Computer Communications 176 (2021) 13–22

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Reinforcement learning-enabled Intelligent Device-to-Device (I-D2D) communication in Narrowband Internet of Things (NB-IoT)☆

Ali Nauman [a], Muhammad Ali Jamshed [b], Rashid Ali [c], Korhan Cengiz [d], Zulqarnain [a], Sung Won Kim [a,*]

[a] Department of Information and Communication Engineering, Yeungnam University, Republic of Korea
[b] Institute of Communication Systems (ICS), Home of 5G Innovation Centre (5GIC), University of Surrey, UK
[c] School of Intelligent Mechatronics Engineering, Sejong University, Seoul, Republic of Korea
[d] Department of Electrical - Electronics Engineering, Trakya University, 22030, Edirne, Turkey

ARTICLE INFO

Keywords:
Reinforcement Learning (RL)
Intelligent communication
Device-to-Device (D2D) communication
Narrowband Internet of Things (NB-IoT)
5th Generation (5G) networks

ABSTRACT

The 5th Generation (5G) and Beyond 5G (B5G) are expected to be the enabling technologies for Internet-of-Everything (IoE). The quality-of-service (QoS) for IoE in the context of uplink data delivery of the content is of prime importance. The 3rd Generation Partnership Project (3GPP) standardizes the Narrowband Internet-of-Things (NB-IoT) in 5G, which is Low Power Wide Area (LPWA) technology to enhance the coverage and to optimize the power consumption for the IoT devices. Repetitions of control and data signals between NB-IoT User Equipment (UE) and the evolved NodeB/Base Station (eNB/BS), is one of the most prominent characteristics in NB-IoT. These repetitions ensure high reliability in the context of data delivery of time-sensitive applications, e.g., healthcare applications. However, these repetitions degrade the performance of the resource-constrained IoT network in terms of energy consumption. Device-to-Device (D2D) communication standardized in Long Term Evolution-Advanced (LTE-A) offers a key solution for NB-IoT UE to transmit in two hops route instead of direct uplink, which augments the efficiency of the system. In an effort to improve the data packet delivery, this study investigates D2D communication for NB-IoT delay-sensitive applications, such as healthcare-IoT services. This study formulates the selection of D2D communication relay as Multi-Armed Bandit (MAB) problem and incorporates Upper Confidence Bound (UCB) based Reinforcement Learning (RL) to solve MAB problem. The proposed Intelligent-D2D (I-D2D) communication methodology selects the optimum relay with a maximum Packet Delivery Ratio (PDR) with minimum End-to-End Delay (EED), which ultimately augments energy efficiency.

1. Introduction

Internet-of-Things (IoT) refers to the assortment of smart objects with the capability of self-reconfiguration, uniquely addressable, interoperable, and flexible that can sense, acquire, and process data [1]. IoT applications are driving the advances for future wireless communication and the smart applications/services, such as smart city. Currently, the number of connected devices to the internet is 23 billion devices [2], and the sum of connected devices is expected to extend up to 75 billion by 2025 [3]. This exponential increase of IoT devices upraises the demand for Machine Type Communication (MTC) [4]. MTC is categorized into three categories, that is long-range (range ≥100 m), medium-range (10 m<range<100 m), and short-range (range ≤10 m). IoT devices have limited resources in the perspective of energy, processing, and memory. Long-range MTC over such devices with limited resources necessitates the standardization of Low Power Wide Area (LPWA) technology [5].

Narrowband-Internet of Things (NB-IoT) introduced by 3rd Generation Partnership Project (3GPP) in Release 13 of Long Term Evolution (LTE) [6]. NB-IoT is designed to improve spectrum efficiency, in-depth and extended coverage [7]. NB-IoT is one of the licensed LPWA technologies which provides a transmission range of more than 3 km in urban and 15 km in open area with strong penetration capabilities for MTC [7]. The main aim of NB-IoT is to support the IoT devices with an intended life expectancy of 10 years [8]. One of the attractive

☆ This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00313) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation) and in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1A09082266).

* Corresponding author.
E-mail addresses: anauman@ynu.ac.kr (A. Nauman), mohammadalijamshed@gmail.com (M.A. Jamshed), rashidali@sejong.ac.kr (R. Ali), korhancengiz@trakya.edu.tr (K. Cengiz), zulqarnain@ynu.ac.kr (Zulqarnain), swon@yu.ac.kr (S.W. Kim).

https://doi.org/10.1016/j.comcom.2021.05.007
Received 27 February 2021; Received in revised form 21 April 2021; Accepted 7 May 2021
Available online 17 May 2021

characteristics of NB-IoT is that it can be directly integrated into LTE or Global System for Mobile Communications (GSM) networks to share spectrum and reuse the same hardware to scale down the deployment cost. NB-IoT requires one Physical Resource Block (PRB) of the LTE spectrum, that is 180 kHz from system bandwidth for downlink and uplink communication. The limited one PRB bandwidth increases the constraints of NB-IoT system resources [9]. Adequate utilization of these deficient resources intensifies the challenges for NB-IoT deployment.

IoT devices are usually focused on the uplink transmission to upload the acquired data to the gateway/sink node or the cloud server. Efficient uplink data transmission is one of the leading research areas in IoT. Repetition of control and data signal is one of the main approaches considered in NB-IoT to augment coverage and reliability. However, network efficiency decreases with the increase of number of transmission repetitions [10]. Besides, NB-IoT devices deployed in deep indoor locations, experience an additional penetration loss of up to 20 dB. To provide extended coverage, repetitions and extended Transmission Time Interval (TTI) augments time delay and energy consumption [11]. Thus, it is critical to have a delay and energy-efficient uplink transmission mechanism for wide deployment applications.

Motivation: 5G and Beyond 5G (B5G) are expected to be the enabling technologies for Internet-of-Everything (IoE), enabled devices. In the context of data delivery for these devices, quality-of-service (QoS) is of prime importance. Currently, LTE-Advanced (LTE-A) supports Device-to-Device (D2D) communication [12]. The D2D communication provides an efficient mechanism to assist the NB-IoT User Equipment (UE) in transmitting the acquired critical data to the Evolved NodeB/Base-Station (eNB/BS). The D2D communication exploit cellular devices within the proximity that can operate act as relay nodes. The integration of NB-IoT within the LTE-A standard is one of the appealing characteristics. Thus the amalgamation of NB-IoT and D2D communication is anticipated to augment the performance of the wireless networks. In this research work, a Reinforcement Learning (RL) based Intelligent-D2D (I-D2D) communication approach for a NB-IoT UE has been presented, which exploits D2D communication as uplink routing approach for NB-IoT UE to upload the critical data to eNB/BS. The work presented in this paper is the extension of our Deterministic D2D (2D2D) approach proposed in [13]. The 2D2D approach selects the relay node in a deterministic manner at BS, which involves multiple control signals towards BS that augments system overheads, energy consumption, and delay. I-D2D reduces the additional overheads by working as a distributed system at the NB-IoT UE, which models the relay selection problem as a Multi-Arm Bandit (MAB) system, as proposed in [14]. The MAB is well-known for working in a real-time channel scenario [15]. One of the best and effective way to solve the MAB problem is Upper-Confidence-Bound (UCB) algorithm that converges quickly in stationary, independent, and identical distributed traffic [16]. The proposed RL model augments the performance of the NB-IoT system by minimizing overheads, optimizing End-to-End Delay (EED), and increasing the Packet Delivery Ratio (PDR) by selecting the best relay node for D2D communication. In summary, this paper aims to make the following contributions:

- This article presents the description and background of NB-IoT, D2D communication, and MAB learning schemes.
- We adapt the D2D communication as a routing extension to upload the urgent NB-IoT UE data to eNB in order to maximize PDR and minimize EED.
- This article presents a simple yet effective two-step RL-enabled intelligent D2D communication model, which considers relay selection as a MAB problem and solves it using the UCB algorithm. The proposed I-D2D algorithm effectively selects the relay with minimum overheads and uploads the UEs data to BS/eNB with minimum delay and maximum PDR.

