own decisions reactively and select a route in a distributed manner. D2D communication is preferred, which helps networks to offload traffic from the control plane to the data plane. In addition to the theoretical analysis, a real testbed was set up for the proof of concept; it was composed of eleven nodes with independent processing units. Experiment results showed improvements in traffic offloading, higher utilization of network channels, and a lower interference level between primary and secondary users. Packet delivery ratio and end-to-end delay were affected due to a higher number of intermediate nodes and the dynamicity of PU activities.

Keywords: hybrid route selection; 5G; reinforcement learning; USRP/GNU radio; Raspberry Pi; centralized route selection; distributed route selection; cognitive radio; D2D communication; femtocell node and base station

1. Introduction

Fifth-generation (5G) has been envisaged as the next-generation cellular network for deploying, supporting, and scaling new technologies, including augmented reality, driver-less vehicles, Internet of Things, smart cities, virtual reality, and 3D video streaming services. Nevertheless, three main characteristics of the next-generation network scenario have posed main challenges to the realization of 5G. Firstly, dynamicity of channel availability, in which the operating channels, particularly the licensed channels, can be randomly occupied by licensed users (or primary users, PUs); consequently, unlicensed users (or secondary users, SUs) must search for and use the licensed channels in an opportunistic manner [1–5]. Secondly, heterogeneity, in which the network consists of different types of network cells (e.g., macrocells and small cells) and different types of nodes (e.g., using different operating channels and transmission power levels); consequently, nodes must adapt to a diversified operating environment [6–10]. Thirdly, ultra-densification, in which there are a large number of base stations (BSs), particularly small cells, and nodes in an area; consequently, nodes must search within nodes and BSs to find a route with the least traffic intensity in order to maximize traffic offloading [7,11–13].

The fifth generation (5G) is a multi-tier cellular network, as shown in Figure 1. There are different types of network cells, such as macrocells and femtocells. Macrocells have the broadest coverage, followed by femtocells (e.g., 30 m radius). In Figure 1, the BSs of
femtocells $f_{c1}, f_{c2}, ..., f_{c8}$ are scattered within the transmission range of a macrocell base station (MC BS). The BSs of the macrocell and femtocells can communicate with each other either directly or via multiple hops. Small cells deployed outside the transmission range of macrocells can help to improve network coverage. Due to the different characteristics of the network cells, the network can be segregated into two main layers, as shown in Figure 2. The **macrocell layer** uses lower frequency bands with a higher transmission power to provide a longer transmission range; however, the channel capacity is lower at lower frequency bands. The small-cell layer (i.e., the **femtocell layer**) uses higher frequency bands (e.g., millimeter wave (mm-wave) or above 30 GHz [14]) with a lower transmission power to provide a higher channel capacity; however, the transmission range is shorter at higher frequency bands. In terms of the dynamicity of channel availability, the macrocell layer may use channels with a higher number of white spaces, given a limited number of available licensed channels. In terms of heterogeneity and ultra-densification, these characteristics are more prevalent in small cells as a large number of small cells must be deployed to provide connections to heterogeneous nodes (or user equipment). Some of the main characteristics of 5G architecture are **control plane** and **data plane** separation as shown in Figure 2. The **control plane**, which includes the macrocell layer, contains a central controller (CC) that (a) collects network-wide information (e.g., network topology comprised of nodes, links, and channel availability) from all nodes in the network and maintains the information (e.g., in a routing table); (b) performs global-level tasks (e.g., determines routes); and (c) disseminates information (e.g., prioritizes routes) to nodes in the network (i.e., source node $f_{cs}$). The **data plane**, which includes the small-cell layer, performs local-level tasks (e.g., selects a route based on a set of available and prioritized routes, and the availability of **device-to-device** (D2D) communication) [15,16].

![Figure 1. A multi-tiered 5G network. The source node $f_{cs}$ establishes a route to the destination node $f_{cd}$. The optional routes are also shown.](image1)

![Figure 2. Control and data plane separation in a multi-tiered 5G network. An equivalent network is shown in Figure 1.](image2)
There are two main features of 5G. Firstly, dynamic channel access, whereby a SU node senses underutilized channels in licensed channels (or white spaces) owned by PUs, and subsequently accesses the white spaces in an opportunistic manner without causing unacceptable interference to PU activities in order to improve the overall spectrum utilization. Secondly, D2D communication, whereby a node can communicate directly with its neighboring nodes without passing through a BS, which helps to (a) offload traffic from highly-utilized BSs, particularly the MC BSs, to improve load balancing and reduce traffic congestion at BSs, and (b) extend coverage.

This paper proposes a cognition-inspired hybrid route selection scheme called CenTri for 5G network scenarios to embrace these two main features. CenTri consists of centralized and distributed route selection mechanisms and establishes multi-hop routes across the macrocell and small-cell layers. The centralized route selection mechanism by the CC adopts dynamic channel access to address the dynamicity of channel availability. In contrast, the distributed route selection mechanism adopts D2D to address heterogeneity and ultra-densification. Cognition enables a decision-maker (or an agent) to observe the respective operating environment and learn about the right decisions on the route selection in the operating environment at any time instant. In CenTri, cognition is embedded in the source node, which is the final ‘decision maker’ to learn about and select the best possible route with a high amount of white spaces to facilitate the traffic offload from the macrocell through traffic distribution to the femtocell layer. The CC sends a list of routes, which are ranked (or prioritized) based on the route length (or the number of hops), to the source node. The source node reranks the routes while considering the traffic offload, using the femtocell layer and the congestion level of the MC BS.

1.1. Why Is a Hybrid Route Selection Scheme Crucial to 5G Networks?

While separate centralized mechanisms [6,7,17] and distributed mechanisms [18–20] for route selection are shown to improve network performance, the benefits of combining both mechanisms are yet to be discovered in the context of 5G, and so a hybrid route selection scheme is the focus of this paper. Intuitively, a hybrid route selection scheme can address the intrinsic limitations of each mechanism. In the centralized mechanism, the computational complexity and routing overhead increase with the number of nodes in the network (or node density) as each source node in the data plane must receive information from the CC and then discover and maintain routes (e.g., update the routing table) [9,21]. In the distributed mechanism, the routing overhead increases as each node in the data plane must report the information availability (i.e., the available channels and the bottleneck link rate) with neighboring source nodes [12,22,23]. The issues observed in both centralized and distributed mechanisms intensify with the increased dynamicity of channel availability, network heterogeneity, and ultra-densification. Our proposed hybrid mechanism uses both centralized and distributed mechanisms. The distributed mechanism performs the traditional centralized tasks, particularly those using highly dynamic information in a distributed manner to minimize high computational complexities at the CC. Naturally, the routing overhead reduces because BSs and nodes do not send dynamic information (e.g., the traffic level) to the CC. The centralized mechanism performs the traditional distributed tasks, particularly those requiring frequent information exchange among BSs and nodes in a centralized manner to minimize high routing overheads at BSs and nodes. Naturally, the routing overhead reduces because BSs and nodes do not exchange dynamic information among themselves.

CenTri consists of centralized and distributed mechanisms for route selection. The centralized mechanism is embedded in a CC in the control plane, gathers and maintains network-wide and less dynamic information (i.e., available routes), and selects routes across different layers (i.e., the macrocell and small-cell layers). Subsequently, the source node makes routing decisions based on the list of routes provided by the CC. The routing decisions prioritize routes with fewer PUs in the data plane and D2D communication. The purpose is to minimize SU interference to PUs at the global level. In addition, being more
stable, the routes can be established in a proactive manner to serve as backbone routes for different source–destination pairs, contributing to a reduced routing overhead required in the route discovery and maintenance. The distributed mechanism, which is embedded in each BS and node in the data plane, gathers and maintains neighborhood (or local) and more dynamic information (i.e., available routes and traffic levels of nodes in range), and selects intra- or inter-cell routes in the small-cell layer. The purpose is to maximize traffic offload from BSs, particularly the MC BSs, at the local level. Being more dynamic, the routes can be established via D2D among BSs and nodes in the small-cell layer in a reactive (or on-demand) manner to offload traffic from the macrocell layer.

1.2. Why Is Cognition Crucial to the Hybrid Route Selection Scheme in 5G Networks?

Due to the need for interaction with the operating environment, the majority of testbeds using USRP have adopted Q-learning to adapt to the environment. Q-learning is the preferred approach because it does not need datasets required in supervised machine learning approaches and does not explore the underlying pattern or relevancy. A clear justification is needed (e.g., utilized or unutilized), which is why an unsupervised ML is not preferred [24–26]. For more details, refer to Section 3.1.

Due to the dynamicity of the operating environment, it is necessary to practice constant learning to achieve optimal network performance as time passes. As different types of network cells have different characteristics, the dynamicity level varies across the macrocell and small-cell layers. For example, the dynamicity of channel availability is lower in the macrocell layer as channels with a higher number of white spaces have been assigned to the layer. Therefore, a single set of rules or policies would less likely be optimal when applied across the entire network. In CenTri, BSs and nodes embrace a popular cognition approach called reinforcement learning (RL).

1.3. CenTri as a Hybrid Route Selection Scheme: An Overview

In CenTri, the centralized mechanism enables a CC to establish backbone routes that minimize SU interference to PUs based on network-wide information (i.e., channel availability of routes). The backbone routes can be used by different source–destination pairs. Nevertheless, the centralized mechanism has two shortcomings: (a) the CC is not sensitive to the dynamicity of the local environment (i.e., the traffic level of each source node towards its destination node) and (b) the routes predetermined and disseminated to source nodes by the CC may suggest the best route being the shortest one, which goes through the backbone route, causing an increased traffic load at the MC BS. The distributed mechanism enables nodes of small cells to revise routing decisions received from the CC to maximize traffic offload from MC BSs based on local information (i.e., available routes and traffic levels of D2D routes).

