computing@computingonline.net
www.computingonline.net

Print ISSN 1727-6209
On-line ISSN 2312-5381
International Journal of Computing

REINFORCEMENT LEARNING BASED ANTI-COLLISION ALGORITHM FOR RFID SYSTEMS

Murukesan Loganathan [1], Thennarasan Sabapathy [1], Mohamed Elobaid Elshaikh [1], Mohamed Nasrun Osman [1], Rosemizi Abd Rahim [1], Muzammil Jusoh [1], Mohd Ilman Jais [1], Badlishah Ahmad [2]

[1] Bioelectromagnetic (BioEM) Research Group, School of Computer and Communication Engineering, University Malaysia Perlis, 02600 Pauh Putra, Perlis, Malaysia
murukesan.loganathan23@gmail.com
[2] Faculty of Informatics and Computing, University Sultan Zainal Abidin, 22200 Besut, Terengganu, Malaysia

Paper history:
Received 2 September 2018
Received in revised form 16 November 2019
Accepted 15 March 2019
Available online 30 June 2019

Keywords:
collision avoidance;
dynamic frame-slotted Aloha;
EPC-C1G2;
reinforcement learning;
Q-learning.

Abstract: Efficient collision arbitration protocol facilitates fast tag identification in radio frequency identification (RFID) systems. EPCGlobal-Class1-Generation2 (EPC-C1G2) protocol is the current standard for collision arbitration in commercial RFID systems. However, the main drawback of this protocol is that it requires excessive message exchanges between tags and the reader for its operation. This wastes energy of the already resource-constrained RFID readers. Hence, in this work, reinforcement learning based anti-collision protocol (RL-DFSA) is proposed to address the energy efficient collision arbitration problem in the RFID system. The proposed algorithm continuously learns and adapts to the changes in the environment by devising an optimal policy. The proposed RL-DFSA was evaluated through extensive simulations and compared with the variants of EPC-C1G2 algorithms that are currently being used in the commercial readers. Based on the results, it is concluded that RL-DFSA performs equal or better than EPC-C1G2 protocol in delay, throughput and time system efficiency when simulated for sparse and dense environments while requiring one order of magnitude lesser control message exchanges between the reader and the tags.

Copyright © Research Institute for Intelligent Computer Systems, 2019.
All rights reserved.

1. INTRODUCTION

Radio frequency identification (RFID) technology had found widespread acceptance in security, logistics, retailing and inventory management [1]. RFID system is the most efficient and reliable way to identify an entity and collect data [2]. An RFID system consists of one or multiple readers with numerous tags that communicate using a shared communication channel. Among the three types of tags that are available in the market – passive, active and semi-active – passive tags are the least complex and cheapest. It uses the backscatter electromagnetic energy from the reader's signal to communicate the ID information. The communication protocol for the RFID system has to be simple since the tags are computationally challenged. Thus, the reader assumes full responsibility for managing or reducing collisions in the network. There are three types of collisions in an RFID system, namely, tag-to-tag, reader-to-tag and reader-to-reader [3]. The focus of this work is to propose a solution for reader-to-tag collisions using reinforcement learning technique.

Collision arbitration protocols for the RFID system can be divided into two categories. The existing protocols are either deterministic (tree-based) or probabilistic (Aloha-based). The tree-based protocols use binary tree search method where tags are continuously split into subsets until each set has only one tag. Tree-based anti-collision algorithms (ACA) are found to be efficient when the number of tags is small. However, long identification delays for a large number of tags and high protocol complexity are the drawbacks of these protocols [4]. On the other hand, Aloha-based ACA uses time slots and random transmission strategy to reduce the collision probabilities. They are known for minimal complexity and ease of implementation in the context of RFID applications. In fact, EPC-Global Class 1- Generation 2 (EPC-C1G2) standard uses a variant of Aloha for its operation. However, the theoretical maximum throughput of slotted aloha ACA is only 37% [5]. Also, these protocols cannot guarantee low identification delay in a dynamic environment such as warehouse [6].

In this paper, we present an efficient Aloha-based ACA which adapts its frame size dynamically using reinforcement learning mechanism. Framed slotted Aloha (FSA) is selected due to its simplicity and ability to handle a large number of tags or nodes when combined with capable algorithms [7]. The performance of the proposed algorithm (RL-DFSA) was evaluated using Monte-Carlo simulations and was compared with algorithms that are currently being used in commercial settings. RL-DFSA reduces collisions and improves throughput and delay significantly as compared to algorithms that are currently being employed in the commercial RFID readers. Besides, it is energy efficient since the control message overhead is an order of magnitude lower than that of the best performing algorithm in the commercial readers.

The remainder of this paper is organized as follows. Section 2.0 discusses the current RFID standard and related works. In Section 3.0, the complete methodology of the proposed RFID anti-collision protocol is presented in detail. Section 4.0 presents results and discussion of the proposed protocol in relation to selected protocols from the literature. Finally, the paper concludes with concluding remarks and future works in Section 5.0.

2. BACKGROUND INFORMATION

In FSA, a frame is divided into slots of the same length. At the beginning of each frame, interrogator or reader broadcasts the frame size to the tags. The tags then select a slot randomly and send the ID information to the reader in that slot. Due to this random slot selection policy, excessive collisions are bound to happen depending on the tag population if a non-optimal frame size is selected by the reader. The average throughput, U of FSA for N tag population and frame size L is,

$$\mathrm{U}=\mathrm{N}\left(1-\frac{1}{\mathrm{L}}\right)^{\mathrm{N}-1}, \tag{1}$$

and the normalized throughput, $\mathrm{U}_{\mathrm{norm}}$ is given by,

$$\mathrm{U}_{\mathrm{norm}}=\frac{\mathrm{N}}{\mathrm{L}}(1-\frac{1}{\mathrm{L}})^{\mathrm{N}-1}. \tag{2}$$

The normalized throughput is maximized when $L = N$. However, readers are not privy of the tag population and FSA has a fixed frame size. Due to these limitations, a variant of FSA called dynamic frame slotted Aloha (DFSA) which adapts frames dynamically based on the backlog tag estimation was proposed in the literature [8]. Depending on the accuracy of the backlog tag estimation method, the number of collisions in the proceeding frames varies for the better or worse. Besides, the throughput of the DFSA also drops when there is a large number of tags to read. Therefore, a variation of DFSA, called Q-algorithm was proposed to be used as the standard protocol in current generation RFID systems.

2.1 EPCGLOBAL CLASS 1 GENERATION 2 STANDARD

EPC-Global Class 1 Generation 2 (EPC-C1G2) is the RFID air interface protocol which enables interoperation of RFID devices across the globe with the help of its standardization [9]. It uses a variant of dynamic frame slotted Aloha (DFSA) known as Q-algorithm which operates per slot basis to arbitrate collisions and dynamically adapt the frame size.