Table 1
Symbols used throughout the paper.

Symbol	Meaning
α	Exploration and exploitation coefficient
β	SINR threshold
γ	Threshold of transmission power of NB-IoT UE
σ	Path loss exponent
μ_k	Stationary mean reward of the kth relay
μ^*	Expected value of the reward of the optimal relay
π	Reinforcement learning policy
a_t	Action of agent at time t
$A_k(t)$	Upper confidence bias
$B_k(t)$	UCB index
BS/eNB	Base Station/evolved NodeB
CUE	Cellular user equipment
E	Total number of data packets received
EED	End-to-end delay
g	Channel gain
K	State-space
$k_n(t)$	State k_n (CUE relay) at time t where $n \in \{1, 2, \ldots, N\}$
k_n	State (CUE relay)
L	Total number of data packets transmitted
M	Total number of UEs
N	Total number of relays in PRS $n \in \{1, 2, \ldots, N\}$
$N_k(t)$	Total number of times CUE k is selected as relay from instant 0 to $(t-1)$
P_t	Transmission power
P_r	Received power
$PRS(K)$	Potential relay set
PDR	Packet delivery ratio
$r_k(t)$	Reward of relay k at time t
R_t^{π}	The regret of the policy π at time t
R_{req}	Required transmission power of NB-IoT UE
$SINR$	Signal-to-Interference-Noise-Ratio
t	Discrete time step $t = \{1, 2, 3, \ldots\}$
τ	Processing time to select relay
$\bar{X}_k(t)$	Sample mean of the relay k

- A comparison of I-D2D with deterministic model [13] and opportunistic model [17] is presented. It shows that I-D2D communication has a higher PDR with minimum EED in terms of processing time.

Paper Organization: The organization of the paper is as follows: Section 2 presents a brief comparison with related work, the description related to the overview of NB-IoT wireless technology, and the D2D mechanism. Section 3 shows how RL and its tools can enable NB-IoT UE to select relay nodes intelligently. Section 4 describes the system model of D2D communication in NB-IoT and the proposed I-D2D model. Performance evaluation and simulation results are presented in Section 5. Section 6 concludes this article. Table 1 presents the list of symbols used throughout the paper.

2. Related work and overview

2.1. Related work

In recent works, vast research has been conducted to enhance the performance of the 5G communication systems in the context of reliability and latency, which corresponds to PDR and EED of the system by using D2D communication. An unorthodox work has been presented in [18]. An effectual D2D communication-based approach to improve the compliant content uploading is proposed. A trust-based approach for the NB-IoT network is developed, which takes the past reputation of a device is into consideration before establishing a D2D communication link for security reasons. The proposed scheme aims to filter out skeptical users and avoid unsuccessful transmissions. *Petrov et al.* [19] proposed NB-IoT enabled opportunistic crowdsensing-based application. The D2D communication link is exploited with the help of the vehicles, which acts as a relaying system. The proposed scheme is an opportunistic model in which energy abundant vehicles

are equipped with advanced communication modules to assist the battery constrained IoT devices. *Osama et al.* in [20] studied the effect of mutual interference in cellular UE (CUE) and D2D NB-IoT UE. As CUE and NB-IoT UE transmits the data in the same resource block to increase the spectral efficiency of the system. *Elsawy et al.* in [21] put forward the analytical model for D2D communication. In this work, an adaptive mode selection model, along with truncated channel inversion power control for uplink cellular communication has been presented. *Liu et al.* in [22] considered the potential gain of D2D communication for enhancing network coverage and spectral efficiency of the communication system. *Yang et al.* [23] designed an approach to harvest energy for relay devices for D2D communication from BS. *Sreedevi et al.* in [24] proposed RL-based latency controlled D2D connectivity for indoor network.

D2D communication model based on dynamic programming has been presented in [17], which uses a UE from the relaying group for D2D communication to transmit the data of NB-IoT device to BS. The research paper formulates optimization problems in order to maximize the reliability in terms of the expected delivery ratio and to optimize EED. In this work [17], an NB-IoT UE is allowed to establish a D2D link with available CUE, which acts as a relay in duty cycles in an opportunistic manner. When the transmission is unsuccessful, the UE re-transmits on the next scheduled relay node from the relaying group. The data packet is dropped after a fixed waiting time interval. In a pragmatical network, such an opportunistic scheme results in a significant increase in the system's overhead and degrades the system's efficiency with an increase in energy consumption and incur a huge delay. Such overheads and delay are not tolerable for time-sensitive applications, for instance, surveillance and monitoring in smart industry, critical readings of a heart patient in smart hospitals, and traffic management system in a smart city. *Nauman et al.* in [13] proposed a deterministic approach instead of an opportunistic model for relay selection. The proposed approach selects the relay for D2D communication at BS, which eliminates the additional delay present in the opportunistic model to wait for Cellular UE (CUE) to operate as a relay. However, to select the relay in a deterministic manner, NB-IoT UE has to transmit a pilot signal every time when it has data to upload to the eNB/BS. The CUEs that qualify and are available for D2D communication transmit the pilot signal to eNB/BS to select the best relay. The eNB/BS selects the best candidate for relaying the data after ranking the relays in decreasing order on the basis of channel gain and residual power. This incorporates additional processing and delay, thereby increasing energy consumption. Therefore, this necessitates an intelligent mechanism based on ML to select the relay dynamically for NB-IoT UE with minimum processing time, ultra-reliability, and low latency.

2.2. NB-IoT overview

2.2.1. NB-IoT deployment modes

NB-IoT uses one PRB of 180 kHz in the frequency domain for downlink and uplink transmission, which splits into 12 sub-carriers of 15 kHz each. NB-IoT can be deployed directly in the LTE or GSM spectrum in three different modes of operations to scale down the deployment costs. When NB-IoT UE is first powered on, it searches for carrier channel, thus the deployment mode should be clear to the NB-IoT UE [25]. Following are the deployment modes of NB-IoT [5].

In-band mode. One of the PRBs of the LTE spectrum is allocated for NB-IoT deployment. The total power of eNB is shared between LTE and NB-IoT.

Stand-alone mode. NB-IoT can also be deployed within 200 kHz of the GSM spectrum. NB-IoT can exploit the power of BS, which significantly improves the coverage of the system.

Guard-band mode. The guard-band of the LTE spectrum is utilized for NB-IoT deployment.

2.2.2. Downlink and uplink transmission

The NB-IoT follows similar numerology and frame structure as of LTE for compatibility with LTE. Each frame of 10 ms is composed of 10 sub-frames. The sub-frame is of 1 ms duration. Each sub-frame is equidivided into two slots of 0.5 ms length, which constitutes seven Orthogonal Frequency Division Multiplexed (OFDM) symbols and a normal Cyclic Prefix (CP). Uplink supports both single-tone and multi-toned transmissions. The single tone occupies 3.75 kHz or 15 kHz bandwidth. The 3.75 kHz numerology uses 2 ms slots, and the 15 kHz numerology is identical to LTE. Multi-tone transmission is based on Single Carrier Frequency Division Multiple Access (SC-FDMA) with the same 15 kHz sub-carrier spacing. Uplink uses Binary Phase Shift Keying (BPSK) or Quadrature Phase Shift Keying (QPSK), while downlink uses only QPSK [26]. The standardized data rate of NB-IoT is 160–250 Kbps for downlink and 160–200 Kbps for uplink transmissions [7]. For coverage enhancement, NB-IoT uses 128 re-transmissions for uplink and 2048 re-transmissions for downlink [5].