In the small-cell layer, a source node selects a route based on four priority levels, which helps to offload traffic from the macrocell layer: the first (or highest) priority is communication via the femtocell layer (through a D2D route), the second priority is communication via the route with minimum interference from PUs, the third priority is communication via the route with the minimum number of hops, and the fourth (or last) priority is communication via the backbone route. A single multi-hop route can consist of links belonging to different layers. For example, the communication through the backbone route connects a source node of the femtocell layer to a BS of the macrocell layer. It then transmits the data to the destination node in the femtocell layer. In Figure 1, the $f_{c_s} - f_{c_2} - f_{c_5} - f_{c_d}$ route is preferred compared to $f_{c_s} - mc - f_{c_d}$ used in the traditional network.

RL is embedded in the source nodes of the femtocell layer so that the right decisions can be made on the route selection at the local level to reduce the global workload of the CC.

1.4. Reinforcement Learning: An Overview

Q-learning, which is a popular RL approach, enables an agent (or decision maker) to gain knowledge independently in order to take the right action at the right time in its
operating environment for individual performance enhancement. At any time instant \( t \), an agent \( i \) observes its operating environment in the form of state \( s_i^t \), and then selects and takes action \( a_i^t \) in the operating environment. At the next time instant \( t+1 \), the agent \( i \) receives a reward \( r_i^{t+1}(s_i^{t+1}) \) and the state changes to \( s_i^{t+1} \leftarrow s_i^t \). The Q-value \( Q_i(s_i^t, a_i^t) \), which represents the long-term reward of a state–action pair \((s_i^t, a_i^t)\) of an agent \( i \), is updated using the Q-function, as follows:

\[
Q_i^{t+1}(s_i^t, a_i^t) = Q_i^t(s_i^t, a_i^t) + \alpha \left[ r_i^{t+1}(s_i^{t+1}) + \gamma (Q_i^{t+1}(s_i^{t+1}, a_i^{t+1}) - Q_i^t(s_i^t, a_i^t)) \right]
\]

where \( 0 \leq \alpha \leq 1 \) represents the learning rate and \( 0 \leq \gamma \leq 1 \) represents the discount factor, which is the next state–action pair emphasizing the future reward. The discount reward always has a lesser weight than the immediate reward.

1.5. Contributions

Common routing mechanisms, such as route requests and route replies, have been well investigated in the literature [27–30]. This paper focuses on route selection and makes three contributions. Firstly, a hybrid route selection scheme called CenTri is proposed to select routes that minimize SU interference to PUs in available licensed channels in 5G networks characterized by the dynamicity of channel availability and ultra-densification. The purposes are to: a) improve load balancing through traffic offload from the macrocell layer to the small-cell layer; and b) the overall spectrum utilization. Secondly, RL models and algorithms are proposed for CenTri. Thirdly, the issues associated with implementing CenTri on a real-world platform consisting of heterogeneous nodes embedded with universal software radio peripheral (USRP) units are presented.

1.6. Organization of This Paper

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 presents the system model. Section 4 presents the CenTri RL model and algorithm. Section 5 presents the results and discussion.

2. Related Work

This section presents related work on routing in 5G networks, hybrid route selection schemes, the application of RL to hybrid route selection schemes, and the implementation of schemes on a real-world 5G platform consisting of USRP units.

2.1. Routing in 5G

Based on the functions of the route selection schemes in 5G networks, there are two main categories: (a) traffic offloading [9,21,23,31,32]; and (b) traffic splitting by selecting routes comprised of nodes in the femtocell layer [21,31,33].

2.1.1. Traffic Offloading

Traffic offloading allows traffic to be offloaded from highly utilized BSs, particularly MC BSs to other BSs (e.g., in femtocells) in order to improve load balancing and reduce traffic congestion at BSs of the upper-tier (i.e., the macrocell layer). This can be achieved via D2D communication.

In [34], a distributed channel allocation scheme is proposed to offload traffic to D2D nodes in small cells to avoid (or reduce) interference between licensed and unlicensed users. The scheme addresses network heterogeneity (i.e., small-cell BSs and D2D users) and potential interference from unlicensed users. For channel allocation, channels are allocated to licensed users, and then the available channels are shared among D2D users communicating within the range of a small-cell network called a coalition. Therefore, the priority of spectrum sharing is higher among D2D devices in a coalition. The proposed
scheme increases the throughput and network spectrum utilization due to better channel allocation and route selection.

In [9], a distributed route selection scheme is proposed to offload traffic from highly congested BSs to less congested BSs via D2D communication in order to achieve load balancing. The scheme addresses network heterogeneity (i.e., macrocell and small-cell BSs) and ultra-densification. Traffic can be offloaded from: (a) congested MS BSs to less congested small cells; (b) congested small-cell BSs to less congested small-cell BSs; and (c) from congested small-cell BSs to less congested MC BSs. Deploying a large number of small cells can lead to ultra-densification. Dijkstra's algorithm [35] is applied to determine the shortest route (which has the least number of nodes and hops) towards a destination node. The proposed scheme increases the throughput due to better traffic distribution among the network cells.

In [32], a distributed route selection scheme is proposed to offload traffic from nodes with lesser resources (i.e., residual energy and memory) to those with more resources via D2D communication in order to improve network performance. The scheme addresses the dynamicity of channel availability and network heterogeneity, whereby nodes have different amounts of resources. Since routes have different resource requirements, nodes with higher residual energy and memory are selected to establish resource-intensive routes to deliver more traffic; however, once their resources drop below a certain threshold, the nodes are selected for less resource-intensive routes. The proposed scheme is shown to increase the packet delivery rate as nodes with more resources are selected for packet forwarding.

In [23], a distributed route selection scheme is proposed to offload traffic from nodes with low residual energy to those with higher residual energy yet less traffic load via D2D communication in order to prolong the network lifetime. The scheme addresses the dynamicity of channel availability and network heterogeneity, whereby nodes have different residual energy levels. In other words, nodes with higher residual energy levels are well utilized, while nodes with lower residual energy levels serve the backup role. Rerouting is performed from time to time as the residual energy levels at different nodes reduce at different rates, and the traffic load is dynamic in nature. Among the available routes, the one with higher residual energy and lower end-to-end delay, in terms of the number of hops, is selected to deliver traffic. The proposed scheme is shown to increase the network lifetime and packet delivery rate as load balancing is achieved among nodes with different residual energy levels throughout the network.

2.1.2. Traffic Splitting among Available Routes

Traffic splitting allows traffic to be split and distributed among different routes in order to improve load balancing and reduce traffic congestion at BSs serving large traffic streams. Traffic splitting is possible in 5G networks as the CC is aware of source nodes containing a particular traffic stream, as well as the available routes towards a destination.

In [21,31], a centralized route selection scheme is proposed to split and distribute large traffic streams among available routes in order to achieve load balancing. The scheme addresses the dynamicity of channel availability. The CC: (a) gathers network-wide information (i.e., network topology comprised of nodes and links, resources comprised of channel availability, and traffic patterns); (b) identifies source nodes that possess a particular traffic stream requested by a destination node; (c) estimates the channel capacity and end-to-end delay of routes established between a pair of source and destination nodes; and (d) dynamically segregates a traffic stream into different fractions and sizes contributed by different source nodes based on the channel capacity and the amount of traffic of the candidate routes. Based on the routing decisions made by the CC, a source node sends parts of a traffic stream in fractions to a destination node. Intermediate nodes that receive the fractions from different source nodes aggregate them, and subsequently split them among available outgoing links towards a destination based on the shortest route, in terms of the number of links and the current traffic amount of each route. The destination node aggregates the fractions to obtain the original traffic stream. The proposed scheme is shown
to reduce end-to-end delay. A similar traffic splitting scheme has been presented in [33], although the intermediate nodes that receive the fractions do not aggregate and split them. In [36], intermediate nodes receive updated routing tables to address the split traffic to the right destination. The split traffic scheme is shown to improve throughput and reduce handover and end-to-end delay.

2.1.3. CenTri for Achieving Traffic Offloading

CenTri enables BSs and nodes of small cells to maximize the traffic offload from MC BSs (see Section 2.1.1). In the next section, the 5G characteristics of CenTri are presented.

2.2. Hybrid Routing Schemes in Wireless Networks

Hybrid route selection schemes consist of centralized and distributed mechanisms that cooperate to determine the least cost routes (e.g., with lower traffic loads and the number of hops).

In [37], a hybrid route selection scheme consists of a centralized mechanism at the control plane and a distributed mechanism at the data plane to shift traffic intensity from congested routes to routes with less traffic. The route update procedure consists of two mechanisms: (a) proactive route updates from the CC with minimum delay; and (b) reactive route updates from the CC after data transmission completes in a node in the data plane. Therefore, the proposed algorithm provides an ingress threshold to determine the congestion level of each node before a route update takes place. The proposed scheme is shown to improve load balancing and increase the network throughput.