Operation of Q-algorithm is shown in Fig. 1. Q algorithm operates using two parameters, namely, a floating-point parameter, Q_{fp} and c_q. The round Q_{fp} value is used to set the frame size, L and the c_q is used to increase or decrease the Q_{fp} value in the event of collision or empty slots, respectively. An interrogation process is initiated by the reader with the broadcast of a $Query$ command which contains the frame size. Upon receiving this command, tags generate a random number in the range between $0-2^{Q-1}$ and set their counter equal to the generated value. Then, the reader interrogates each slot of the frame one by one using the Query_repeat command. For each Query_repeat command, tags decrease their counter by one. Tag with counter equals to zero transmits its ID information to the reader. However, if there are more than one tag with counter equals zero for current slot, a collision would be detected by the reader. Consequently, the Q_{fp} is increased by some pre-determined $c_q$value. In the case of empty slot, Q_{fp} would be decreased by the same c_q value. The round Q_{fp} value would be updated continuously for each slot until a change is detected upon which the reader would exit the current frame and broadcasts a new frame size using the Query_adjust command. This process repeats until all tags are identified. The standard limits the round Q_{fp} value in the range between $0\ to\ 15$ for delay concerns. Besides, the reader has the autonomy to decide whether to exit the current frame or continue interrogating it even when the round Q_{fp} value had changed.

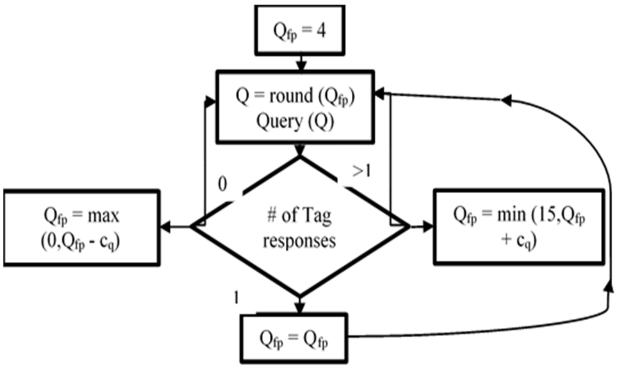


Figure 1 – Q-algorithm of EPC-C1G2 protocol where c_q values are between 0.1 to 0.5 [9]

One unique feature of the EPC-C1G2 algorithm is that it has different time durations for success, collision and empty slots as per the standard. Thus, the claim that the throughput of FSA maximized when $L = N$ is not applicable even though EPC-C1G2 is a variant of FSA. This has been verified analytically by [10] and the optimal frame size, L for EPC-C1G2 was calculated as,

$$L=1.46\times N\text{-}1, \tag{3}$$

where, $N - 1$ is the contending tag population.

However, Q-algorithm has several drawbacks as follow. The initial selection of the Q value affects its performance significantly. The reader has no means to know the population of tags in the network a priori to set the Q value appropriately. Besides, Q adjustment strategy using c_q produces excessive protocol overheads and also performs poorly in dense tag environment.

2.2 RELATED WORKS

In this section, we discuss some representative past studies on DFSA based anti-collision algorithms for RFID systems. The objective of the proposed algorithms can be either solving for optimal frame size or estimating the tag population. More often, the proposed algorithms try to achieve both these objectives as can be seen from the reviewed protocols in this section. We also explain some shortcomings of these algorithms.

Floerkemeier [11] and Bueno-Delgado et al. [12] proposed a solution for optimal frame size in the RFID system. Authors from both papers asserted that $L = N$ is the optimal frame size. However, we know from [10] that optimal frame size for RFID system is not same as in the traditional networks due to the different slot durations for the success, empty and collision slots of the RFID networks. On the other hand, Zhen et al. [13] proposed that the optimal frame size should be set as 1.4 times the tag population based on their own experimentation.

Eom et al. [14] proposed an anti-collision protocol which updates the frame size using the estimated backlog tag population. An estimation algorithm is used to calculate the number of collided tags (γ) in each collision slot. The author reported that the proposed protocol exhibits improved tag estimation accuracy while reducing the total number of slots required for an interrogation round. However, the author failed to distinguish between the three types of slots (success, empty and collision) when evaluating the total number of slots. Therefore, it is safe to assume that the reported comparison with the rest of the protocol is not valid.

Chen [15] introduced an anti-collision protocol which dynamically adjusts the frame length by examining only one slot per frame. This reduces the total number of examinations needed for setting the optimal frame size. The protocol updates the frame length using the estimated number of tag population. The author evaluated the protocol through simulations and reported that the normalized throughput of the protocol is higher as compared to the EPC-C1G2 protocol. However, the comparison is not valid since the author assumed all three types of time slots to have the same duration.

Even though there are numerous anti-collision algorithms available in the literature, in this paper, we only compared our proposed algorithm with the EPC-C1G2 protocol and its variants due to reasons stated as follows. Since we already know the optimal frame size from the literature, we can create an upper bound for performance (Ideal algorithm) as we had explained in Section 4. Therefore, there is no need to compare the proposed anti-collision algorithm with any other algorithms from the literature except the EPC-C1G2 algorithm and its variants. Besides, we can compare the results reported in this paper with other protocols by getting the percentage of improvement from the EPC-C1G2 protocol.

3. REINFORCEMENT LEARNING BASED DYNAMIC FRAME SLOTTED ALOHA (RL-DFSA)

In this section, the proposed RL-DFSA anti-collision algorithm is explained in detail. The primary motivation for pursuing RL based frame adaptation method is inspired by the work of Shaheen [16]. In this work, the author had used Markov decision process (MDP) which is the framework for most reinforcement learning algorithms [17] to analyze the slotted Aloha protocol. However, the author dropped the idea of solving the MDP for a large number of tags due to the need for an enormous number of computations. In turn, a heuristic-based method was adopted in the work. Accordingly, in this work, we approached the problem from a different point-of-view. Rather than calculating the transition probabilities, we used the Q-learning algorithm which updates the Q-value for each state based on its interaction with the environment. As a result, the computational complexity drops with the convergence time as a tradeoff. We used the Q-learning algorithm since it is known to be one of the most effective and popular algorithms to find an optimal policy in the absence of transition probability and reward function [18].

3.1 INTRODUCTION TO Q-LEARNING

Q-learning is a model-free reinforcement learning algorithm which learns by interacting with the environment and receiving Q-value for the state-action pair. The Q-value denotes the preference of taking an action over all other available actions when the system is at a certain state. Formally, for each state $s_t \in S$ and action $a_t \in A$ we define Q-value by,

$$Q(s_t, a_t) \longleftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a} Q(s_{t+1}, a') - Q(s_t, a_t)] , \quad (4)$$

where α is the learning rate, γ is the discount factor and r_{t+1} is the delayed reward. The $\alpha \in [0;1]$ value controls how quickly learning occurs. Besides, $\gamma \in [0;1]$ controls the willingness or deferment for delayed rewards. The objective of a reward function is to lead the learning agent towards the goal by properly rewarding or punishing the agent for the action taken at a certain state. A carefully defined reward function will lead the Q algorithm towards convergence in a relatively short amount of time depending on the application. Q-learning pseudo code for a single agent is presented in Algorithm 1.

Algorithm 1 Q-learning

```
Set t=0 and initialized Q-values $Q(s_t,a_t)$ for all $s_t \in S$
and $a_t \in A$.
while t<max_iteration do
    Observe the current state $s_t$.
    Select next action $a_t = \arg\max_{a' \in A} Q(s_t,a')$.
    Apply $a_t$, observe the next state $s_{t+1}$ and
    reward $r_t \triangleq r(s_t,a_t)$.
    Update Q-value
    $Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_t + \gamma \max_{a'} Q(s_{t+1},a') - Q(s_t,a_t)]$
    $t = t + 1$.
end while
```

The goal of the learning agent is to map each state to an action that maximizes its expected discounted reward over the time. However, a policy which chooses only the known maximal action without occasional exploration may succumb to locally optimal solutions. Therefore, there are numerous exploitation-exploration strategies available in the literature to tackle this problem. As for this work, we selected the well-known epsilon-greedy method [19] to balance between the exploitation and exploration. An agent following this learning strategy would occasionally choose actions which have lower Q-values with ε probability.