2.2.3. Device-to-device communication

D2D refers to a direct communication link with nearby devices without considering the intervention of the cellular networks. D2D communication is also defined as Proximity-based Service (ProSe) [27]. D2D was included in the LTE release 12 in 2012 [28]. The 3GPP standardizing community has approved the proposal of integrating D2D communication into LTE-Advanced (LTE-A). D2D communication has been classified into out-band D2D and in-band D2D communication. In-band, also referred to as LTE Direct, uses a licensed spectrum while out-band D2D exploits an unlicensed spectrum of other wireless enabling technologies that supports D2D communication such as IEEE 802.11 (WiFi) or IEEE 802.15 (Bluetooth) [29]. The D2D UE can access the licensed spectrum in shared mode (also refer as non-orthogonal/underlay mode) or dedicated mode (also known as orthogonal/overlay mode) [29]. The use of D2D communication leads to multiple advantages such as high packet delivery rate, minimum delay, better spectrum re-usability, and low energy consumption. Fig. 1 depicts the D2D communication scenario in a NB-IoT systems.

3. Reinforcement learning enabled relay selection

Instead of selecting relay node to upload NB-IoT UE data to eNB/BS in a deterministic or opportunistic manner [13,17], we propose in this article a dynamic relay selection approach to learn about the relay, which is more likely to be available and provide the best PDR. The learning process to select an optimum relay can be modeled as a Multi-Arm Bandit (MAB) system, as presented in [30]. The quality of the relay changes based on the location of the relay and the channel condition (Signal-to-Interference-Noise power Ratio SINR). Therefore, the selection of the optimum relay potentially leads towards the minimum delay and a reliable PDR, less costly in terms of system overheads and energy consumption.

3.1. Multi-arm bandit framework

The RL is the type of ML in which the learner (agent) has no prior knowledge of which action to perform in order to maximize the numerical reward (to move in the direction of the main objective). However, the agent has to discover which action to perform that yield maximum reward by hit and trial methodology. The motivation to use RL is also the same, as the full dynamics of the network environment is not known. If the dynamics of the network are known prior, heuristic algorithms such as dynamic programming are used to find the optimal solution. However, in real-time network scenarios dynamics are not known. Therefore, RL algorithms are used to find the optimal solution in real-time networks. The RL has three main elements: agent, environment, and reward. The MAB problem is the form of RL techniques in which a agent (player) repeatedly decides to

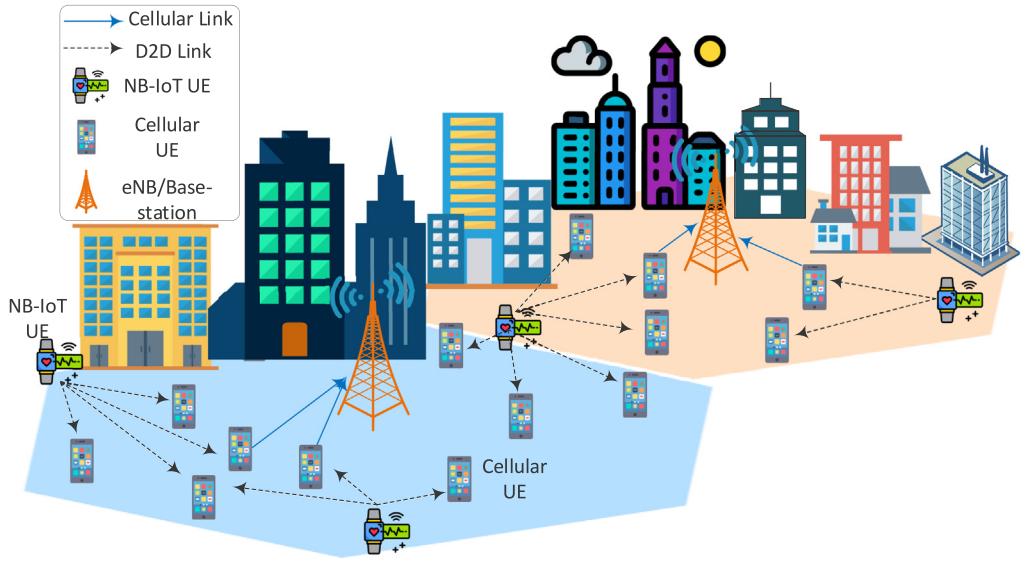

Fig. 1. D2D communication in NB-IoT system.

choose a state k_n (machine) from K number of states (machines), that is $k_n \in \{k_1, k_2, \ldots, k_N\}$ at discrete time $t = \{0, 1, 2, \ldots, \}$ based on their corresponding reward. Where, N is the total number of states, such that $n \in \{1, 2, \ldots, N\}$. The agent (player) is interested in choosing the state (machine), which maximizes the reward. The associated rewards with the states (machines) are independent and identically distributed (i.i.d) and accompany an unknown and fixed distribution law d_k. The reward distributions $\{d_1, d_2, \ldots, d_K\}$ vary from state to state, and the player has no prior knowledge about the distribution.

State-space and action: In this article, the player (agent) is the NB-IoT UE, and the state-space contain K states (machines) of the environment, which are the Cellular UEs (CUEs) used as a relay node to upload NB-IoT data. The action a_t is defined as the selection of relay CUE by the player (NB-IoT UE) with highest PDR, which maximizes the reward. Fig. 2 shows the environment and mechanism of the proposed I-D2D model with its elements.

Reward: Let $r_k(t)$ be the reward of successful data transmission for a relay node k at instant t. In this paper, two values for reward have been considered, i.e., 1 or 0. The reward is equal to 1 if the selected relay node successfully uploads the data to the eNB/BS, and acknowledgment is received and with good PDR. Note that the channel quality of the relay node is initially checked when cellular UE receives the pilot signal from NB-IoT UE with the request of D2D communication. The proposed algorithm is explained in Section 4.

Incentive for CUE Relay: The incentive for CUEs acting as the relay is defined as proposed in the Smart Media Pricing (SMP) framework [31]. In reality, the CUEs are selfish and reluctant to share their energy and communication resources. In order to deal with selfishness, SMP relay framework proposes that the relay device will price its energy, computation and communication resource used on the relay transmission, and the source which is eNB/BS in downlink scenario or NB-IoT device in the uplink scenario will pay the price for the incentive to the relay. The incentives can be defined as free service time or virtual points to be used later. Providing more incentives from the source to the relay motivates the relay device to increase its resources offered to use on the relay transmission. In this article, the incentive is provided by eNB/BS to CUEs acting as a relay, but the price will be paid by NB-IoT UE for these incentives.

Exploration and exploitation trade-off: At each time step, the agent has a choice to either exploit or explore the action. The *exploitation* refers to choose an action with the prior knowledge of the action values whose estimated value indicates the highest mean reward. The *exploration* refers to choose an action with no prior knowledge, which is to choose an action randomly from a set of actions to search for a better reward. The exploitation maximizes the immediate mean reward on the one step. However, exploration may yield better-accumulated reward in the long run. The uncertainty in exploration is that it is unknown which action produces a better reward. It is better to explore non-greedy actions if there are many time steps ahead to exploit them later on. However, it is not possible to select an action using exploitation and exploration at a single time step. This dilemma refers to as exploration and exploitation trade-off [16]. The UCB algorithm automatically balanced the exploration and exploitation as explained in the next subsection.

Regret: It refers to the loss experienced by the difference between the expected reward associated with the sub-optimal cellular relay node selected by the NB-IoT UE and the ideal reward associated with the optimal relay. As the NB-IoT UE does not have prior knowledge about the distribution of reward, it cannot avoid the loss when selecting a cellular relay UE.

Let π denote the learning policy for the best relay selection, and let μ_k=E$[d_k]$ be the stationary mean reward of the kth relay, where E[.] denotes the expectation function. The regret of the policy π is defined as

$$R_t^{\pi} = t.\mu^* - \sum_{l=0}^{t-1} r_l, \tag{1}$$

where, μ^* is the expected value of the reward of the optimal cellular relay.