In [38,39], a hybrid route selection scheme, which consists of proactive and reactive mechanisms, is proposed to establish routes among two types of nodes, namely static and mobile nodes. Firstly, static nodes, which have lower dynamicity and energy constraints, perform: (a) proactive route selection to establish backbone routes towards a gateway node for internet access; and (b) reactive route selection to establish routes with mobile nodes. Secondly, mobile nodes, which have higher dynamicity and energy constraints, perform reactive route selections to establish better routes (i.e., with lower traffic loads and higher residual energy) towards gateway nodes. The proposed approach is shown to increase throughput, and reduce packet loss rate and end-to-end delay. A similar hybrid route selection scheme was proposed in [39]. The mobile nodes perform the reactive route selection to establish routes with a higher number of static nodes (or a lower number of mobile nodes) in order to improve route stability; the proposed scheme is shown to improve throughput and reduce the delay incurred in the route selection and the energy consumption of mobile nodes.

In [40,41], a hybrid route selection scheme, which consists of centralized and distributed mechanisms, is proposed to offload traffic from routes with busier nodes to routes with less occupied nodes. Nodes are segregated into groups, and several nodes in a group are selected as group leaders to communicate with neighboring groups. For intra-group communication, each node selects a route with a lower number of hops and traffic load in a distributed manner. For inter-group communication, with more than a single group leader, a source node can establish more than a single route towards a destination node. The CC selects a route with the lowest number of hops or traffic load and communicates the selected route to the source node. The proposed scheme is shown to increase throughput and channel utilization, as well as reduce the packet loss rate caused by route breakages.

CenTri is a hybrid route selection scheme that consists of centralized and distributed mechanisms. The centralized mechanism adopts a proactive approach to establish backbone routes using network-wide information in order to search for and use white spaces. In contrast, the distributed mechanism adopts a reactive approach to further improve the routes using local information in order to perform traffic offloading. In comparison with existing works, (a) CenTri considers heterogeneous nodes and networks as seen in 5G, it is different from [37–41] in terms of the control plane and data plane separation and the platform experimental setup with up to 11 nodes; (b) the traffic offloading mechanism makes it different
from [9,23,32,34] as the platform offers a coexistence of proactive–centralized and reactive distributed; (c) the use of D2D communication of the femtocell with a backbone route and the capability of learning the most efficient route with the least presence of PUs and dynamicity of channel availability, which distinguish the platform from [20,23,37,42–46].

2.3. Application of Reinforcement Learning to Hybrid Routing

In [42], a hybrid route selection scheme comprised of proactive and reactive mechanisms is proposed to establish stable routes from a source node to a destination node. There are two zones: (a) the intra-zone consists of nodes located within the transmission range of the node, and (b) the inter-zone consists of nodes located out of the transmission range of the node. The proactive mechanism establishes routes to the inter-zone nodes, while the reactive mechanism establishes routes to the intra-zone nodes. The traditional Q-learning approach is embedded in the proactive mechanism of each node to adjust the routing period, which is the time period of the next route discovery since the last one. A longer routing period is suitable for a network with lower dynamicity since existing routes are likely to function. In comparison, a shorter routing period is suitable for networks with higher dynamicity since existing routes are likely to be broken, so new routes must be established. By adjusting the routing period, the number of route discoveries can be adjusted according to network dynamics (i.e., mobility, residual energy level, and channel availability). In this RL scheme, the state represents the destination node, the action represents the selection of the routing period, and the reward represents the network dynamics. The proposed scheme is shown to increase the packet delivery rate and reduce routing overhead.

In [47], a hybrid route selection scheme, composed of learning- and non-learning-based mechanisms, is proposed to establish stable routes from a source node to a destination node. The non-learning-based mechanism establishes least-cost routes using Dijkstra’s algorithm during normal operation, while the learning-based mechanism estimates the link cost when it changes. Subsequently, the estimated link cost is used by the non-learning-based mechanism to revise its routes. The traditional Q-learning approach is embedded in the learning-based mechanism of each node to provide accurate estimations of the link cost in order to minimize route oscillation, whereby routes oscillate, between less utilized and favorable and highly utilized and unfavorable as time goes by. In this RL scheme, the state represents a link, the action represents the selection of the link cost, and the reward represents factors affecting the link cost. The proposed scheme is shown to reduce the packet loss rate and end-to-end delay.

In CenTri, the CC in the control plane establishes backbone routes in order to minimize SU interference to PUs. The nodes and BSs in the control plane have less dynamicity while the nodes and BSs in the data plane have high dynamicity. Therefore, the RL model is embedded in the distributed mechanism to enable BSs and nodes in the data plane to revise the proactive routes in order to offload traffic from highly utilized nodes and BSs, particularly MC BSs, to less utilized BSs, particularly small-cell BSs. The RL model of the distributed mechanism uses more dynamic local information (i.e., traffic levels at the BSs). In Section 4, the CenTri model and algorithm are presented.

2.4. Testbed Implementation of a Hybrid Route Selection Scheme

Universal software radio peripheral (USRP) testbeds have been deployed to perform route selection for 5G and cognitive radio networks. Together with an open-source software toolkit, namely GNU radio [48], signals are generated and various processes, including encoding, decoding, and modulation are performed. In this work, CenTri is implemented on a USRP/GNU radio testbed for 5G networks. While most experimental studies focus on data links and physical layer investigations, including interference mitigation [13,14,49], channel sensing [50], and power allocation [51,52], this paper focuses on network layer implementation. There are six nodes used in [43,44], eight nodes in [20], and ten nodes in [46].
In [20,43,44,46], the route selection scheme enables a SU to establish stable routes in the presence of PU activities in a multi-hop cognitive radio network. The routing metrics are channel availability [20,44,46,53], channel quality [43,44,46], interference among SUs [43], and the number of hops of the route [20]. In [46], RL is embedded in each node to establish stable routes with the highest possible channel available time. The SU must perform channel sensing from time to time to measure channel availability based on the number of available channels and the number of route switches. The SUs can learn about the availability of white spaces via channel sensing in opportunistic channel access, or via direct communication with PUs in channel leasing. To implement a larger multi-hop network with more nodes (e.g., >10 nodes), the USRP/GNU radio units are connected to a single computer via a switch [46], which helps to reduce hardware and software delays. The proposed schemes have been shown to increase throughput [20,43,44] (or packet delivery rate [46]), as well as reduce the end-to-end delay [20,43,44], the number of route breaks [46], and routing overhead caused by rerouting [43,44].

In this work, a heterogeneous network is considered in which there are up to eleven USRP/GNU radio nodes from different layers, including the macrocell and femtocell layers. Hence, the nodes have different characteristics in terms of the choice of operating channels, transmission power, and transmission range. Each node consists of a USRP/GNU radio unit, which serves as the RF front end, connected to a mini-computer Raspberry Pi3 B+, which is programmed to run CenTri. Thus, the testbed of this work extends those in the previous works [43,44,46], whereby the USRPs/GNU radio units are wire-connected with a single computer via a switch to emulate a common control channel. To the best of our knowledge, this is the first testbed implementation for a hybrid route selection scheme.

3. System Model

There is a set of channels \( C = \{c_1,c_2,...,c_{|C|}\} \), each \( c_i \) is occupied by PU \( p_j \in P = \{p_1,p_2,...,p_{|P|}\} \). A multi-tier 5G network shown in Figures 1 and 2 is considered. The MC BSs \( MC = \{mc_1,mc_2,...,mc_{|N|}\} \) and femtocell nodes and BSs, \( FC = \{fc_1,fc_2,...,fc_{|N|}\} \) can communicate among themselves directly. For instance, a femtocell BS \( fc_1 \) can communicate with a MC BS \( mc_1 \) directly via \( fc_1 - mc_1 \). A BS can establish a multi-hop route \( k_s \in K = \{k_1,k_2,...,k_{|K|}\} \) across different layers. For instance, a femtocell source node \( fc_s \) establishes a multi-hop route \( fc_s - mc_1 - fc_q \) to MC BS \( mc_1 \) and then to the femtocell destination node \( fc_d \). The traffic intensity \( \psi_{ts}^{k_s} \in \Psi = \{\psi_{t_1}^{k_1},\psi_{t_2}^{k_2},...\psi_{t_{|N|}}^{k_{|K|}}\} \) is one of the parameters that source nodes consider before selecting a route. The CC and the MC BSs are located in the control plane \( C^C \), while the FC BSs are located in the data plane \( D^C \).

A source node selects a route based on the PU activities, which can either be ON (busy) or OFF (idle). The ON/OFF duration of PUs in channel \( c_C \in C \) and link \( L = \{l_1,l_2,...,l_{|l|}\} \) follows the Poisson model, which is exponentially distributed with rates \( \lambda_{ON}^{c_p,pu} \) for active time and \( \lambda_{OFF}^{c_p,pu} \) for idle time of the PU activities, respectively. The channel availability \( V_{c_{l_1}}^{c_{L_{OFF}}} \) determines the average availability of the link in the time instant \( t \) as follows [54]:

\[
\Phi_{c_{L_{OFF}}}^{c_{L_{OFF}}} = \frac{\lambda_{ON}^{c_{pu}}}{\lambda_{ON}^{c_{pu}} + \lambda_{OFF}^{c_{pu}}} + \frac{\lambda_{OFF}^{c_{pu}}}{\lambda_{ON}^{c_{pu}} + \lambda_{OFF}^{c_{pu}}} \times e^{-(\lambda_{ON}^{c_{pu}} + \lambda_{OFF}^{c_{pu}})t} \tag{2}
\]

The PU and SU avoid any possible collision, but if that happens, it has to be less than the IEEE requirement [45]. In this setup, the appearance of PUs follows the Poisson model that creates a random pattern for ON and OFF. The SU node estimates the average channel available time \( \Phi_{c_{L_{OFF}}}^{c_{L_{OFF}}} \) of channel \( c_C \in C \) on a link of a route \( k_s \in K \). This ON–OFF time assignment is exponentially distributed and shows the duration of the PU transitions.
(the traffic in each channel). The duration of this random appearance follows the ON–OFF time period shown in Table 2. For example, the PU ON time period can be 50 s and the PU OFF time period can be from 50 to 250 s. This allows PUs to utilize their licensed channels whenever they want during their OFF time. Meanwhile, SUs utilize white spaces opportunistically.