3.2 RL-DFSA

This subsection describes the methodology used to adapt FSA using Q-learning algorithm for the RFID systems. We are well aware of the computational restriction of the RFID tags and the complexity of the Q-learning algorithm. Thus, the proposed algorithm is created to run on the readers only. There are numerous high-end readers like GAORFID, RapidRadio etc. which have a powerful ARM processor and memory card supports [20], [21] that can run the proposed algorithm without any trouble. Besides, the algorithm also can be made to function in online or offline mode. In online mode, the reader would continuously update the Q-matrix until the end of the interrogation round. This mode also supports dynamic tag number population since the algorithm is actively learning. In the offline mode, the algorithm would be made to run on a reader for a certain tag population until convergence is achieved. After convergence, the Q-matrix can be transferred to low-end readers using memory cards so that they can function optimally for a certain tag population. This reduces the computational complexity since selecting an action with maximal Q-value from a matrix requires a smaller number of operations. The downside of the offline mode is that the low-end reader would produce errors when the tag population changes way beyond what it did during the training period in the high-end reader. We used offline mode for evaluating RL-DFSA due to the following reasons.

The proposed RL-DFSA has two phases, namely, learning (exploration + exploitation) and testing (exploitation only) due to the technical difficulty in running both the learning and testing, concurrently. For 1000 tags, RL-DFSA requires 20,000 iterations (~ 12 minutes) to converge to a near optimal policy as shown in Fig. 2. Learning an optimal policy is not possible due to the stochastic nature of our application. Fig. 3 shows the downward trend of cumulative reward as the exploration probability, ε is decayed over time settles around 20,000 iterations.

Therefore, learning and testing phases were conducted separately for time concerns. The slow convergence of the algorithm is due to the stochastic nature of our application and Q-learning itself is slow as rightly observed in [22].

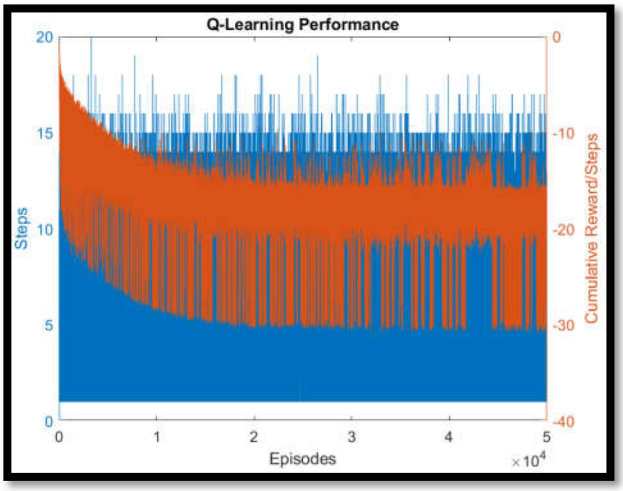


Figure 2 – Q-learning performance for 1000 tags

3.2.1 BASIC SETUP

In this work, the Q-learning algorithm was integrated into FSA to solve the reader-to-tag collision problem in the RFID networks. In FSA, a frame is comprised of multiple time slots of the same length. During each timeslot, the reader would interrogate the tags to get their ID information. Only the tag which had selected current time slot for transmission would reply in that particular timeslot. However, if the frame size is much smaller than the tag population, severe collisions would happen at the reader's side and depletes its energy. To make the matter worse, readers are not privy of the tag population to set the frame size to be optimal. Therefore, there is a need to estimate the tag population and determine the optimal frame size for the RFID networks. In this regard, Q-learning can help the reader to adapt its frame size dynamically using the feedback it got from the network. Besides, it can also solve the tag estimation problem through experimenting with the various tag estimation methods by having them as its possible actions as explained in the rest of this subsection.

The problem of determining optimal frame size for RFID network was solved analytically by [10] and it was found that the frame size should be set 1.46 times the tag population. Therefore, in this work, we focused on creating a policy for the reader so that it can adjust its frame size by alternating between the various tag estimates. The tag estimates were calculated based on a rational intuition which was based on the fact that the number of collided tags in a timeslot can be equal to or greater than two only. Thus, we defined the action space of the Q-learning algorithm as follow,

$$\text{Action 1}=1.46\times 2.0\times \text{number\_of\_collision}. \quad (5)$$

$$\text{Action 2}=1.46\times 2.2\times \text{number\_of\_collision}. \quad (6)$$

And so on until,

$$\text{Action 11}=1.46\times 4.0\times \text{number\_of\_collision}. \quad (7)$$

The number of actions was limited to eleven since increasing it further introduces additional time complexity which is exponential. As for the state of the learning agent, it was set to be equal to the number of collisions in the previous frame. A reward function was defined using reward shaping methodology to assist the learning agent to achieve its goal. A metric called collision ratio $\frac{\text{number\_of\_collision}}{\text{frame\_size}}$ was used to define the reward function as follows,

$$\text{reward}=\begin{cases} -1 \text{ ,ratio}>0 \text{ and}< \frac{1}{4} \\ -2,\ \text{ratio}\geq\frac{1}{4} \text{ and}< \frac{1}{2} \\ -4,\ \text{ratio}\geq\frac{1}{2} \text{ and}< \frac{3}{4} \\ -8, \qquad \text{ratio}\geq\frac{3}{4} \end{cases}. \quad (8)$$

It is clear from the reward function that the goal of the learning agent is to reduce the number of collisions to receive higher rewards.

3.2.2 LEARNING AND TESTING PHASES

The number of actions space must be small so that the Q-learning can converge in a reasonable amount of time. Therefore, an initial study was performed to identify the dominant actions based on the cumulative sum of their Q values. Using this criterion, three actions (1, 2 and 4) were identified as dominant and a new simulation was done using the identified actions. Through this new simulation, an optimal policy and Q-matrix for a tag population of 1000 were obtained. The number of tags was limited at 1000 since increasing it further would increase the simulation time exponentially. Besides, the policy obtained using 1000 number of tags can be used for tag population up to 2500 based on our own experiments. Beyond that error is produced since the state space exceeds the index of the Q-matrix. The parameters of RL-DFSA algorithm for the initial study are presented in Table 1.

Table 1. RL-DFSA parameters for the initial study

Parameter	Value
Initial state	2
Action	11
Learning rate, α	0.1
Discount rate, γ	0.9
Exploration, ε	0.3
Epsilon decay rate	0.99971
Maximum iteration	30,000
Number of tags	1000
Initial frame size	16

The initial state of the agent can be any arbitrary value except one since state one is the goal state. The timing parameters given in Table 3.1 were used for all our simulations. The pseudocode of RL-DFSA is presented in Algorithm 2.