Based on (1), the expected cumulative mean regret is

$$E[R_t^{\mu}] = \sum_{k=1}^{K} (\mu^* - \mu_k) E[N_k(t)], \tag{2}$$

where, $N_k(t)$ is the total number of times cellular relay k has been selected from instant 0 to instant $t-1$.

The MAB problem can be solved by many RL algorithms. Among them, UCB is the most efficient way to solve the MAB problem. In the following subsection, the use of UCB is briefly defined for dynamic relay selection.

3.2. Upper confidence bound algorithm

The policy in this article is based on UCB algorithm to aid the NB-IoT UE in the selection of a cellular relay node to upload its data to eNB/BS. The UCB algorithm requires few resources for processing and

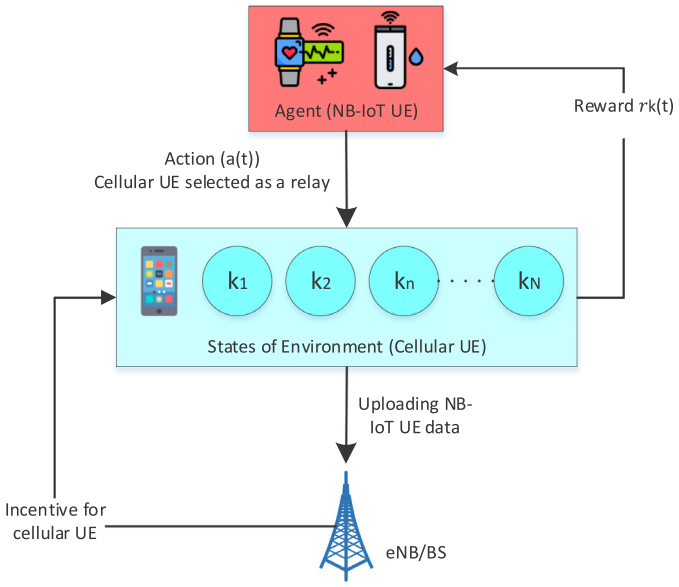

Fig. 2. MAB based I-D2D with its elements.

storage and guarantees the optimal performance asymptotically [16]. The UCB index $B_k(t)$, is calculated for each relay k_n and at each instant t. The UCB index reflects an estimation of the expected reward of a relay k_n. The UCB index is defined as

$$B_k(t) = \bar{X}_k(t) + A_k(t), \tag{3}$$

$$\bar{X}_k(t) = \frac{1}{N_k(t)} \sum_{t=0}^{t-1} r_k(t) \mathbf{1}_{(\alpha_{t=k})}, \tag{4}$$

$$A_k(t) = \sqrt{\frac{\alpha \ln(t)}{N_k(t)}}, \tag{5}$$

where, $\bar{X}_k$ denotes the sample mean of the relay k reward and A_k is upper confidence bias. $\mathbf{1}$ is the indicator function, and α in (5) is an exploration coefficient for relay availability and good PDR. If α gets small, NB-IoT UE will exploit the already chosen cellular relays, and if α gets larger, the UCB algorithm will explore more cellular relays for better availability and PDR. As the number of times a specific selected state increases, the upper confidence bias decreases. Therefore, the UCB moves towards the state with the highest upper confidence bias to explore. The upper confidence bias refers to the uncertainty of the state that is not explored. However, the UCB index is the summation of sample mean reward and upper confidence bias. The UCB ensures that the state with the highest upper confidence bias also has the highest accumulated reward in the past, which makes the exploration more beneficial. The model at each time step continuously explores and acts greedily simultaneously to select the state with the highest accumulated reward and highest upper confidence bias. The value of function indication $\mathbf{1}_{\alpha_{(t=k)}} = 1$ when the cellular relay k has been chosen at instant t. The output at each time step is the state with a maximum UCB index.

$$a_t = \arg\max_k (B_k(t)), \tag{6}$$

where, a_t is the selected CUE relay using the policy π at the tth transmission resulting from the UCB algorithm with the highest UCB index.

4. System model and proposed scheme

In this section, the network model and assumptions made in this article are detailed.

4.1. Network model

In this paper, a two-tier network model is considered with a single eNB/BS in a network cell, and NB-IoT UEs, operating in In-band mode. The eNB/BS is assumed to be located in the center of the cell surrounded by cellular devices, as shown in Fig. 1. The network model depicts the scenario where the NB-IoT UE comprises IoT devices such as smart ambulances, smart watches, or implants, which have critical data to transmit. Timely and reliable data transmission is of utmost importance. This paper is focused on uplink transmissions, where an NB-IoT UE has to transmit the acquired data to the eNB/BS. The uploading of the data is limited to two hops, i.e., NB-IoT UE can directly upload the data to the eNB/BS or first transmitting the data to a nearby cellular UE (CUE) that can upload the data to the eNB/BS. The former case is using single-hop, and the latter case is using two-hop communication via D2D link and cellular link, as shown in Fig. 1. The uplink cellular spectrum is divided into C sub-channels, and each CUE communicates with eNB/BS using one of the sub-channels. Moreover, CUE and NB-IoT D2D link use the same uplink resources of the eNB/BS. Any CUE available in the cell can act as a relay in D2D communication.

4.1.1. Channel model

The general power-law propagation is considered to characterize the path loss effect of all cellular and D2D link transmissions. The channel noise is assumed to be Additive White Gaussian Noise (AWGN). A Rayleigh fading channel is considered between transmitter and receiver pair. The gain for Rayleigh fading channel is normally distributed. The power P_r received on the receiver from the transmitter with power P_t at a distance d can be calculated as:

$$P_r = P_t.g^{-\sigma}, \tag{7}$$

where g is the channel gain and $\sigma > 2$ is the path loss exponent. It is assumed that all the fading coefficients are independent, and the D2D and cellular link share the same path-loss exponent σ.

4.1.2. Signal-to-interference-noise-ratio

Since all the communication links, i.e., cellular and D2D, are allowed to use the same uplink resources of the eNB/BS, the cross-tier interference among all the communication links is avoided assuming that a unique sub-carrier is allocated to each UE. The cross-tier interference could also be ignored, assuming that the distance between the D2D pair is small and transmitting power required for the D2D link is lower than the CUE direct link. As per the 3GPP specifications, the Signal-to-Interference-Noise-Ratio (SINR) for LTE-A link is not defined [32]. SINR is calculated by the UE internally and reported by UE to eNB during uplink transmission to determine the link quality of each UE. The SINR is calculated from Reference Signal Received Quality (RSRQ), which is determined by Reference Signal Received Power (RSRP) [32]

$$RSRQ = N_{PRB} \cdot \frac{RSRP}{RSSI}, \tag{8}$$

where, $RSSI$ is the Received Signal Strength Indicator, and N_{PRB} is the number of Physical Resource Blocks. The SINR is then measured by

$$SINR = \frac{12.RSRQ}{x}, x = \frac{RE}{RB}, \tag{9}$$

where RE indicates Resource Element and RB indicates Resource Block.