The time horizon in a channel for SU is segregated into the sense–transmit time window [55]. The sensing time is the duration of the channel sensing and processing time. The processing time is the duration that the USRP/GNU radio takes for hardware and software initiation, such as packet encoding and decoding, digital conversion, and transmission or reception. The transmit time is the duration that a SU node takes to send or receive a data packet.

3.1. Reinforcement with Static Learning

In traditional RL (refer to Equation (3)), the learning mechanism has a constant rate that can be determined based on the importance of the current or discount values. The SU receives its reward based on the channel availability (or white spaces, or the idle status of PU). Specifically, the learning rate can be higher (lower) when the PU activity level is lower (higher), as represented by the equation below [46]:

$$Q_{i+1}^{t}(s_i, a_i) = [(1 - \alpha) \times Q_i^{t}(s_i, a_i) + \alpha \times \min(r_{i+1}^{t}(s_{i+1}))]$$  \hspace{1cm} (3)

where $r_{i+1}^{t}(s_{i+1})$ is the traffic intensity, which refers to the channel available time of a route $k_t \in K$. The traditional RL approach can be ideal for a less dynamic network with low PU activity levels so that no adjustment is required for learning.

3.2. Enhanced Reinforcement with Dynamic Learning

For a more dynamic network with frequent changes in the PU appearance in channels, an enhanced RL approach is required to adapt its learning to the dynamicity of the channel. For instance, the dynamicity of MC is less than that of the FC. Therefore, from MC to FC, the dynamicity increases, and the learning rate $\alpha$ is decreased. Equation (3) is enhanced as follows [46]:

$$Q_{i+1}^{t}(s_i, a_i) = (1 - \alpha(s_i, a_i)) \times Q_i^{t}(s_i, a_i) + \alpha \times \min(r_{i+1}^{t}(s_{i+1}))$$ \hspace{1cm} (4)

As the bottleneck link in a route has the least channel capacity, it helps in determining the priority of the D2D route over other routers in the network. A link with a lower channel capacity has a lower priority compared to other routes. The dynamic learning rate is $\alpha_{\text{min}} \leq \alpha(s_i, a_i) \leq \alpha_{\text{max}}$ and it is determined by the availability of channels (or white spaces) as follows:

$$\alpha_i(s_i, a_i) = \Phi_i^{\text{c}n_{\text{min}}}$$ \hspace{1cm} (5)

For $\alpha_i(s_i, a_i)$, a higher value shows a higher dependency of the Q-value on the current knowledge and a lower value shows a higher dependency of the Q-value on previous knowledge. The dynamic learning of $\alpha_i(s_i, a_i)$ can be varied based on the immediate state–action pair and previously learned rewards. In this study, the agent is myopic and relies on the immediate state–action pair with some consideration of the previously learned Q-value (rather than the next state–action pair (or discounted reward) due to the random appearance of PUs; hence, the discount value is $\gamma = 0$. Due to the need for interaction with the operating environment, the majority of the testbeds using USRP [33,37,46] have adopted Q-learning to adapt to the environment. Q-learning is the preferred approach because: (a) it does not need datasets required in supervised machine learning approaches [24–26] and (b) it provides definite outcomes, particularly whether a channel is utilized or unutilized, which
is preferred compared to unsupervised machine learning. The non-learning approach, called non-RL, selects routes based on their priority (e.g., based on the number of hops of the routes specifically routes $k_2 > k_3 > k_4$). Hence, the proposed dynamic Q-learning approach is compared with the traditional Q-learning approach and the non-learning approach.

In the following sections, the learning model and algorithm of CenTri are presented.

4. CenTri: Reinforcement Learning Model and Algorithm

In CenTri, the BSs of both macrocell and small-cell layers collaborate to perform the route selection. A proactive link selection mechanism is deployed in the control plane and a reactive route selection mechanism is deployed in the data plane.

Figure 3 presents the route selection mechanism of the RL model, which serves as the decision-making engine. The CC in the control plane selects routes based on factors with less dynamicity (i.e., a lower number of intermediate nodes and delay); while the nodes in the data plane (i.e., the source node $f_{c_1}$) select routes based on the priority levels of D2D communication (or with less PU interference). The amount of the delay is higher in D2D routes, so the distributed RL model embedded in the BSs and nodes of small cells aims to improve the routing decision made by the CC and select routes with lower PU activity levels and the number of intermediate nodes.

Figure 3. Route selection decision made by the source node based on RL.
The CC establishes backbone routes in order to provide an always available route for small-cell nodes. The backbone routes are shared among BSs in the data plane and, subsequently, the BSs share the backbone routes with source nodes. The distributed RL model determines the traffic intensity of available routes and establishes routes with the least traffic congestion via traffic offloading. BSs in small cells are part of the backbone routes from the macrocell, and so they inform the CC about their usage of white spaces. Routes are assumed to be disjointed in this paper for simplicity. The rest of this section presents the RL models and algorithms for CenTri.

4.1. Reinforcement Learning Models

This section presents the centralized proactive route selection mechanism with the distributed reactive RL model. While the CC establishes more stable routes and backbones, the distributed RL model learns to offload traffic from MC BSs in a collaborative manner.

4.1.1. Centralized Route Selection

In the centralized mechanism, a proactive route selection mechanism is deployed in the CC to establish backbone routes among MC BS and femtocell nodes $f_c$. Backbone routes, which do not have PU activities, help to maximize the packet delivery ratio in networks. The CC evaluates the available routes based on network-wide information gathered from the BSs and femtocell nodes $f_c$. The routes are prioritized as $k_1 = f_c - m_c - f_{d}$, $k_2 = f_d - f_{d}$, $k_3 = f_c - f_{d}$, $k_4 = f_{d} - f_{d}$, and $k_4 = f_c - f_{d} - f_{d} - f_{d}$. The evaluated routes from the CC are prioritized based on the number of hops without taking PU activities into consideration. The CC tends to give priority to the shortest route (i.e., the backbone route), which goes through MC BS. The routes are given to source nodes proactively to save processing times incurred in determining routes. Figure 3 shows that the proactive routes are provided by the CC to distributed source nodes in the data plane.

The source node re-prioritizes (or reranks) the given routes based on D2D communication with the minimum number of intermediate nodes (or the number of hops) and traffic intensity. Hence, the source node gives the lowest priority to the backbone route since it does not use D2D communication; this helps in distributing traffic from the MC BS. The presence of PUs has a direct impact on the throughput and delays the performance of packet transmission. Routes with lesser channel switches have lower signaling overheads, leading to higher stability and bandwidth availability [56].

4.1.2. Distributed Reinforcement Learning Model

In the data plane, the distributed RL model is embedded in all small-cell BSs and their corresponding nodes in the network. Hence, a source node in the network can select a route towards the destination in a reactive manner. Route selection in the data plane of the small cells follows the priority levels (refer to Section 1.3). The selected route has lower intermediate nodes and traffic intensity for maximizing traffic offload and achieving a higher throughput through increasing packet delivery rate. The channel capacity of a link in a route is determined by $\Phi_{c,n}^{l,k,n}$ from Equation (2) and it shows the utilization of links $l_{n}$ of route $k_{n}$ is defined as follows [57]:

$$U_{l_{n}}^{c,n} = \frac{\sum p_{i}^{c,n} c_{n}}{BW_{c,n}}$$

where $BW_{c,n}$ is the available bandwidth of the channel $c_{n}$ for the link $l_{n}$ in the femtocell layer. It is noteworthy that all D2D routes in this study have the same bandwidth but different frequencies, which means the link between two nodes uses different channels. The channel utilization of a route is defined as follows:
where $\Sigma(U_{f_c}^{k_n})$ is the sum of the channel utilization by PUs in the links $l_n$ of route $k_n$ with channels $c_n$ at time instant $t_n$. The channel utilization of a route includes the PU activities, and it is important for source nodes in the network to be aware of the traffic intensity of each route to reduce the possibility of interference from SUs to PUs.

Traffic intensity is particularly important when multiple routes from a source node have the same number of hops (or intermediate nodes). In the D2D type of communication, including node-to-node, node-to-BS, and BS-to-BS communications, the traffic intensity of a route can be calculated as long as the backbone routes are not utilized. For instance, in Figure 2, the source node $f_{c_s}$ and the destination node $f_{c_d}$ are three hops away in route $k_2 (f_{c_2} - f_{c_3} - f_{c_6} - f_{c_9})$. However, for routes $k_3 (f_{c_2} - f_{c_3} - f_{c_4} - f_{c_7} - f_{c_9})$ and $k_4 (f_{c_5} - f_{c_2} - f_{c_5} - f_{c_8} - f_{c_9})$, both routes have the same number of intermediate nodes and links (i.e., four hops). Hence, traffic intensity is used to select a route with a lower congestion level. In this example, PUs appear in route $k_3$, then route $k_4$ is used. In such a case, route $k_4$ continues to be used until communication ends or interference occurs.