Algorithm 2 RL-DFSA

```
Set t=0, max_iteration =30000 and initialize
  α,γ,ε and Q-values Q(s_t,a_t) for
  all $s_t \in S$ and $a_t \in A$.
while t<max_iteration do
$s_t = 2$ and select random action, $a_t$
Frame size = 16
Collision   = 0; Success = 0; Empty = 0;
while $s_t \neq 1$ do
    Broadcast frame size and get C, S, E.
    Next state $s_{t+1}$= C + 1;
    Get reward $r_t \triangleq r(s_t,a_t)$.
    Update Q-value
  $Q(s_t,a_t) \leftarrow Q(s_t,a_t)+\alpha\left[r_t+\gamma \max_{a'} Q(s_{t+1},a')-Q(s_t,a_t)\right]$
    if $\varepsilon >$ random_float_between 0 and 1
        Select random action, $a_t$
    else
    Select next action $a_t$= arg $\max_{a' \in A}$ Q($s_t$,a')
    end
    Frame size = $a_t$
    $s_t$= Next state $s_{t+1}$
end while
t=t+1  and  ε= ϵ×decay_rate
end while
```

During the testing phase, the learned Q-matrix was used to select an optimal action in each state. Monte-Carlo simulations with 5000 iterations were done for a various number of tags and the results are presented in Section 4.

Table 2. RL-DFSA parameters for the initial study [23]

Parameter	Duration		Parameter	Duration
Tari	6.5μs		PRT	57.594μs
RTcal	16.25μs		TFS	35.25μs
BLF	394kHz		T_{Query}	236.34μs
T_1	20.84μs		T_{ACK}	181.5μs
T_2	7.61μs		T_{QRep}	67.75μs
TRext	1		$T_{QueryAdj}$	96.68μs
M	1		T_S	1.1ms
T_{RN16}	126.9μs		T_C	223.11μs
T_{EPC}	695.43μs		T_E	113.97μs
T_3	25.381μs			

4. SIMULATION SETTINGS, RESULTS, AND DISCUSSION

In this section, simulation results and discussions for all five algorithms (EPC-Fixed, EPC-Q-Frame, EPC-Q-Slot, Ideal, and RL-DFSA) are presented. In EPC-Fixed, the fixed frame size of 16 (for sparse) and 128 (for dense) were used to simulate commercial readers with similar characteristics such as Symbol, ThingMagic Mercury 4, Samsys and Intermec [12]. Fixed frame size commercial readers are available in two variants which are the non-customizable and user customizable readers. The Q value for non-customizable tag readers is fixed at 4 while for the user-customizable tag readers, the user can select Q value from a range of 1 to 7 at the start of the interrogation round [12]. Therefore, in this simulation, the frame size of 16 and 128 were selected for simulating fixed frame size commercial readers in the sparse and dense environment, respectively. In the case of EPC-Q-Frame, initial frame size was set to be 16 as per the EPC-Gen2 standard requirement. However, there are no clear rules available in the EPC-Gen2 standard for fixing the c_q value. Nevertheless, c_q value of 0.3 was selected since it is found to perform most stable for sparse and dense networks [24]. EPC-Gen2 also allows the reader to decide whether to continue interrogating the current frame or abandon it when the round (Q_{fp}) value varies due to collision and empty slots. In this regard, EPC-Q-Frame (Algorithm 3) simulates the situation where reader decides to continue interrogating current frame even though the round (Q_{fp}) value had changed mid-frame. New Q value is only broadcasted at the beginning of next frame. As for the EPC-Q-Slot, reader abandons the current frame as soon as it detects a variation in round (Q_{fp}) value. In Ideal case, initial frame size was set 16 and it is assumed that the reader knows exactly the number of remaining tags in the system after the expiration of the first frame. Subsequent frame sizes were set to

1.46 ×remaining_tags which is the optimal frame size as explained in Section 2. The Ideal case is treated as the upper bound of performance that can be achieved by an optimal algorithm. Finally, in RL-DFSA, initial frame size was set to 16 and the subsequent frames were adjusted dynamically based on the optimal action (3 actions available) selected by the reader at each state.

```
Algorithm 3 EPC-Q-Frame Implementation
Require: $Q_{fp}$, $c_q$ and total_unidentified_tags
Ensure: total_unidentified_tags = 0
while (collision ≠ 0) do
    collision = 0;
    success  = 0;
    empty    = 0;
    Q = round ($Q_{fp}$)
    current_frame = $2^Q$
    slot_for_each_tag = random_slot (1, $2^Q$)
    for (each slot in current_frame)
        if (more than one tag selects current slot)
            collision = collision + 1
            $Q_{fp}$    = min (15, Qfp + $c_q$)
    elseif (only one tag selects current slot)
            success = success + 1
    elseif (current slot is unselected by tags)
            empty = empty + 1
            $Q_{fp}$ = max (0, Qfp - $c_q$)
    end
end
total_unidentified_tags = total_unidentified_tags - success
end while
```

Simulations were performed for a single reader with a various number of tags (10 - 1000) using Matlab 2017 software. Also, our simulation used the 394-kbps tag-to-reader link rate which obeys the regulation set by EPCGlobal [9]. The timing details presented in Table 3.1 were obtained from [23] since the same frequency as in the present work was used. Besides, the simulation scenario was divided into two – sparse (10 – 100 tags) and dense (100 – 1000 tags) environments – for an easier interpretation of the results. In order to get more reliable and accurate results, Monte-Carlo simulations with 5000 iterations were conducted for each algorithm and the following five performance metrics were recorded.

4.1 TIME SYSTEM EFFICIENCY (TSE) [10]

This metric gives the percentage of time successfully spend in identifying tags. It is calculated as follow:

$$\text{TSE}=\frac{\text{Success}\times T_s}{\text{Success}\times T_s+\text{Empty}\times T_e+\text{Collision}\times T_c}, \quad (9)$$

where, Success, Collision, and Empty denote the number of successful, collided and empty slots in the frame, respectively. T_s, T_e, and T_c are the duration of successful, empty and collision slots, respectively.

4.2 THROUGHPUT (tag per second)

This metric gives average tags per second that can be identified using the given algorithm. It is calculated as follow:

$$\frac{\text{Tag}}{\text{s}}=\frac{\text{Success}}{\text{Success}\times T_s+\text{Empty}\times T_e+\text{Collision}\times T_c+T_{\text{query}}}, \quad (10)$$

where, T_{query} denotes duration of the query command issued by the reader.

4.3 AVERAGE FRAME PER ROUND

This metric gives us an average number of frames issued by the reader for each interrogation round.