4.1.3. Definitions

Packet delivery ratio:. The PDR is a key metric for evaluating the performance in terms of reliability. The PDR is defined as the ratio of the number of packets that originated at the transmitter to the number of packets received at the receiver end [33]. The following expression defines the PDR as

$$PDR = \frac{\sum_{e=0}^{E}(E_e)}{\sum_{l=1}^{L}(L_l)}, \tag{10}$$

where L is the total number of packets transmitted and E is the total number of packets received. As stated in [17], average Bit Error Rate (BER) based on SINR for binary signal detection in AWGN is

$$BER = Q(\sqrt{SINR}). \tag{11}$$

Here $Q(.)$ denotes the standard Gaussian error function $Q(f) = (1/\sqrt{2\pi}) \int_f^\infty e^{\frac{-1}{i^2}} dt$. It is assumed that the error in bits occurs independently. For J number of bits in a packet, the PDR is determined by considering the probabilities of receiving all the bits correctly at eNB/BS. The PDR of ith $J - bit$ packet can be expressed as

$$PDR(i) = \prod_{j=1}^{J} (1 - BER). \tag{12}$$

The estimated PDR of a transmission link can be calculated by averaging (12) over L packets

$$PDR = \frac{1}{L} \sum_{l=1}^{L} \prod_{j=1}^{J} (1 - BER). \tag{13}$$

End-to-end delivery ratio:. The End to End Delivery ratio (EDR) is a performance metric in multi-hop transmissions. It is the summation of PDR from NB-IoT UE to CUE relay and from relay CUE to BS/eNB. The EDR is calculated using the following expression:

$$EDR = \sum_{n=1}^{N} (PDR^{(k_n)}_{UE \rightarrow CUE_{k_n} \rightarrow BS/eNB}), \tag{14}$$

End-to-end delay:. The EED for an NB-IoT UE is the EED for the packet sent by NB-IoT UE and received by eNB/BS over two hops (using CUE as a relay for D2D communication). The EED in this paper is optimized by minimizing the processing time to select the CUE relay.

Processing time:. The processing time τ is the time for NB-IoT UE to select the relay to upload the data to the eNB/BS. Minimum τ reduces EED.

Potential relay set:. The potential relay set (PRS) for NB-IoT UE is the number of N CUEs which are within the range of NB-IoT UE for D2D communication and assist NB-IoT UE to forward the packet to the eNB. In [17], these relays are sorted according to the availability of time-slot reserved for CUE to be used as a relay in the proposed opportunistic model. A CUE can act as a relay in duty cycles, and NB-IoT UE has to wait for the scheduled time-slot to transmit the data. This approach promotes dropping the packet after the defined threshold time if NB-IoT UE is not able to find the opportunity to transmit the data.

However, in 2D2D [13], this uncertainty is eliminated by working in a centralized manner at eNB/BS. The NB-IoT UE broadcasts a pilot signal to all CUE in its range to confirm the availability and eligibility to act as a relay. Every CUE evaluates itself by comparing its channel gain and residual energy with predefined threshold parameters, to find its eligibility and updates the eNB/BS. The channel gain is based on SINR of communication link with eNB/BS for uplink transmission and residual energy, which indicates either the CUE has enough energy resources to utilize it for relaying the data in uplink transmission. The PRS is sorted in decreasing order at the eNB/BS that selects the highest-ranked relay. However, I-D2D works in a distributed manner at NB-IoT UE. Instead of broadcasting the pilot signal every time NB-IoT UE has data to upload, the NB-IoT UE broadcasts the pilot signal only the first time when it has data to upload. The PRS is formed formally, i.e., $K = \{k_1, k_2, k_n, \ldots, k_N\}$, where N is the total number of relays, i.e., $N = \{1, 2, \ldots, N\}$. After the learning period, which is explained in the next section, the NB-IoT UE selects the best relay determined by UCB. This approach significantly reduces the control overheads, which in turn augments PDR and EED.

4.2. Proposed model

The MAB learning consists of an agent that learns (that is an NB-IoT UE), an environment (that is the number of CUEs available for relay), a policy (that is to select relay which maximizes the PDR), a reward (that is 0 or 1), and function $X_k(t)$ (a accumulated reward in term of sample mean). The learning and behavior of an NB-IoT UE at a given time t depends on the policy π it follows. A policy π refers to a set of specified rules to determine prospective actions that are mapped with the perceived states of the environment. The reward refers to the main objective of the NB-IoT UE, which determines the quantitative value of the situation at each time step. In the RL based MAB problem, an agent's objective is to select the state (CUE) that maximizes the accumulated mean reward over the long run. Where reward $r_k(t)$ is the quantitative value for any single immediate action of a specific state. The value $X_k(t)$ denotes the accumulated mean reward achieved till the current time state. Perhaps, it is likely that a CUE achieves a low immediate reward but still possesses a high mean reward.

For the uplink transmission, NB-IoT UE decides its association as follows. When NB-IoT UE turns on, it establishes a link with its nearest eNB/BS. Each NB-IoT UE sends Channel Quality Indicator (CQI) to the eNB/BS to check channel quality [32]. The CQI indicates SINR, as explained in the previous section. If SINR is over a stated threshold value, i.e., β, it forms a direct communication link with the eNB/BS. Otherwise, the UE operates in D2D mode. In this paper, when NB-IoT UE is under deep fading, it requires to upload critical data to eNB/BS and will always transmit its data in a two-hop manner with the help of CUE. To start sending the data via D2D link, this paper proposes an I-D2D mechanism based on the RL model. The proposed I-D2D mechanism is elucidated in algorithm 1. The I-D2D works in two stages. The following two steps are followed by NB-IoT UE to select the best relay for D2D communication:

- *Step 1:* The NB-IoT UE has to formulate PRS for which a pilot signal is broadcasted within its range. The CUE analyzes the pilot signal by comparing and computing SINR of the communication link with eNB/BS using Eq. (9) with SINR threshold (β), and the transmission power of NB-IoT UE R_{req} threshold γ. If the CUE is eligible for D2D communication, it sends a response signal to update NB-IoT UE about the availability for D2D communication. Else it will withdraw from the selection process. The NB-IoT UE formulates a PRS with N number of CUE relays and considers PRS as the state-space K of the environment.
- *Step 2:* NB-IoT UE models the relay selection process as a reinforcement learning MAB problem and solves it by exploiting the UCB algorithm as follows:
  - The UCB works by initializing state-space K and exploration coefficient α.
  - The inputs of the UCB algorithm are $\{a_{t=0}, r_{t=0}, a_{t=1}, r_{t=1}, \ldots, a_{t-1}, r_{t-1}\}$. In UCB algorithm, player (NB-IoT UE) plays each k_n machine (CUEs) in the PRS one by one in an iterative manner to determine the upper confidence bias using (5) and accumulates the mean reward $r_t(k)$ against each action $a_t(k)$ using (4). The reward $r_t(k)$ is +1 if CUE successfully transmits NB-IoT UE data with good PDR. Otherwise the reward $r_t(k)$ is 0. Every time a CUE k_n is selected and given +1 reward, it decreases the upper confidence bias because $N_k(t)$ is in the denominator of (5) and decreases the uncertainty against an action to select a state. On the other hands, every time an action is taken to select the state other than k_n, the $\ln(t)$ in numerator increases and augments uncertainty.

Algorithm 1: Intelligent-D2D (I-D2D).

Step 1:
Input: (β, γ, R_{req}, $SINR$)
for ($M = \{1, \ldots, m\}$) **do**
  if $SINR \leq \beta$ **then**
    if $R_{req} \leq \gamma$ **then**
      Push $m(t)$ UE in a new array PRS
    else
      withdraw from selection process
    end if
  else
    withdraw from selection process
  end if
end for
Output: (PRS $K = \{k_1, k_2, ..., k_N\}$)
Step 2:
Input: $\{a_0, r_0, a_1, r_1, ..., a_{t-1}, r_{t-1}\}$
Initialize parameters: PRS $K = \{k_1, k_2, ..., k_N\}$, α
Output: action (a_t)
for $L = \{1, 2, ..., l\}$ **do**
  if $t \leq K$ **then**
    select the relay from PRS one by one in an iterative manner
    if uplink transmission successful **then**
      reward $r_k(t) = 1$
    else
      reward $r_k(t) = 0$
    end if
  else
    select the relay from PRS with maximum UCB index
    if uplink transmission successful **then**
      reward $r_k(t) = 1$
    else
      reward $r_k(t) = 0$
    end if
    update reward $r_k(t)$ and $\bar{X}_k(t)$ using Eq. (4)
    update UCB confidence bias $A_k(t)$ using Eq. (5)
    update UCB index $B_k(t)$ using Eq. (3)
    update $a_t = \arg\max_k (B_k(t))$ using Eq. (6)
  end if
  return action a_t with maximum UCB index
end for

– The UCB algorithm checks if ($t \leq (K)$), then selects the next relay in the PRS to compute the upper confidence index using (3)–(5) and return a_t by Eq. (6). It means that I-D2D will select every CUE relay one by one and determine it's UCB index. When ($t > (K)$), the UCB selects the relay with maximum UCB index directly using (6) and update Eq. (3)–(5).