However, the presence of PUs in a route (i.e., $k_3$) makes the route unavailable and recorded as occupied. For a source node that is required to send a data stream to a destination node, it has the initial roadmap comprised of available routes toward the destination node. These routes include nodes and BSs from both macrocell and femtocell layers. Based on the communication priority levels, the source node rearranges the given list by the CC and prefers routes using the femtocell layer. The higher the number of intermediate nodes, the greater the number of links in the route and the more possibility of PUs appearing in its channels. Hence, the route with a lower number of intermediate nodes tends to be selected.

The traffic intensity of a route is proportional to its channel capacity. A greater channel capacity at the bottleneck link of a route determines the channel capacity in the absence of PUs, as follows:

$$\Phi_{t_n, k_n}^{s_n, k_n, \beta, \text{OFF}} = \max_{\forall h_n \in K} \min_{l_n \in L} \left( \Phi_{t_n, l_n, k_n}^{s_n, k_n, \beta, \text{OFF}} \right)$$

where $\beta$ signifies the bottleneck link of a route. The traffic intensity $\psi_{t_n}^{k_n}$ of a route $k_n$ in the femtocell layer at time instant $t_n$ is the cumulative of the bottleneck channel capacity in the absence of PUs, as follows:

$$\psi_{t_n}^{k_n} \leftarrow \Sigma(\Phi_{t_n, k_n}^{s_n, k_n, \beta, \text{OFF}})$$

The traffic intensity of a route is proportional to its channel capacity. A greater channel capacity in a route has lesser PU interference and vice-versa. The source node learns about the dynamic changes of traffic intensity and adapts the learning rate accordingly. For instance, the greater traffic intensity in route $k_1$ at the time instant $t_1$ causes the source node to reduce the learning rate $\alpha(s_{t_1}^{k_1}, a_{t_1}^{k_1}) \rightarrow 0$ closer to zero, which makes the Q-function more dependent on the previous Q-value. The lower traffic intensity causes the source node to increase the learning rate $\alpha(s_{t_1}^{k_1}, a_{t_1}^{k_1}) \rightarrow 1$ closer to one, which makes the Q-function more dependent on the current Q-value. The source node makes a route selection decision by choosing a route with the maximum Q-value, which has the minimum PU activities as follows:

$$a_{t_n}^{k_n} = \max_{k_n \in A} \left( Q_{t_n}^{k_n} \left( s_{t_n}^{k_n}, a_{t_n}^{k_n} \right) \right)$$

The dynamic learning rate $\alpha_{\text{min}} \leq \alpha(s_{t_n}^{k_n}, a_{t_n}^{k_n}) \leq \alpha_{\text{max}}$ is as follows:

$$\alpha(s_{t_n}^{k_n}, a_{t_n}^{k_n}) \leftarrow \psi_{t_n}^{k_n}$$
A source node is equipped with the Q-routing model to rank the D2D routes. The best route toward the destination node is based on the number of intermediate hops and traffic intensity of available routes. Table 1 shows the RL model of the reactive mechanism embedded in BSs and nodes. In this model, state $s^{k_n}_t \in S = \{ k^{MC}, k^{fc_1}, k^{fc_2}, ..., k^{MC, fc_n} \}$ represents the given route from the CC at the time instant $t_1$. Action $a^{k_n}_t \in A = \{ k^{fc_1}, k^{fc_2}, ..., k^{fc_n} \}$ represents a set of actions in which a source node $fc_\alpha$ selects a D2D route $k_n \in K$ with the highest priority level at the time instant $t_1$. Reward $r^{k_n}_{t+1}$ represents the cost reflecting the traffic intensity of a route from the source node to the destination node.

| Table 1. The RL model for the route selection embedded in the BS of the small cell and its corresponding nodes in the data plane. |
| --- |
| State | $s^{k_n}_t \in S = \{ k^{MC}, k^{fc_1}, k^{fc_2}, ..., k^{MC, fc_n} \}$ represents the given route from the CC at the time instant $t_1$. |
| Action | $a^{k_n}_t \in A = \{ k^{fc_1}, k^{fc_2}, ..., k^{fc_n} \}$ represents a set of actions in which a source node $fc_\alpha$ selects a D2D route $k_n \in K$ with the highest priority level at the time instant $t_1$. |
| Reward | $r^{k_n}_{t+1}$ represents the traffic intensity $\Psi^{k_n}_t$ of the selected route when PU is in the OFF state at the time instant $t + 1$. |

Based on Table 1, three criteria are checked between source and destination nodes prior to communication in the data plane: (a) the type of communication (i.e., D2D); (b) the number of intermediate nodes; and (c) traffic intensity. The first criterion uses the available white spaces in channels and intends to offload traffic from MC BS. The second criterion helps to make route decisions more efficient by looking at routes with a lesser number of intermediate nodes, which helps to reduce the possibility of the PU appearance, leading to a higher successful transmission rate. The third criterion selects the route with fewer PU activities when two routes are identical in terms of the number of hops and channels. The route with fewer PU activities has a lower traffic intensity and a higher Q-value, so it is preferred.

In Figure 4, an example of the random appearance of PUs in routes is illustrated for three transmission cycles, namely A, B, C. The PUs can occupy a channel of routes $k_2$, $k_3$, and $k_4$. The presence of PUs in one of the link channels of a D2D route can cause the entire route to be blocked, and the source node must select another route. In this figure, at time $t$, the source node selects a route to the destination node. Since routes $k_3$ and $k_4$ are occupied by PUs, route $k_2$ is selected. At time $t_1$, both routes $k_2$ and $k_4$ are occupied; therefore, route $k_3$ is selected. At time $t_2$, all D2D routes $k_2$, $k_3$, and $k_4$ are occupied, so the source node uses the backbone route through MC BS. At time $t_3$, route $k_3$ is occupied by PUs, but both routes $k_2$ and $k_4$ are available to the source node. The source node selects route $k_2$ as it has a higher priority due to a smaller number of intermediate hops and nodes compared to route $k_4$. At time $t_4$, route $k_2$ is occupied by PUs, and both routes $k_3$ and $k_4$ are available to the source node. In this case, the source node selects route $k_3$ as it has a higher priority over route $k_4$. During the communication between the source node and the destination node through D2D routes, the source node learns about the appearance of PUs and the availability of D2D routes. This makes route selection more accurate as time goes by. For instance, both D2D routes $k_3$ and $k_4$ have the same number of hops, but the source node has learned that route $k_3$ has a higher channel capacity, and so it has a higher Q-value compared to route $k_4$. Therefore, route $k_3$ is selected.
4.2. Reinforcement Learning Algorithm

In this section, the RL algorithm for the distributed mechanism is presented. All the nodes and BSs of small cells in the data plane are equipped with the RL algorithm and receive a route map from the CC proactively. This enables them to select the best route proactively based on priority levels (see Section 4.1.2), which helps to offload traffic from MC BS. In the data plane of the small cell, a source node transmits a data stream to a destination node in a selected route out of the given routes by the CC. During the transmission, if the route is interrupted by PUs, a second prioritized route is selected and the previous route is identified as a route with a high traffic intensity level. By continuing this process, a table is updated with route scores based on channel capacity at the end of each transmission cycle, which gives a clearer pattern of the random appearance of PUs.

Algorithm 1 provides a general route selection scheme in which the CC sends the initial prioritized routes through the MC BS to the source node in the data plane proactively. As for CC, it is assumed that routes are readily available and prioritized based on the number of hops.

Algorithm 1 General description of the route selection by the source node.

```
1: procedure HYBRID ROUTE SELECTION
2: CC sends initial prioritized routes to MC BS proactively
3: MC BS sends prioritized routes $K$, where $k_1$ has the highest and $k_k$ has the lowest priority, to a source node $f_{cs}$
4: for $(k_{D2D} \in K)$ do
5: Select route $k_1$ if it has the least hops and traffic intensity $\Psi$ at time instant $t$ when PU is OFF
6: if all $k_{D2D}$ have ON PUs then
7: Select the backbone route $k_{bb}$
8: end if
9: end for
10: end procedure
```

Algorithm 2 shows the distributed route selection mechanism for traffic offloading. Based on the flowchart shown in Figure 3, the source node receives a route map (i.e., a list of routes created proactively), which is prioritized based on the number of hops, from
the CC. The source node rearranges the list and gives priority to D2D routes in order to offload traffic from MC BS. The Q-value is dependent on traffic intensity, which is based on the channel capacity of the links of a route and the learning rate $\alpha$ that changes dynamically based on the PU activity level. Therefore, routes with lesser PU activities tend to be selected compared to those with higher PU activities.