4.4 AVERAGE SLOTS PER ROUND

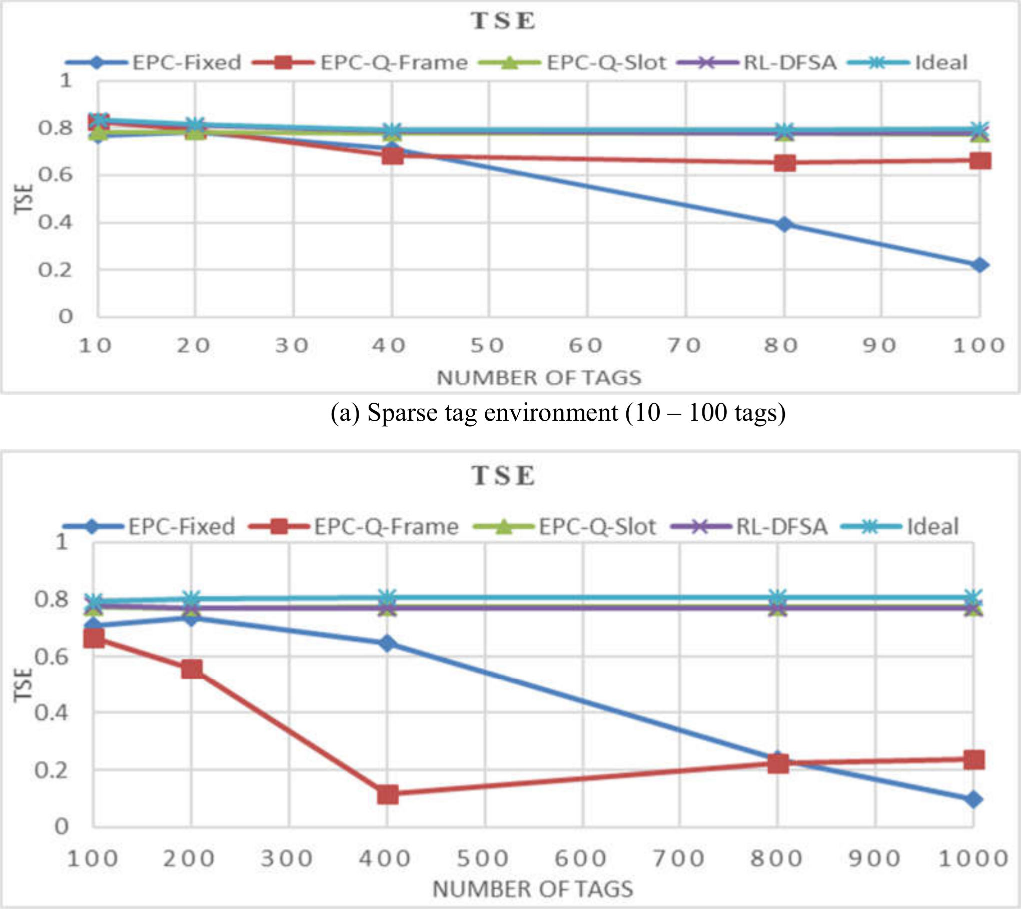
This metrics shows an average number of slots required for each interrogation round.

4.5 AVERAGE DELAY PER ROUND

This metric gives the average time taken by the reader to finish each interrogation round.

4.6 RESULTS AND DISCUSSION

The performance of RL-DFSA in terms of TSE was evaluated by comparing it with the other four algorithms for a various number of tags as shown in Fig. 4. As expected, EPC-Fixed performed the worst since the frame size was fixed for both the sparse and dense environments. As the number of tags increases, TSE drops abruptly due to the increase in collisions. One persistent trend in TSE and throughput results pertaining to EPC-Q-Frame is its performance deteriorate from 100 to 400 tags then increases gradually. This is because that Q-algorithm is slow to adapt to the rapid changes in the tag number population. Such behavior of Q-algorithm had also been reported by other researchers [25]. In contrast, EPC-Q-Slot performed far better since it abandons current frame as soon as it detects variation in the Q value. The performance of RL-DFSA and EPC-Q-Slot are almost identical to the Ideal case in sparse tag environment. However, unlike EPC-Q-Slot, RL-DFSA adapts to the changes in the frame with an order of magnitude fewer message exchanges as presented in Fig. 4.2. Its superior performance is due to the efficient learning method using feedback received in the form of reward/cost. In dense tag environment, there is a small gap in TSE for Ideal case and EPC-Q-Slot and RL-DFSA algorithms which denote there is still some room for improvements. Overall, RL-DFSA is 6.3% – 250 %, 0.4% - 18.6% and 0.4% - 5.7% better at TSE for sparse tag environment as compared to EPC-Fixed, EPC-Q-Frame and EPC-Q-Slot algorithms, respectively. Also, for dense tag environment, RL-DFSA performs 5.3% - 707.4% and 17% - 578.8% better as compared to EPC-Fixed and EPC-Q-Frame algorithms, respectively. The performance increment or decrement is insignificant (less than 1%) as compared to EPC-Q-Slot algorithm.

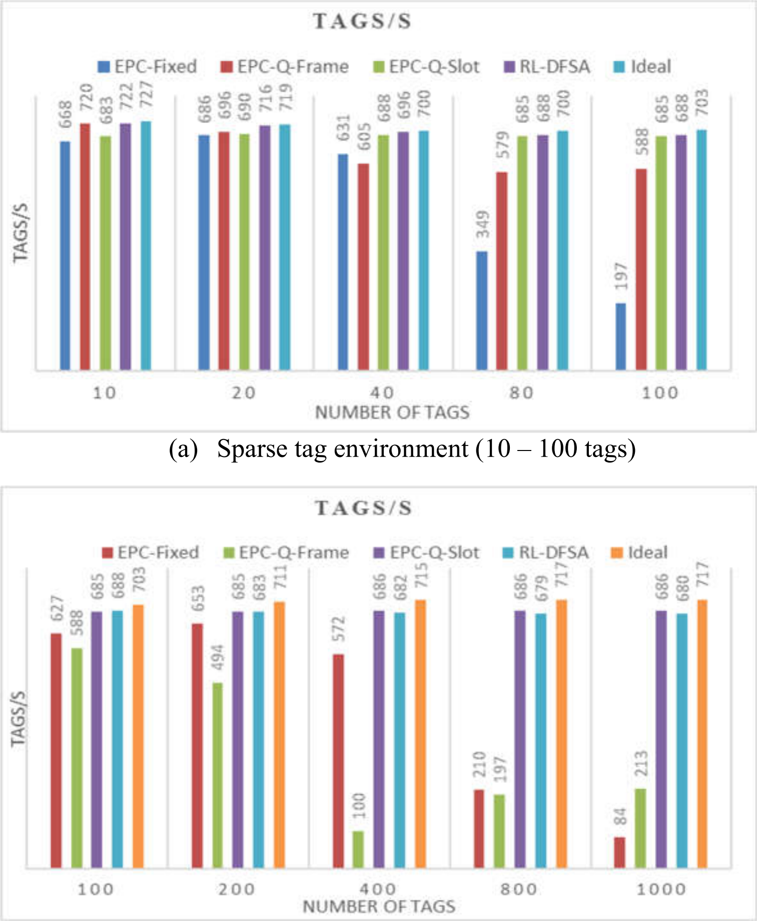


(a) Sparse tag environment (10 – 100 tags)

(b) Dense tag environment (100 – 1000 tags)

Figure 4 – TSE of the algorithms for various number of tags

The conventional normalized throughput of the algorithms fails to give an accurate picture on how it may translate to the real-life applications. Hence, for RFID systems, the throughput of an algorithm is given as the number of tags that can be identified per second as shown in Fig. 4.1. A similar trend as in the TSE can be observed here. RL-DFSA and EPC-Q-Slot perform almost identical on both sparse and dense environments except at a very low number of tags where RL-DFSA performed better. However, the performance of EPC-Fixed drops rapidly for the dense environment as the fixed 128 frame size is insufficient to accommodate large tag numbers. EPC-Q-Frame performs better than EPC-Fixed when the number of tags is small. In dense tag environment, its performance is unstable for the similar reasons mentioned during the discussion of TSE. Overall, RL-DFSA performs far better than EPC-Fixed and EPC-Q-Frame algorithms in both the sparse and dense tag environments with significant performance gap when the number of tags is large as can be seen in Fig. 4.1 (b).