The proposed I-D2D algorithm works in a distributed manner at the edge of the network and significantly reduces the processing time τ in selecting the relay node for D2D communication with maximum PDR at NB-IoT UE.

5. Performance evaluation

This section outlines the performance evaluation of the proposed I-D2D scheme and provides a comparison with the state-of-art techniques available in the literature.

5.1. Simulation deployment scenario

The setup comprises of a single cell-based cellular network, with M users being deployed randomly, such that, each m user undergoes small scale fading. Each user is then segregated based on their received signal strength at the eNB/BS into NB-IoT (unable to transmit the signal to the eNB/BS) and relay users (able to transmit the signal to the eNB/BS). Initially, a PRS is formed, i.e., the N users have the capability of uploading the traffic of NB-IoT users. Precisely, the users in PRS K have good channel conditions with the eNB/BS as well as the NB-IoT node. It is assumed that the NB-IoT is working in in-band deployment mode within the LTE system spectrum. Lastly, this information is fed into the MAB-based RL algorithm, as explained in the previous section. The detailed simulation parameters are listed in Table 2.

Table 2
Simulation parameters.

Symbol	Value
No. of UEs M	50
Max. communication range	130 m
SINR threshold (β)	13 dB
Path-loss exponent (σ)	3.5
Noise power density	−174 dBm/Hz
Max. transmission power of NB-IoT UE (γ)	14 dB
Size of NB-IoT data packet (L)	32 bytes
Exploration coefficient (α)	1.5
Simulation iterations	100

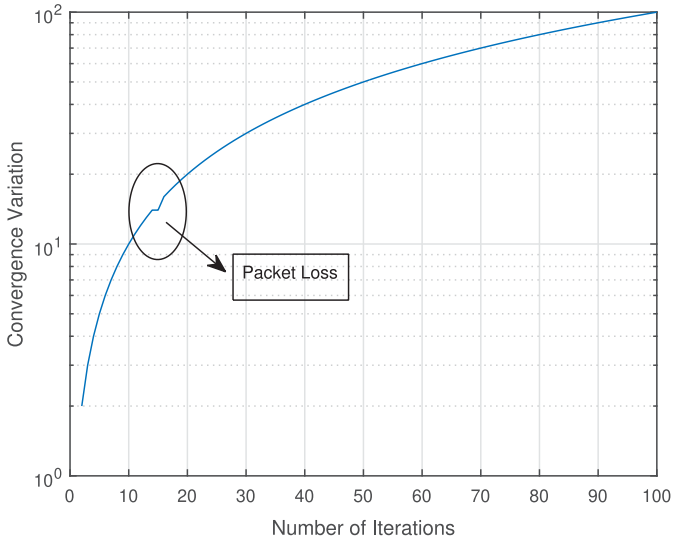

Fig. 3. Convergence of the proposed I-D2D algorithm over a fixed number of iterations.

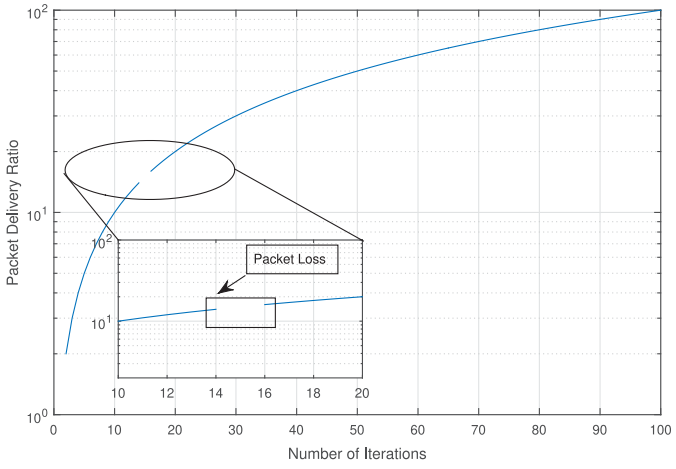

Fig. 4. PDR of the proposed I-D2D algorithm over a fixed number of iterations.

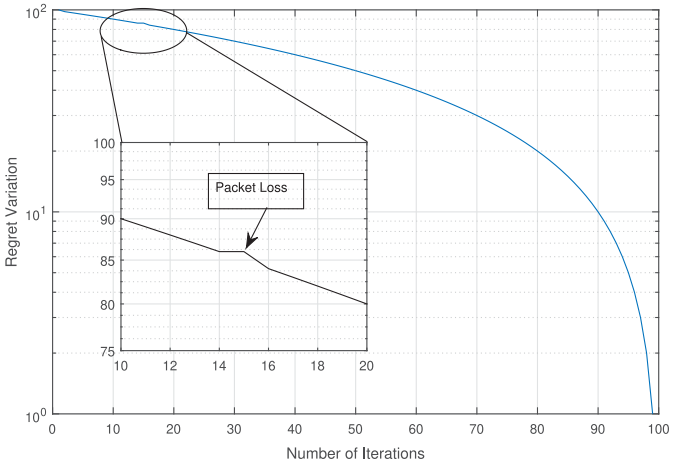

Fig. 5. Regret variation of the proposed I-D2D algorithm over a fixed number of iterations.

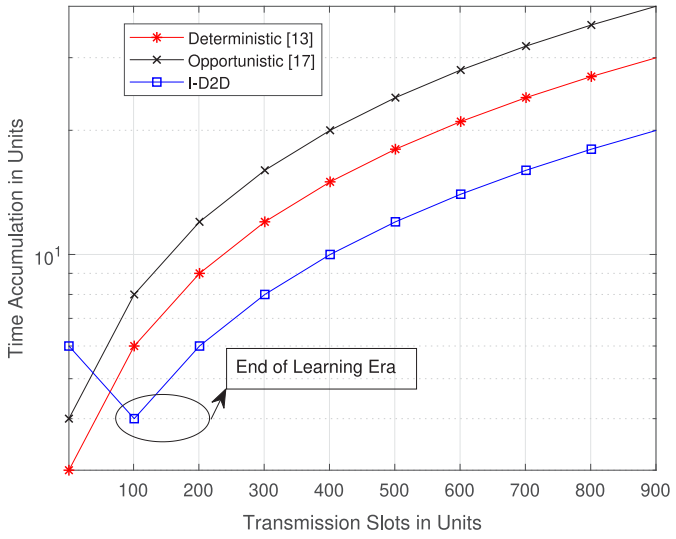

Fig. 6. A comparison of the transmission time accumulation of the proposed I-D2D algorithm with some state-of-art techniques over a varying number of time slots.