![Experimental setup for the hybrid route selection with three D2D routes and a backbone route via MC BS.](image)

**Figure 5.** Experimental setup for the hybrid route selection with three D2D routes and a backbone route via MC BS.

**Algorithm 2** RL mechanism at distributed nodes and BSs.

```plaintext
1: procedure D2D ROUTE SELECTION
2: /* Source node selects a route to offload traffic from MC BS */
3: MCBS receives a route map (i.e., $\{k_{bb}, k_2, k_3, k_4\}$) from the CC and sends it to $fcs$
4: /* Stage 1 */
5: for time $t_1$; $fcs$ reprioritize D2D routes in the route map do
6:   \[ k = \{k_2 > k_3 > k_4 > k_{bb}\} \]
7:   if $k_2$ or $k_3$ or $k_4$ is not available then
8:       $fcs$ use backbone route $k_{bb}$
9:   end if
10: end for
11: /* Stage 2 */
12: for time $t_{n+1}$, $n \in |N|; fcs$ reprioritize D2D routes based on traffic intensity $\Psi_{tn}^{k_n}$ and Equation (4) do
13:   Estimate $\Phi_{tn,a_n}^{k_n}$ using Equations (2) and (8)
14:   if $\Psi_{tn}^{k_n} < \Phi_{tn,a_n}^{k_n} /\text{traffic intensity is not updated }$/ then
15:       update $\Psi_{tn}^{k_n} \leftarrow \Sigma(\Phi_{tn,a_n}^{k_n})$
16:   end if
17:   Calculate dynamic learning rate $\alpha(s_{tn}^{k_n}, a_{tn}^{k_n})$ using Equation (10)
18:   Update the Q-value using Equation (4)
19:   for $fcs$, do select a D2D route with the maximum Q-value from routes $\{k_2 > k_3 > k_4\}$
20: end for
21: end for
22: end procedure
```
4.2.1. Implementation Requirements and Parameters

The implementation has eleven USRP/GNU radio units as nodes and BSs. Each of the ten USRP/GNU radio units is connected with a Raspberry Pi3 B+ unit equipped with 30 GB of external memory for storing and running algorithms. The USRP unit, specifically USRP N200, is equipped with the VERT900 antenna, and the GNU radio runs an open-source software-defined radio (SDR). A personal computer, which is equipped with the core i7 processor and 16 GB RAM, serves as MC BS. The D2D nodes have closer proximity among themselves compared to MC BS, so the transmission power is 10 dBm (10 mW) among themselves and 20 dBm (100 mW) with the MC BS.

Table 2 presents the parameters. In this platform, the user datagram protocol (UDP) is the preferred transport layer protocol for multimedia applications because it is connectionless and it does not perform retransmission during packet loss, which reduces delay at the expense of the acceptable packet loss. Figure 5 shows the platform with USRPs equipped with RP3. Nodes are located in the MC BS proximity and receive route information proactively from the CC via MC BS.

| Category                  | Parameter                             | Value       |
|---------------------------|---------------------------------------|-------------|
| Experiment                | Duration                              | 900 s       |
|                           | Number of channels                    | 11          |
|                           | Number of USRP/GNU radio nodes        | 10          |
| USRP/GNU radio nodes      | Transport layer                       | UDP         |
| USRP                      | Channel bandwidth                     | 40 MHz      |
|                           | Sample rate                           | 1.1 MB      |
| USRP antenna              | D2D transmission power among nodes    | 10 dBm      |
|                           | Transmission power between node and MC BS | 20 dBm    |
|                           | Carrier frequency                     | 850 MHz     |
| RP3                       | Operating system                      | Ubuntu-Mate |
| PU activities             | PU ON time                            | 50 s        |
|                           | PU OFF time                           | \{50, 100, 150, 200, 250\} s |

4.2.2. Assumptions

The platform performs multi-hop communication from the source node to the destination node. There are a few assumptions in this setup as follows:

- The delay incurred in multi-hop communication is not considered in order to focus on routes with less traffic (i.e., with low PU activities).
- The backbone and D2D routes are readily available, and the source node re-prioritizes them.
- A D2D route is up to three hops, and the source and destination nodes do not have direct communication.

4.2.3. Appearance of PUs on Channels

Three PUs reappear in the operating channels of the D2D communication randomly. The backbone route $k_1$, which serves as a backup, is free from PU activities. When the channel of a route has PU activities, the route breaks and the source node must select another available route following the priority mechanism explained in Section 4.1.1. There are three scenarios related to the presence of PU activities in routes $k_2$, $k_3$, and $k_4$. In all scenarios, the destination node is beyond the transmission range of the source node, and so the traffic stream must go through the intermediate nodes of the network.
Scenario 1

In the first scenario, as shown in Figure 6, PUs reappear in route \(k_3\) \((f_{c_3} - f_{c_1} - f_{c_4} - f_{c_7} - f_{c_4})\) in a random manner. PU_1, PU_2, and PU_3 interfere with channels \(c_2, c_6, \) and \(c_9\), respectively. The source node selects either D2D routes \(k_2\) or \(k_4\), or the backbone route \(k_1\). Since route \(k_2\) has a higher priority due to a lower number of hops, it is selected.

Figure 6. Scenario 1: The appearance of PUs in route \(k_3\) interfere with channels \(c_2, c_6, \) and \(c_9\).

Scenario 2

In the second scenario, as shown in Figure 7, PUs reappear in routes \(k_3\) \((f_{c_3} - f_{c_1} - f_{c_4} - f_{c_7} - f_{c_4})\) and \(k_4\) \((f_{c_3} - f_{c_2} - f_{c_5} - f_{c_8} - f_{c_3})\). PU_1, PU_2, and PU_3 interfere with channels \(c_2, c_6, \) and \(c_3\), respectively. Specifically, two channels, \(c_2\) and \(c_6\) of route \(k_3\) and channel \(c_3\) of route \(k_4\), are occupied by PUs. The source node selects the D2D route \(k_2\) rather than the backbone route \(k_1\).

Figure 7. Scenario 2: The appearance of PUs in routes \(k_3\) and \(k_4\) interfere with channels \(c_2, c_3, \) and \(c_6\).

Scenario 3

In the third scenario, as shown in Figure 8, PUs reappear in all D2D routes, including routes \(k_2\) \((f_{c_3} - f_{c_3} - f_{c_6} - f_{c_4})\), \(k_3\) \((f_{c_3} - f_{c_1} - f_{c_4} - f_{c_7} - f_{c_4})\), and \(k_4\) \((f_{c_3} - f_{c_2} - f_{c_5} - f_{c_8} - f_{c_3})\). PU_1, PU_2, and PU_3 interfere with channels \(c_2, c_3, \) and \(c_4\), respectively. Specifically, channel \(c_2\) of route \(k_3\), channel \(c_3\) of route \(k_4\), and channel \(c_4\) of route \(k_2\) are occupied by PUs. Only the backbone route \(k_1\) is available to the source node.

Figure 8. Scenario 3: The appearance of PUs in routes \(k_2, k_3, \) and \(k_4\) interfere with channels \(c_2, c_3, \) and \(c_4\).
5. Results and Discussion

Simulation results, including the packet delivery ratio, end-to-end delay, throughput, and the number of route breakages, are presented.

5.1. Packet Delivery Ratio

The packet delivery ratio (PDR) is the ratio of the number of packets received by the destination node to the number of packets sent by the source node. In Figure 9, PDR increases for D2D routes \( k_1, k_2, \) and \( k_3 \) as the PU OFF time increases. When the PU OFF time increases from 50 to 250 s, the PDR of: (a) route \( k_2 \) increases from 0.891 (89.1%) to 0.929 (92.2%); (b) route \( k_3 \) increases from 0.853 (85.3%) to 0.915 (91.5%); and (c) route \( k_4 \) increases from the lowest at 0.844 (84.4%) to 0.912 (91.2%).

![Figure 9. Average packet delivery ratio comparison among D2D routes with dynamic \( \alpha \) at different PU OFF time.](image)

Route \( k_2 \) achieves a better PDR compared to routes \( k_3 \) and \( k_4 \) since it has a lower number of hops and PU activities. For routes \( k_3 \) and \( k_4 \), their PDRs are very close to each other and their gap reduces as the PU OFF time increases. This is because both routes have the same number of hops; however, route \( k_3 \) has a lesser presence of PUs, explaining why it is a preferred route over route \( k_4 \).

5.2. End-to-End Delay

The end-to-end delay is the time taken by a data stream to be transmitted from the source node \( f_c_s \) to the destination node \( f_c_d \). The three D2D routes have different numbers of intermediate nodes (or hops). Route \( k_2 = f_c_s - f_c_3 - f_c_6 - f_c_d \) has four nodes with two intermediate nodes and routes \( k_3 = f_c_5 - f_c_4 - f_c_7 - f_c_d \) and \( k_4 = f_c_5 - f_c_2 - f_c_5 - f_c_8 - f_c_d \) have five nodes with three intermediate nodes. Route \( k_2 \) has the lowest end-to-end delay among the D2D routes since it has the lowest number of intermediate nodes. Since routes \( k_3 \) and \( k_4 \) have the same number of intermediate nodes, the source node selects one over the other through the learning mechanism, explained in Section 4.2.3.

Figure 10 shows a comparison of the end-to-end delay incurred between the traditional reinforcement learning (TRL) mechanism with \( \alpha = 0.5 \) and the dynamic reinforcement learning (DRL) mechanism. In the traditional RL mechanism, the learning rate is constant at \( \alpha = 0.5 \) for the entire experiment. The end-to-end delay reduces with increasing PU OFF time. Compared to traditional reinforcement learning, dynamic reinforcement learning shows a lower end-to-end delay for routes \( k_2, k_3, \) and \( k_4 \) when the PU OFF time increases from 50 to 250 s.
5.3. Throughput

Throughput is the rate of the successful data stream delivered to the destination node through a selected route in a specific time frame. Figure 11 shows a comparison of the throughput achieved by the three D2D routes $k_2$, $k_3$, and $k_4$ for different PU OFF times. Based on the experimental results, the average throughput of the routes increases when the PU appearance reduces. Therefore, increasing the PU OFF time has a positive effect on the rate of the data stream transmission. The throughput of the dynamic reinforcement learning (DRL) mechanism, which has a dynamic learning rate $\alpha$ that varies with rewards, is affected by the PU appearance in a route. Specifically, when the PU OFF time increases from 50 to 250 s: (a) the throughput of route $k_2$ increases from 1.487 to 1.68 Mbps, (b) the throughput of route $k_3$ increases from 1.473 to 1.652 Mbps, and (c) the throughput of route $k_4$ increases from 1.468 to 1.642 Mbps. A similar trend is observed in: (a) traditional reinforcement learning (TRL) with a fixed learning rate of $\alpha = 0.5$; and (b) the non-learning approach, called non-RL (NRL), which selects routes using their priority $k_2 > k_3 > k_4$, which is based on the number of hops of the routes. Overall, DRL outperforms both TRL and NRL.