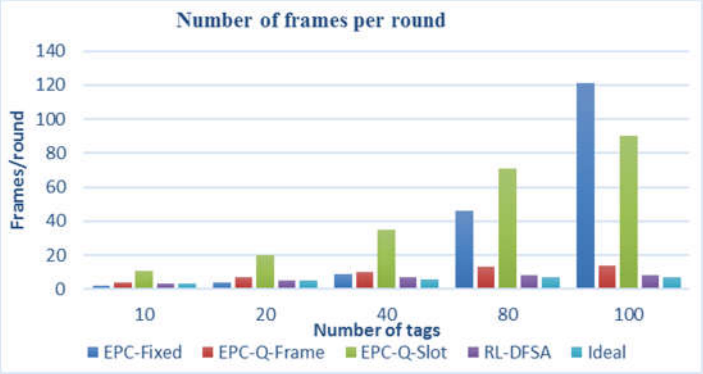
(a) Sparse tag environment (10 – 100 tags)

(b) Dense tag environment (100 – 1000 tags)

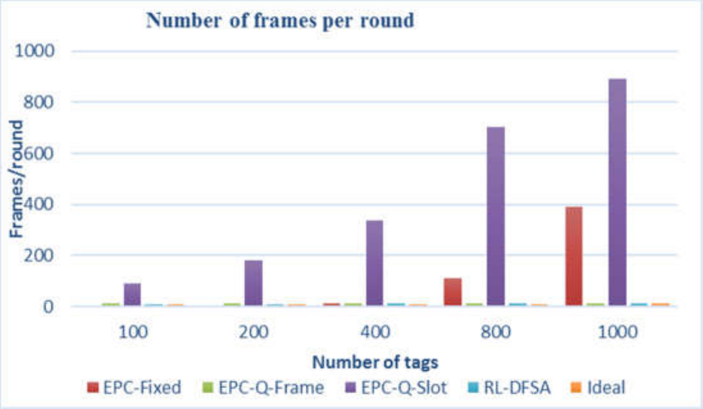
Figure 4.1 – Throughput (tags/s) performance of the algorithms for various number of tags

Energy efficiency is critical in RFID systems as the readers are battery operated [26]. Therefore, an efficient ACA should be able to reduce the collisions while guaranteeing fast tag identification time. In addition, the number of frames required per interrogation round also need to be kept at a minimum for energy and delay concerns. Fig. 4.2 shows an average number of frames per interrogation round issued by the reader for all five algorithms. EPC-Fixed and EPC-Q-Slot require an order of magnitude higher number of frames as compared to the other three algorithms. In the case of EPC-Fixed, the number of frames required increases with the number of tags due to lack of dynamic frame size adaptivity in the algorithm. It performs much better at dense tag environment since the frame size is 128 as compared to 16 in sparse tag environment. As for the EPC-Q-Slot, the decision to abandon a frame as soon as there is a difference in the Q value leads to excessive frame adjustment queries which is much more pronounced in dense tag environment. Even though EPC-Q-Slot performs similarly to RL-DFSA in TSE and throughput, this excessive overhead makes it ill-suited for RFID application since it is not energy efficient. Finally, RL-DFSA performed the best due to its efficient learning capability even in a stochastic environment.

In contrast to the number of frames, number of slots in each frame should be larger than the contending tags population so the tags can find a unique transmission slot. However, the number of slots cannot be increased indefinitely as a large number of empty slots would increase the system delay. The performance of EPC-Fixed and EPC-Q-Frame follows the same trend as in the earlier metrics with the performance of EPC-Q-Frame is better in the sparse environment. Since EPC-Q-Slot is utilizing a slot-by-slot frame updating mechanism, its performance should be the upper bound for this metric. In both sparse and dense tag environments, RL-DFSA performs almost identical to EPC-Q-Slot which shows its superior performance even though it utilizes a frame-by-frame updating mechanism. This is mainly because the actions of the Q-learning algorithm are optimal as they had been carefully selected from the initial study. Consequently, the performance of RL-DFSA is equal to EPC-Q-Slot algorithm as shown in Fig. 4.3.

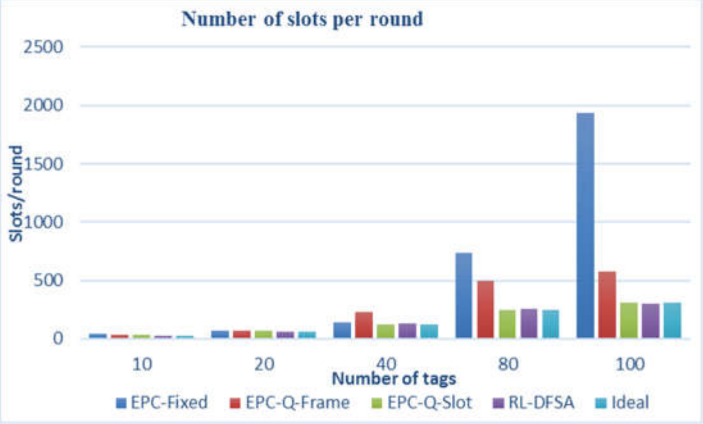


(a) Sparse tag environment (10 – 100 tags)

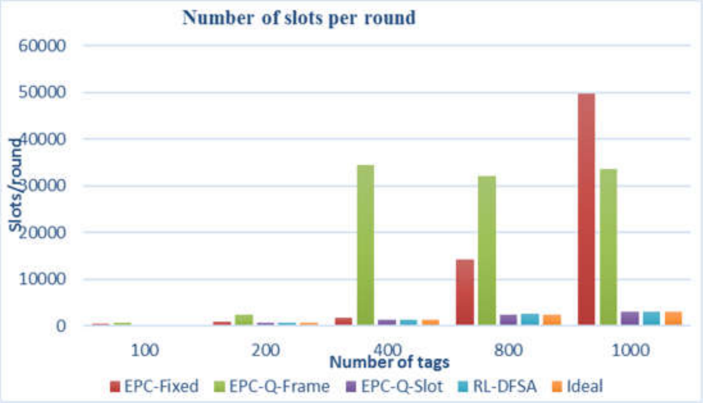


(b) Dense tag environment (100 – 1000 tags)

Figure 4.2 – Average number of frames required per interrogation round for various algorithms