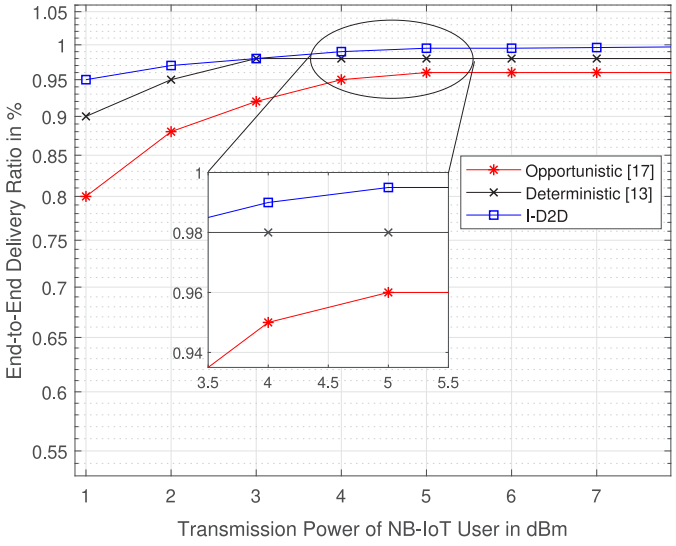

Fig. 7. A comparison of EDR ratio of the proposed I-D2D algorithm with some state-of-art techniques over a variation in transmit power of a NB-IoT user.

When the NB-IoT UE has to upload the data, it tries to connect to nearby relay devices for D2D communication that maximizes the reward, as explained in the previous section. The reward matrix is updated as follows:

- If the NB-IoT UE receives an acknowledgment for successful packet transmission from eNB/BS, it will update the reward for the CUE relay node with 1 and proceeds to the data transmission.
- Otherwise, the reward for selected CUE relay node is updated to 0.

5.2. Results and discussion

In Fig. 3, the convergence of the RL-based I-D2D algorithm has been shown. The significant findings from this result are that, as the number of iterations increases, the system converges towards optimality in selecting the best relay selection, which corresponds to the CUE relay with good PDR and minimum τ. This result was expected from the target and behavior of the policy. During the initial iterations, the algorithm begins exploring the CUEs from PRS with an objective to calculates the UCB index. The notch between iteration 10 and 20 indicates packet loss, i.e., CUE relay with reward 0 is selected during exploration, because the relay with the highest probability to be free does not necessarily yield the best PDR. Moreover, when CUE relay fails to allocate resources, the NB-IoT device experience packet loss.

In Fig. 4, the PDR of the RL-based I-D2D algorithm has been shown. It can be seen that over the increasing number of iterations, the PDR increases. It can be seen in Fig. 4, that between iteration number 10 and 20, there is a disconnection between the two points, which shows the loss of packets during the transmission time interval because CUE relay was not able to allocate the resources to NB-IoT device. The loss in packet occurs during the learning period of the I-D2D algorithm, which shows the wrong selection of CUE relay node over a fixed number of time slots during exploration.

Fig. 5 represents the cumulative regret variation of the RL-based I-D2D algorithm. The regret is because of a random selection of CUE relay as there is no prior knowledge about the PDR associated with the CUE relay, and the availability of CUE relay for D2D communication is also unknown. It can be seen that over the increasing number of iterations, the cumulative regret of the algorithm starts to decrease and finally approaches zero as the system converges.

The principal motivation behind the I-D2D algorithm is to cutoff the processing time τ required to select the relay for each transmission duration. As explained in Section 4.1.3, the 2D2D algorithm in [13] works in a centralized manner at eNB/BS, which requires an additional control signal to be transmitted by NB-IoT UE and CUE to eNB/BS. Whereas [17] is an opportunistic manner that waits for the CUE to act as a relay in reserved time slots in duty cycles. When an NB-IoT UE cannot find a relay node after a specified time, it drops the data packet, which promotes the packet loss rate. However, I-D2D reduces τ by working in a distributed manner and selecting the CUE relay using RL based MAB problem, which learns the availability and PDR associated with CUE. I-D2D automatically selects the best CUE relay as the system converges. Fig. 6 shows the comparison of the processing time τ accumulation over multiple transmission time slot of the I-D2D scheme with the state-of-art schemes, i.e., 2D2D and the opportunistic model proposed in [13] and [17], respectively. It can be seen that at the 100th iteration, the learning period of I-D2D ends, i.e., UCB system converges, after which it selects the CUE relay in significantly less time with good PDR as compared to other models. From the fact that the I-D2D requires significantly less τ, it can be concluded that I-D2D reduces EED by reducing τ.

The comparison of the end-to-end delivery ratio (EDR) versus the varying transmission power of NB-IoT users has been shown in Fig. 7.

In comparison to similar schemes in [13] and [17], the I-D2D scheme shows a considerable improvement in the EDR. This improvement can be clarified using the I-D2D algorithm by incorporation of the UCB algorithm in it, which ensures the reliability of the packet being transmitted successfully by selecting the CUE relay. Ideally, after the learning period, the I-D2D will ensure a 100% EDR, but it is impractical. In order to justify 100% PDR, additional losses (mobility of relay user, shadowing, etc.) have been considered. The learning period of I-D2D is not considered while making the comparison shown in Fig. 7.

6. Potential application areas of the proposed method

Smart cities which includes smart grid, smart industries and smart healthcare have gained significant recognition in the past decade. IoT is one of the key enabler for future smart cities. The next generation communication requires devices to be adaptive and intelligent to provide ultra reliable and low latency communication. For reliable and time sensitive application such as healthcare, tele-surgery, drone applications, and autonomous industry, data delivery is of prime importance. Intelligent D2D communication can assist in achieving high PDR, EDR with minimum time. The proposed mechanism enhances PDR and reduces time delay by reducing number of retransmissions with optimal cellular relay selection, which ultimately increases the energy efficiency. The proposed model can be incorporated in other scenarios where reliability and time-delay is of prime importance. The future steps of our research work target to investigate the multi-player scenario where number of NB-IoT UEs need to select a CUE relay for uploading the data. Furthermore, cooperative ML techniques also known as federated learning need to be investigated for large IoT deployments.

7. Conclusion and future work

The emergence of massive MTC requires ultra-reliability in the context of data delivery with extended in-depth coverage. NB-IoT fulfills these requirements by repetitions of control and data signals. Reducing energy utilization is one of the prominent aspects of NB-IoT. However, the fundamental solution of increased repetitions of control and data signals consumes more energy. In order to improve data delivery, a novel D2D communication link is used as a routing approach for NB-IoT, which offers the NB-IoT UE a two-hop route to reduce repetitions. An Intelligent D2D (I-D2D) relay selection model based on RL is designed, which selects the cellular UE relay with the highest probability to be available with the maximum PDR and the minimum EED. Simulation results depict that the proposed I-D2D algorithm outperforms the available state-of-the-art techniques in the literature.

CRediT authorship contribution statement

Ali Nauman: Conception and design of study, Acquisition of data, Writing - original draft. **Muhammad Ali Jamshed:** Conception and design of study, Acquisition of data. **Rashid Ali:** Analysis and/or interpretation of data, Writing - original draft. **Korhan Cengiz:** Writing - review & editing. **Zulqarnain:** Analysis and/or interpretation of data, Writing - original draft. **Sung Won Kim:** Conception and design of study, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

All authors approved the version of the manuscript to be published.