Figure 11. Average throughput comparison among routes $k_2$, $k_3$, and $k_4$ for dynamic RL with dynamic $\alpha$, traditional RL with $\alpha = 0.5$ and non-RL at a different PU OFF time.
5.4. Number of Route Breakages

The source node selects a route based on the priority given by the CC. However, the priority of routes with D2D communication changes with the presence of PUs and successful data transmission to the destination node. For each data transmission cycle, a route breakage occurs when a PU reappears in a selected route, and this causes the source node to switch to another route.

Figure 12 shows a comparison of the cumulative route breakage between DRL and TRL with a fixed $\alpha = 0.5$. Although TRL has a better performance with less route breakage at the beginning, the source node learns more about routes with higher successful transmission rates as time goes by, contributing to the improvement in DRL. When the PU OFF time increases from 50 to 250 s: (a) the route breakage of TRL reduces from 15.6 to 5.5 and (b) the route breakage of DRL reduces from 16.4 to 5.1.

![Figure 12. Cumulative number of route breakages between TRL with $\alpha = 0.5$ and the DRL mechanism at a different PU-OFF time.](image)

6. Conclusions and Future Work

This paper proposes CenTri, which is a hybrid route selection scheme that uses white spaces to offload traffic from macrocell to small-cell base stations with heterogeneous nodes in 5G network scenarios. It caters to important characteristics of 5G network scenarios, including the dynamicity of channel availability, heterogeneity, and ultra-densification. In this paper, device-to-device (D2D) communication uses traditional reinforcement learning (TRL) and dynamic reinforcement learning (DRL) approaches. While TRL uses a constant learning rate, DRL uses a dynamic learning rate that changes with primary user (PU) activity levels. Our work was tested in a testbed with eleven USRP/GNU radio units. Each USRP unit was embedded with a mini-computer called RP3 to provide more realistic scenarios. Compared to TRL, experimental results show improvement in different quality of service (QoS) metrics, including a higher packet delivery ratio, throughput, lower end-to-end delay, and the number of route breakages. Routes with higher intermediate nodes also achieved higher end-to-end delay but a lower packet delivery ratio and throughput.

In the future, CenTri will require more testing with a higher number of routes and intermediate nodes. Better processing units can relax the assumptions made, including the processing delay. Moreover, a cross-layer design for studying the physical data link and network layers will help to provide a more realistic testing environment.
Author Contributions: Conceptualization, M.K.C., K.-L.A.Y., M.H.L. and Y.-W.C.; methodology, M.K.C., K.-L.A.Y., M.H.L. and Y.-W.C.; investigation, M.K.C.; resources, M.K.C., K.-L.A.Y., M.H.L. and Y.-W.C.; data curation, M.K.C.; writing—original draft preparation, M.K.C.; writing—review and editing, K.-L.A.Y., M.H.L. and Y.-W.C.; visualization, M.K.C.; project administration, K.-L.A.Y. and Y.-W.C.; funding acquisition, K.-L.A.Y. and Y.-W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a publication fund under the Research Creativity and Management Office, Universiti Sains Malaysia. This research was also supported by Universiti Tunku Abdul Rahman.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chaudhari, A.; Murthy, C.S.R. Femto-to-Femto (F2F) Communication: The Next Evolution Step in 5G Wireless Backhauling. In Proceedings of the 2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Paris, France, 15–19 May 2017.

2. Boviz, D.; Member, S.; Chen, C.S.; Member, S.; Yang, S. Effective Design of Multi-User Reception and Fronthaul Rate Allocation in 5G Cloud RAN. *IEEE J. Sel. Areas Commun.* 2017, 36, 1825–1836. [CrossRef]

3. Jalil Piran, M.; Tran, N.; Suh, D.; Song, J.B.; Hong, C.S.; Han, Z. QoE-Driven Channel Allocation and Handoff Management for Seamless Multimedia in Cognitive 5G Cellular Networks. *IEEE Trans. Veh. Technol.* 2016, 66, 6569–6585. [CrossRef]

4. Sucasas, V.; Radwan, A.; Mumtaz, S.; Rodriguez, J. Effect of noisy channels in MAC-based SSDF counter-mechanisms for 5G cognitive radio networks. In Proceedings of the International Symposium on Wireless Communication Systems, Brussels, Belgium, 25–28 August 2015. [CrossRef]

5. Font-Bach, O.; Bartzoudis, N.; Mestre, X.; López-Bueno, D.; Mège, P.; Martinod, L.; Ringset, V.; Myrvoll, T.A. When SDR meets a 5G candidate waveform: Agile use of fragmented spectrum and interference protection in PMR networks. *IEEE Wirel. Commun.* 2015, 22, 56–66. [CrossRef]

6. Öhlén, P.; Skubic, B.; Rostami, A.; Ghebretensä, Z.; Mårtensson, J.; Fiorani, M.; Monti, P.; Wosinska, L. Data Plane and Control Architectures for 5G Transport Networks. *J. Light. Technol.* 2016, 34, 1501–1508. [CrossRef]

7. Carrasco, Ó.; Miatton, F.; Díaz, S.; Herzog, U.; Frascolla, V.; Briggs, K.; Miscopene, B.; Domenico, A.D.; Georgakopoulos, A. Centralized Radio Resource Management for 5G small cells as LSA enabler. *arXiv* 2017, arXiv:1706.08057.

8. Huang, H.; Guo, S.; Liang, W.; Li, K.; Ye, B.; Zhuang, W. Near-Optimal Routing Protection for In-Band Software-Defined Heterogeneous Networks. *IEEE J. Sel. Areas Commun.* 2016, 34, 2918–2934. [CrossRef]

9. Ambriz, S.J.G.; Mendez, R.M.; Angeles, M.E.R. 5GTraDis: A novel traffic distribution mechanism for 5G Heterogeneous Networks. In Proceedings of the 2016 13th International Conference on Electrical Engineering, Computing Science and Automatic Control, CCE 2016, Mexico City, Mexico, 26–30 September 2016. [CrossRef]

10. Jaber, M.; Imran, M.A.; Tafazolli, R.; Tukmanov, A. 5G Backhaul Challenges and Emerging Research Directions: A Survey. *IEEE Access* 2016, 4, 1743–1766. [CrossRef]

11. Tran, T.X.; Hajisami, A.; Pompili, D. Ultra-Dense Heterogeneous Small Cell Deployment in 5G and Beyond Cooperative Hierarchical Caching in 5G Cloud Radio Access Networks. *IEEE Netw.* 2017, 31, 35–41. [CrossRef]

12. Ge, X.; Tu, S.; Mao, G.; Wang, C.X.; Han, T. 5G Ultra-Dense Cellular Networks. *IEEE Wirel. Commun.* 2016, 23, 72–79. [CrossRef]

13. Wassie, D.A.; Berardinelli, G.; Catania, D.; Tavares, F.M.L. Experimental Evaluation of Interference Suppression Receivers and Rank Adaptation in 5G Small Cells. In Proceedings of the 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Boston, MA, USA, 6–9 September 2015.

14. Wassie, D.A.; Berardinelli, G.; Tavares, F.M.L. Experimental Verification of Interference Mitigation techniques for 5G Small Cells. In Proceedings of the 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, UK, 11–14 May 2015. [CrossRef]

15. Lai, W.K.; Shieh, C.S.; Chou, F.S.; Hsu, C.Y. Handover Management for D2D Communication in 5G Networks. In Proceedings of the 2020 2nd International Conference on Computer Communication and the Internet, ICCCI 2020, Nagoya, Japan, 26–29 June 2020; pp. 64–69. [CrossRef]

16. Ouali, K.; Kassar, M.; Nguyen, T.M.T.; Sethom, K.; Kervella, B. An efficient D2D handover management scheme for SDN-based 5G networks. In Proceedings of the 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 10–13 January 2020; pp. 1–6.