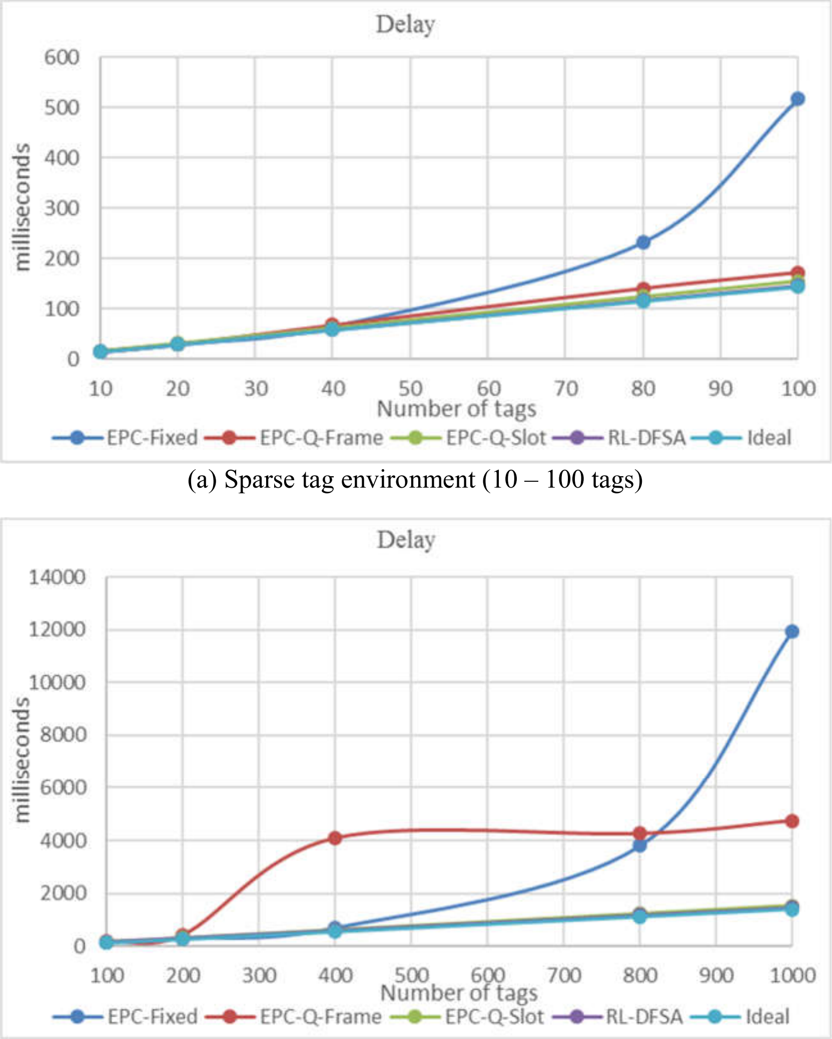
(a) Sparse tag environment (10 – 100 tags)



(b) Dense tag environment (100 – 1000 tags)

Figure 4.3 – Average number of slots required per interrogation round for various algorithm

Fig. 4.4 shows the delay performance of all five algorithms. The list of the algorithms ranked from best to worst for the delay performance is as follow: Ideal, RL-DFSA, EPC-Q-Slot, EPC-Q-Frame, and EPC-Fixed. Due to its static frame size, the EPC-Fixed algorithm took so much longer time as compared to the other four algorithms to finish an interrogation round. The difference is much more pronounced in the dense tag environment. As for the EPC-Q-Frame it performed much better since it has dynamic frame adaptation mechanism. RL-DFSA algorithm performed much better than EPC-Q-Slot algorithm despite using a frame-by-frame adaptation mechanism. In fact, RL-DFSA is 7.9% - 9.3% and 4.9% - 5.5% better in delay performance as compared to EPC-Q-Slot in sparse and dense tag environments, respectively. This shows the superiority of the employed learning algorithm which was able to learn an optimal policy even in a dynamic environment.



(a) Sparse tag environment (10 – 100 tags)

(a) Dense tag environment (100 – 1000 tags)

Figure 4.4 – Average time required per interrogation round for various algorithms

5. CONCLUSION

Energy efficiency is crucial in the internet of things application and RFID systems are no exception. Hence, a great amount of care must be taken when designing the algorithm so it has low overheads while being efficient in doing the intended task. In this work, we proposed an ACA which utilizes the Q-learning algorithm for selecting optimal frame size based on the number of collisions detected in the previous frame. The proposed RL-DFSA was trained with 1000 tags during the learning period and the resultant Q-matrix was used for evaluating the performance of the algorithm for varying number of tags from 10 to 1000. Its performance was compared with four algorithms, namely, EPC-Fixed, EPC-Q-Frame, EPC-Q-Slot, and Ideal. Through extensive simulations, it is concluded that RL-DFSA performs equal to or better than commercial algorithms at various performance metrics. Specifically, the number of frames required by RL-DFSA is an order of magnitude lower than the best performing algorithm that is currently being utilized by the commercial readers. Hence, RL-DFSA is proven to be an efficient anti-collision algorithm which is also energy efficient. The energy efficiency claim is valid since computational cost is more than 70 times cheaper as compared to the communication cost depending on the processor architecture. However, RL-DFSA still has some room for improvements as follow. The algorithm has slow convergence speed and its computational cost is relatively higher than the commercial algorithms. Thus, the application of another Q-learning derivative such as Speedy-Q can be investigated to speed up the learning process. Finally, RL-DFSA should be implemented on a software-defined radio platform and evaluated in real life applications.

ACKNOWLEDGEMENT

The author would like to acknowledge the support from the Fundamental Research Grant Scheme (FRGS) under a grant number of 9003-00608 from the Ministry of Higher Education Malaysia.

6. REFERENCES

[1] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and near-Field Communication*, Third Edition. Wiley, 2010.

[2] R. Want, “Enabling ubiquitous sensing with RFID,” *Computer (Long. Beach. Calif).*, vol. 37, no. 4, pp. 84–86, 2004.

[3] S. Ahson, *RFID Handbook: Applications, Technology, Security, and Privacy*. 2008.

[4] D.J. Deng and H.W. Tsao, “Optimal dynamic framed slotted ALOHA based anti-collision algorithm for RFID systems,” *Wirel. Pers. Commun.*, vol. 59, no. 1, pp. 109–122, 2011.

[5] C.H. Liao, T.K. Woo, C.C. Chen, and I.J. Su, “A novel grouping slotted Aloha scheme to enhance throughput performance for wireless networks,” *Wirel. Pers. Commun.*, vol. 96, no. 1, pp. 1229–1243, 2017.

[6] Y.I. Joo, D.H. Seo, and J.W. Kim, “An efficient anti-collision protocol for fast identification of RFID tags,” *Wirel. Pers. Commun.*, vol. 77, no. 1, pp. 767–775, 2014.

[7] Y. Chu, P.D. Mitchell, and D. Grace, “ALOHA and Q-Learning based medium access control for wireless sensor networks,” *Proceedings of the 2012 Int. Symp. Wirel. Commun. Syst.*, 2012, pp. 511–515.

[8] F.C. Schoute, “Dynamic frame length ALOHA,” *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 565–568, 1983.

[9] EPCglobal, *Specification for RFID Air Interface EPC TM Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID*, 2008.

[10] S. Dhakal and S. Shin, “Precise-optimal frame length based collision reduction schemes for frame slotted Aloha RFID systems,” *KSII Trans. Internet Inf. Syst.*, vol. 8, no. 1, pp. 165–182, 2014.

[11] C. Floerkemeier, “Transmission control scheme for fast RFID object identification,” *Proceedings of the Fourth Annual IEEE International Conference on Pervasive Comput. Commun. Work. PerCom Work. 2006*, vol. 2006, pp. 457–462, 2006.

[12] M.V. Bueno-Delgado and J. Vales-Alonso, “On the optimal frame-length configuration on real passive RFID systems,” *J. Netw. Comput. Appl.*, vol. 34, no. 3, pp. 864–876, 2011.

[13] B. Zhen, M. Kobayashi, and M. Shimizu, “Framed ALOHA for multiple RFID objects identification,” *IEICE Trans. Commun.*, vol. E88–B, no. 3, pp. 991–999, 2005.

[14] J.B. Eom and T.J. Lee, “Accurate tag estimation for dynamic framed-slotted ALOHA in RFID systems,” *IEEE Commun. Lett.*, vol. 14, no. 1, pp. 60–62, 2010.

[15] W.T. Chen, “A fast anticollision algorithm for the EPCglobal UHF class-1 generation-2 RFID standard,” *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1519–1522, 2014.

[16] G. Shaheen, *RFID Tag Identification Protocol Implementing Threshold-Based Dynamic Framed Slotted Aloha Policy*, Carleton University, 2010.