References

[1] P.P. Ray, A survey on Internet of Things architectures, J. King Saud Univ. Comput. Inform. Sci. 30 (3) (2018) 291–319, http://dx.doi.org/10.1016/j.jksuci.2016.10.003.
[2] Statista, Number of Mobile Phone Users Worldwide from 2015 to 2020 (In Billions), 2015.
[3] A. Nordrum, et al., Popular Internet of Things forecast of 50 billion devices by 2020 is outdated, IEEE Spectrum 18 (2016).
[4] S. Sharma, X. Wang, Towards massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions, IEEE Commun. Surv. Tutor. (2019) 1, http://dx.doi.org/10.1109/COMST.2019.2916177.
[5] C.B. Mwakwata, H. Malik, M. Mahtab Alam, Y. Le Moullec, S. Parand, S. Mumtaz, Narrowband Internet of Things (NB-IoT): From physical (PHY) and media access control (MAC) layers perspectives, Sensors 19 (11) (2019) http://dx.doi.org/10.3390/s19112613.
[6] A. Nauman, Y.A. Qadri, M. Amjad, Y.B. Zikria, M.K. Afzal, S.W. Kim, Multimedia Internet of Things: A comprehensive survey, IEEE Access 8 (2020) 8202–8250, http://dx.doi.org/10.1109/ACCESS.2020.2964280.
[7] J. Chen, K. Hu, Q. Wang, Y. Sun, Z. Shi, S. He, Narrowband Internet of Things: Implementations and applications, IEEE Internet Things J. 4 (6) (2017) 2309–2314, http://dx.doi.org/10.1109/JIOT.2017.2764475.
[8] S. Landström, J. Bergström, E. Westerberg, D. Hammarwall, NB-IoT: A sustainable technology for connecting billions of devices, Ericsson Technol. Rev. 4 (2016) 2–11.
[9] S. Popli, R.K. Jha, S. Jain, A survey on energy efficient Narrowband Internet of things (NB-IoT): Architecture, application and challenges, IEEE Access (2018) http://dx.doi.org/10.1109/ACCESS.2018.2881533.
[10] C. Yu, L. Yu, Y. Wu, Y. He, Q. Lu, Uplink scheduling and link adaptation for Narrowband Internet of Things systems, IEEE Access 5 (2017) 1724–1734, http://dx.doi.org/10.1109/ACCESS.2017.2664418.
[11] J. Lianghai, B. Han, M. Liu, H.D. Schotten, Applying device-to-device communication to enhance IoT services, IEEE Commun. Stand. Mag. 1 (2) (2017) 85–91, http://dx.doi.org/10.1109/MCOMSTD.2017.1700031.
[12] A. Algedir, H.H. Refai, Energy-efficient D2D communication under downlink HetNets, in: 2019 IEEE Wireless Communications and Networking Conference, WCNC, 2019, pp. 1–6, http://dx.doi.org/10.1109/WCNC.2019.8885967.
[13] A. Nauman, M.A. Jamshed, Y. Ahmad, R. Ali, Y.B. Zikria, S. Won Kim, 2019 15th International Wireless Communications Mobile Computing Conference, IWCMC, IEEE, 2019, pp. 2111–2115, http://dx.doi.org/10.1109/IWCMC.2019.8766786.
[14] M. Chafii, F. Bader, J. Palicot, Enhancing coverage in narrow band-IoT using machine learning, in: 2018 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2018, pp. 1–6, http://dx.doi.org/10.1109/WCNC.2018.8377263.
[15] L. Melián-Gutiérrez, N. Modi, C. Moy, I. Pérez-Álvarez, F. Bader, S. Zazo, Upper Confidence Bound learning approach for real HF measurements, in: 2015 IEEE International Conference on Communication Workshop, ICCW, 2015, pp. 381–386, http://dx.doi.org/10.1109/ICCW.2015.7247209.
[16] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, second ed., The MIT Press, 2018, http://incompleteideas.net/book/the-book-2nd.html.
[17] Y. Li, K. Chi, H. Chen, Z. Wang, Y. Zhu, Narrowband Internet of Things systems with opportunistic D2D communication, IEEE Internet Things J. 5 (3) (2018) 1474–1484, http://dx.doi.org/10.1109/JIOT.2017.2782323.
[18] L. Militano, A. Orsino, G. Araniti, M. Nitti, L. Atzori, A. Iera, Trusted D2D-based data uploading in in-band narrowband-IoT with social awareness, in: Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016 IEEE 27th Annual International Symposium on, IEEE, 2016, pp. 1–6, http://dx.doi.org/10.1109/PIMRC.2016.7794568.
[19] V. Petrov, A. Samuylov, V. Begishev, D. Moltchanov, S. Andreev, K. Samouylov, Y. Koucheryavy, Vehicle-based relay assistance for opportunistic crowdsensing over narrowband IoT (NB-IoT), IEEE Internet Things J. 5 (5) (2018) 3710–3723, http://dx.doi.org/10.1109/JIOT.2017.2670363.
[20] O. ElGarhy, L. Reggiani, Increasing efficiency of resource allocation for D2D communication in NB-IoT context, Proc. Comput. Sci. 130 (2018) 1084–1089, http://dx.doi.org/10.1016/j.procs.2018.04.160.
[21] H. ElSawy, E. Hossain, M. Alouini, Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks, IEEE Trans. Commun. 62 (11) (2014) 4147–4161, http://dx.doi.org/10.1109/TCOMM.2014.2363849.
[22] J. Liu, N. Kato, Device-to-device communication overlaying two-hop multi-channel uplink cellular networks, in: Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc '15, New York, NY, USA, 2015, pp. 307–316, http://dx.doi.org/10.1145/2746285.2746311.
[23] H. Yang, J. Lee, T.Q.S. Quek, Heterogeneous cellular network with energy harvesting-based D2D communication, IEEE Trans. Wirel. Commun. 15 (2) (2016) 1406–1419, http://dx.doi.org/10.1109/TWC.2015.2489651.
[24] A. Sreedevi, T. Rama Rao, Reinforcement learning algorithm for 5G indoor device-to-device communications, Trans. Emerg. Telecommun. Technol. 30 (9) (2019) e3670.

[25] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, A. Ghosh, NB-IoT System for M2m communication, in: Wireless Communications and Networking Conference (WCNC), 2016 IEEE, IEEE, 2016, pp. 1–5, http://dx.doi.org/10.1109/WCNC.2016.7564708.

[26] G. Tsoukaneri, M. Condoluci, T. Mahmoodi, M. Dohler, M.K. Marina, Group communications in narrowband-IoT: Architecture, procedures, and evaluation, IEEE Internet Things J. 5 (3) (2018) 1539–1549, http://dx.doi.org/10.1109/JIOT.2018.2807619.

[27] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, M.A. Javed, A survey of device-to-device communications: Research issues and challenges, IEEE Commun. Surv. Tutor. (2018) http://dx.doi.org/10.1109/COMST.2018.2828120.

[28] D. Feng, L. Lu, Y. Yuan-Wu, G.Y. Li, G. Feng, S. Li, Device-to-device communications underlying cellular networks, IEEE Trans. Commun. 61 (8) (2013) 3541–3551, http://dx.doi.org/10.1109/TCOMM.2013.071013.120787.

[29] P. Mach, Z. Becvar, T. Vanek, In-band device-to-device communication in OFDMA cellular networks: A survey and challenges, IEEE Commun. Surv. Tutor. 17 (4) (2015) 1885–1922, http://dx.doi.org/10.1109/COMST.2015.2447036.

[30] W. Jouini, D. Ernst, C. Moy, J. Palicot, Upper confidence bound based decision making strategies and dynamic spectrum access, in: 2010 IEEE International Conference on Communications, 2010, pp. 1–5, http://dx.doi.org/10.1109/ICC.2010.5502014.

[31] W. Wang, Q. Wang, Price the qoe, not the data: SMP-economic resource allocation in wireless multimedia Internet of Things, IEEE Commun. Mag. 56 (9) (2018) 74–79, http://dx.doi.org/10.1109/MCOM.2018.1701219.

[32] J. Parikh, A. Basu, Effect of mobility on SINR in long term evolution systems, ICTACT J. Commun. Technol. 7 (1) (2016) 1239–1244, http://dx.doi.org/10.21917/ijct.2016.0182.

[33] E. Fazeldehkordi, I.S. Amiri, O.A. Akanbi, Chapter 2 - literature review, in: E. Fazeldehkordi, I.S. Amiri, O.A. Akanbi (Eds.), A Study of Black Hole Attack Solutions, Syngress, 2016, pp. 7–57, http://dx.doi.org/10.1016/B978-0-12-805367-6.00002-8, http://www.sciencedirect.com/science/article/pii/B9780128053676000028.