17. Okasaka, S.; Weiler, R.J.; Keusgen, W.; Pudeyev, A.; Maltsev, A.; Karls, I.; Sakaguchi, K. Proof-of-concept of a millimeter-wave integrated heterogeneous network for 5G cellular. *Sensors* 2016, 16, 1362. [CrossRef]
18. Jia, M.; Gu, X.; Guo, Q.; Xiang, W.; Zhang, N. Broadband Hybrid Satellite-Terrestrial Communication Systems Based on Cognitive Radio toward 5G. IEEE Wirel. Commun. 2016, 23, 96–106. [CrossRef]
19. Su, Z.; Xu, Q. Content distribution over content centric mobile social networks in 5G. IEEE Commun. Mag. 2015, 53, 66–72. [CrossRef]
20. Soltani, S.; Sagduyu, Y.; Shi, Y.; Li, J.; Feldman, J.; Matyjas, J. Distributed Cognitive Radio Network Architecture, SDR Implementation and Emulation Testbed. In Proceedings of the MILCOM 2015—2015 IEEE Military Communications Conference, Tampa, FL, USA, 26–28 October 2015; pp. 438–443. [CrossRef]
21. He, J.; Song, W. Evolving to 5G: A Fast and Near-optimal Request Routing Protocol for Mobile Core Networks. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014; pp. 4586–4591.
22. Taleb, T.; Samdanis, K.; Mada, B.; Flinck, H.; Dutta, S.; Sabella, D. On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Architecture & Orchestration. IEEE Commun. Surv. Tutor. 2017, 19, 1657–1681. [CrossRef]
23. Zhang, G.A.; Gu, J.Y.; Bao, Z.H.; Xu, C.; Zhang, S.B. Efficient Signal Detection for Cognitive Radio Relay Networks Under Imperfect Channel Estimation. Eur. Trans. Telecommun. 2015, 25, 294–307.
24. Hwang, R.H.; Peng, M.C.; Huang, C.W.; Lin, P.C.; Nguyen, V.L. An Unsupervised Deep Learning Model for Early Network Traffic Anomaly Detection. IEEE Access 2020, 8, 30387–30399. [CrossRef]
25. Caron, M.; Bojanowski, P.; Joulin, A.; Douze, M. Deep clustering for unsupervised learning of visual features. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 132–149.
26. Tsipi, L.; Karavolos, M.; Vouyioukas, D. An unsupervised machine learning approach for UAV-aided offloading of 5G cellular networks. Telecom 2022, 3, 86–102. [CrossRef]
27. Saleem, Y.; Yau, K.L.A.; Mohamad, H.; Ramli, N.; Rehmani, M.H. SMART: A SpectrUM-Aware ClusteR-based routing scheme for distributed cognitive radio networks. Comput. Netw. 2015, 91, 196–224. [CrossRef]
28. Saleem, Y.; Yau, K.L.A.; Mohamad, H.; Ramli, N.; Rehmani, M.H.; Ni, Q. Clustering and Reinforcement - Learning-Based Routing for Cognitive Radio Networks. IEEE Wirel. Commun. 2017, 24, 146–151. [CrossRef]
29. Tang, F.; Tang, C.; Yang, Y.; Yang, L.T.; Zhou, T.; Li, J.; Guo, M. Delay-Minimized Routing in Mobile Cognitive Networks for Time-Critical Automation Applications. IEEE Trans. Ind. Inform. 2017, 13, 1398–1409. [CrossRef]
30. Khatib, R.F.E.; Salameh, H.B. A Routing Scheme for Cognitive Radio Networks with Self-Interference Suppression Capabilities. In Proceedings of the 2017 Fourth International Conference on Software Defined Systems (SDS), Valencia, Spain, 8–11 May 2017; pp. 20–25.
31. He, J.; Song, W. Optimizing Video Request Routing in Mobile Networks with Built-in Content Caching. IEEE Trans. Mob. Comput. 2014, 15, 1714–1727. [CrossRef]
32. Guo, J.; Orlik, P.; Parsons, K.; Ishibashi, K.; Takita, D. Resource aware routing protocol in heterogeneous wireless machine-to-machine networks. In Proceedings of the 2015 IEEE Global Communications Conference, GLOBECOM 2015, San Diego, CA, USA, 6–10 December 2015. [CrossRef]
33. Rakshit, K.; Rao, M. Routing Protocol for Device-to-Device Communication in SoftNet Towards 5G. In Proceedings of the Information and Communication Technology for Intelligent Systems (ICTIS 2017), Ahmedabad, India, 25–26 March 2017; Volume 84. [CrossRef]
34. Liu, Y.; Wang, Y.; Sun, R.; Miao, Z. Distributed resource allocation for D2D-Assisted small cell networks with heterogeneous spectrum. IEEE Access 2019, 7, 83900–83914. [CrossRef]
35. Dijkstra, E.W. A Note on Two Problems in Connexion with Graphs. In Edsger Wybe Dijkstra: His Life, Work, and Legacy; ACM: New York, NY, USA, 1959; Volume 271, pp. 269–271.
36. Giambene, G.; Kota, S.; Pillai, P. Integration of Satellite and 5G Networks Satellite-5G Integration: A Network Perspective. IEEE Netw. 2019, 32, 25–31. [CrossRef]
37. Xu, H.; Yu, Z.; Li, X.Y.; Huang, L. Joint Route Selection and Update Scheduling for Low-Latency Update in SDNs. IEEE/ACM Trans. Netw. 2017, 25, 3073–3087. [CrossRef]
38. Chai, Y.; Shi, W.; Shi, T.; Yang, X. An efficient cooperative hybrid routing protocol for hybrid wireless mesh networks. Wirel. Netw. 2016, 23, 1387–1399. [CrossRef]
39. Triviño, A.; Ariza, A.; Casilari, E.; Cano, J.C. Cooperative layer-2 based routing approach for hybrid wireless mesh networks. China Commun. 2013, 10, 88–99. [CrossRef]
40. Caria, M.; Jukan, A.; Hoffmann, M. SDN Partitioning: A Centralized Control Plane for Distributed Routing Protocols. IEEE Trans. Netw. Serv. Manag. 2016, 13, 381–393. [CrossRef]
41. El-Sawy, H.; Dahrouj, H.; Al-Naffouri, T.Y.; Alouini, M.S. Virtualized cognitive network architecture for 5G cellular networks. IEEE Commun. Mag. 2015, 53, 78–85. [CrossRef]
42. Walikar, G.; Biradar, R.; Geetha, D. Topology based adaptive hybrid multicast routing in mobile ad hoc networks. In Proceedings of the 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Pune, India, 19–21 December 2016; pp. 19–21. [CrossRef]
43. Sun, L.; Zheng, W.; Rawat, N.; Sawant, V.; Koutsonikolas, D. Performance comparison of routing protocols for cognitive radio networks. IEEE Trans. Mob. Comput. 2014, 14, 1272–1286. [CrossRef]
44. Huang, X.; Lu, D.; Li, P.; Fang, Y. Coolest Path: Spectrum Mobility Aware Routing Metrics in Cognitive Ad Hoc Networks. In Proceedings of the 2011 31st International Conference on Distributed Computing Systems, Minneapolis, MN, USA, 20–24 June 2011. [CrossRef]

45. Ihara, Y.; Kremo, H.; Altintas, O.; Tanaka, H.; Ohtake, M.; Fujii, T.; Yoshimura, C.; Ando, K.; Tsukamoto, K.; Tsuru, M.; et al. Distributed autonomous multi-hop vehicle-to-vehicle communications over TV white space. In Proceedings of the 2013 IEEE 10th Consumer Communications and Networking Conference, CCNC 2013, Las Vegas, NV, USA, 11–14 January 2013; pp. 336–344. [CrossRef]

46. Syed, A.; Yau, K.L.; Qadir, J.; Mohamad, H.; Ramli, N.; Keoh, S. Route selection for multi-hop cognitive radio networks using reinforcement learning: An experimental study. IEEE Access 2016, 4, 6304–6324. [CrossRef]

47. McAuley, A.; Sinkar, K.; Kant, L.; Graff, C.; Patel, M. Tuning of reinforcement learning parameters applied to OLSR using a cognitive network design tool. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Paris, France, 1–4 April 2012; pp. 2786–2791. [CrossRef]

48. Briand, A.; Albert, B.B.; Gurjao, E.C. Complete software defined RFID system using GNU radio. In Proceedings of the 2012 IEEE International Conference on RFID-Technologies and Applications (RFID-TA), Nice, France, 5–7 November 2012; pp. 287–291. [CrossRef]

49. Wei, X.; Liu, H.; Geng, Z.; Zheng, K.; Xu, R.; Liu, Y.; Chen, P. Software Defined Radio Implementation of a Non-Orthogonal Multiple Access System towards 5G. IEEE Access 2016, 4, 9604–9613. [CrossRef]

50. Zhao, Y.; Pradhan, J.; Huang, J.; Luo, Y.; Pu, L. Joint energy-and-bandwidth spectrum sensing with GNU radio and USRP. ACM SIGAPP Appl. Comput. Rev. 2015, 14, 40–49. [CrossRef]

51. Elsayed, M.H.M. Distributed interference management using Q-Learning in Cognitive Femtocell networks: New USRP-based Implementation. In Proceedings of the 2015 7th International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, 27–29 July 2015.

52. Xiong, X.; Xiang, W.; Zheng, K.; Shen, H.; Wei, X. An open source SDR-based NOMA system for 5G networks. IEEE Wirel. Commun. 2015, 22, 24–32. [CrossRef]

53. Chamran, M.K.; Yau, K.I.A.; Noor, R. An Experimental Study on D2D Route Selection Mechanism in 5G Scenarios. Electronics 2021, 10, 387. [CrossRef]

54. Yarnagula, H.K.; Deka, S.K.; Sarma, N. Distributed TDMA based MAC protocol for data dissemination in ad-hoc Cognitive Radio networks. In Proceedings of the 2013 IEEE International Conference on Advanced Networks and Telecommunications Systems, ANTS 2013, Kattankulathur, India, 15–18 December 2013; pp. 1–6. [CrossRef]

55. Elwhishi, A.; Ho, P.H.; Naik, K.; Shihada, B. ARBR: Adaptive reinforcement-based routing for DTN. In Proceedings of the 2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob'2010, Niagara Falls, ON, Canada, 11–13 October 2010; pp. 376–385. [CrossRef]

56. Habak, K.; Abdelatif, M.; Hagrass, H.; Rizc, K. A Location-Aided Routing Protocol for Cognitive Radio Networks. In Proceedings of the 2013 International Conference on Computing, Networking and Communications (ICNC), San Diego, CA, USA, 28–31 January 2013; pp. 729–733.

57. Bouhaba, M.; Hafid, A.; Belbekkouche, A.; Gendreau, M. Reinforcement learning based routing in wireless mesh networks. Wirel. Netw. 2013, 19, 2079–2091. [CrossRef]