[17] L. Matignon, G.J. Laurent, and N. Le Fort-Piat, “Reward function and initial values: Better choices for accelerated goal-directed reinforcement learning,” *Artif. Neural Networks - ICANN 2006, Pt 1*, vol. 4131, pp. 840–849, 2006.

[18] L. Zhenzhen and E. Itamar, “RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks,” *Int. J. Sens. Networks*, vol. 1, no. 3, pp. 117–124, 2006.

[19] K.-L.A. Yau, H.G. Goh, D. Chieng, and K.H. Kwong, “Application of reinforcement learning to wireless sensor networks: models and algorithms,” *Computing*, vol. 97, no. 11. pp. 1045-1075, 2015.

[20] GAORFID, “Android Based UHF Gen 2 RFID Handheld Data Terminal 246029.” pp. 1–3, 2018.

[21] Rapidradio, “UHF Handheld Reader RRUHFHH2.” pp. 1–2, 2018.

[22] M. Ghavamzadeh, H. J. Kappen, M. G. Azar, and R. Munos, “Speedy Q-Learning,” *Adv.*

Neural Inf. Process. Syst., pp. 2411–2419, 2011.

[23] P. Šolić, J. Radić, and N. Rožić, "Energy efficient tag estimation method for ALOHA-Based RFID Systems," *IEEE Sens. J.*, vol. 14, no. 10, pp. 3637–3647, 2014.

[24] P. Šolić, M. Šarić, and M. Stella, "RFID reader-tag communication throughput analysis using Gen2 Q-algorithm frame adaptation scheme," *Int. J. Circuits, Syst. Signal Process.*, vol. 8, pp. 233-239, 2014.

[25] J. Wang, D. Wang, Y. Zhao, and T. Korhonen, "Fast anti-collision algorithms in RFID systems," *Proceedings of the Int. Conf. Mob. Ubiquitous Comput. Syst. Serv. Technol. UBICOMM 2007*, pp. 75–80, 2007.

[26] D. Klair, K. W. Chin, and R. Raad, "On the energy consumption of pure and slotted Aloha based RFID anti-collision protocols," *Comput. Commun.*, vol. 32, no. 5, pp. 961–973, 2009.

Murukesan Loganathan *obtained his B.E (2012) and M.Sc (2015) in Mechatronic Engineering from the University Malaysia Perlis (UniMAP), Malaysia. Starting from October 2015, he is doing his Ph.D. in Computer Engineering under the supervision of Dr. Thennarasan Sabapathy at UniMAP. His research interests are related to the development of energy efficient medium access control (MAC) protocols for the internet of things and wireless sensor networks.*

Thennarasan Sabapathy *received the B.Eng. degree in electrical-telecommunication engineering from Universiti Teknologi Malaysia in 2007, the M.Eng. degree from Multimedia University, Malaysia, in 2011, and the Ph.D. degree in communication engineering from Universiti Malaysia Perlis in 2014. In 2007, he joined Flextronics as a Test Development Engineer, focusing on the hardware and software test solutions for the mobile phone manufacturing. He was a Research Officer at Multimedia University from 2008 to 2010. From 2012 to 2014, he was a Research Fellow with Universiti Malaysia Perlis, where he is currently a Senior Lecturer with the School of Computer and Communication Engineering. His current research interests include antenna and propagation, millimeter-wave wireless communications, and fuzzy logic for wireless communications.*

Mohamed Elobaid Elshaikh *is a Senior Lecturer under School of Computer and Communication Engineering, University Malaysia Perlis, Malaysia. He received his Ph.D. in Computer Engineering from University Malaysia Perlis, Malaysia, M.Sc Electrical and Electronics Engineering from University Technology Petronas (UTP), Malaysia and B.Sc Engineering Technology (Computer Engineering), University of Gezira, Sudan. His research mainly is on computer networking related.*

Mohamed Nasrun Osman *was born Jitra, Malaysia, in 1987. He received the electrical engineering degree in telecommunication and the Ph.D. degree in electrical engineering from Universiti Teknologi Malaysia, in 2010 and 2016, respectively. He is currently a Senior Lecturer with Universiti Malaysia Perlis, Malaysia. His research interests include reconfigurable antenna design, RF design, and wireless multi-in multi out systems.*

Rosemizi Abd Rahim *was born in Kedah, Malaysia in 1976. He received the B.Eng. degree in Electrical Engineering from the Universiti Teknologi Mara, Malaysia, in 2000 and the M.Sc. degree in Electronic System Design Engineering from the Universiti Sains Malaysia, in 2004. In 2013, he received the Ph.D. degree in Communication Engineering from the Universiti Malaysia Perlis. From 2000 to 2004, he was a failure analysis engineer at a multinational electronic manufacturing company in Penang, Malaysia. His task was to resolve any failure that occurs during the production process of electronic products. Then, since 2005 he has been moved to the Universiti Malaysia Perlis as an academician. His research interest includes the analysis and development of new sources of energy harvesting system and techniques, antenna design and microwave engineering.*

Muzammil Jusoh *received the bachelor's degree in electrical, electronic and telecommunication engineering and the M.Sc. degree in electronic telecommunication engineering from Universiti Teknologi Malaysia (UTM), in 2010 and 2006, respectively, and the Ph.D. degree in communication engineering from Universiti Malaysia Perlis (UniMAP) in 2013. He was an RF and*

Microwave Engineer with the Telekom Malaysia Berhad (TM) Company from 2006 to 2009. He was an Engineer (Team Leader) with the Specialized Network Services Department, TM Senai, Johor. He was involved in the preventive and corrective maintenance of ILS, NDB, DVOR, repeater, microwave system, VHF, and UHF based on contract wise. He was with the Department of Civil Aviation, TUDM, PDRM, ATM, Tanjung Pelepas Port, MCMC, and JPS (Hidrologi Department). He is currently an Associate Professor and a Researcher with the School of Computer and Communication Engineering, UniMAP. He is supervising a number of Ph.D. and M.Sc. students and also managing a few grants under the Ministry of Higher Education Malaysia. He has published a number of quality journals such as the IEEE Antennas and Wireless Propagation Letters, Microwave and Optical Technology Letters, the International Journal of Antennas and Propagation, Progress in Electromagnetics Research, and Radio Engineering and over 70 conference papers. His research interests include antenna design, reconfigurable antennas, multi-in multi-out, ultra-wideband, and wireless communication systems.

***Badlishah Ahmad** obtained B.E. in Electrical and Electronic Engineering from Glasgow University in 1994. He obtained his M.Sc. and Ph.D. in 1995 and 2000 respectively from University of Strathclyde, UK. His research interests are on computer and telecommunication network modeling using discrete event simulators, optical networking and coding, and embedded system based on GNU/Linux for vision. He has five (5) years teaching experience in University Sains Malaysia. Since 2004 until 15 March 2017, he was working with University Malaysia Perlis (UniMAP) as Dean at School of Computer and Communication Engineering. Currently as the Deputy Vice-Chancellor (Research and Innovation), Universiti Sultan Zainal Abidin (UniSZA).*

***Mohd Ilman Jais** received the bachelor's, M.Sc. (Hons.) and Ph. D degrees in communication engineering from Universiti Malaysia Perlis, Malaysia, in 2011, 2014 and 2018, respectively, where he is currently working as a senior lecturer. His research interests include reconfigurable antennas, fuzzy inferences systems, embedded systems, and Internet of Things.*