Design and Analysis of Framed Aloha based RFID Anti-collision Algorithms

Lei Zhu and Tak-Shing Peter Yum
Department of Information Engineering, The Chinese University of Hong Kong
{lzhu6, tsyum}@ie.cuhk.edu.hk

***Abstract*—The anti-collision mechanism is a very important part in RFID systems. Among all the algorithms, the Framed Aloha based (FA) ones are most widely used due to its simplicity and robustness. Previous works mainly focused on the tag population estimation, but determined the reading strategy based on the classical results of Random Access (RA) systems. We show that a new theory is needed for the optimization of the RFID systems as they have characteristics very different from the RA systems. We model the reading process as a Markov Chain and derive the optimal reading strategy through first-passage-time analysis. We show that the optimal strategy can be easily incorporated into the EPCglobal standards to give significant performance improvement.**

I. Introduction

In Radio-frequency Identification (RFID) systems, tags share a common communication channel. Therefore, if multiple tags transmit at the same time, their packets will collide and get lost [1]. Passive tags have bare-bone functionality and no imbedded power supply. They cannot sense the media or cooperate with one another. The RFID reader needs to coordinate their transmissions to avoid collisions. Depending on working principles, RFID anti-collision algorithms in literature can be divided into three main classes: Tree based algorithms [5][6], Framed Aloha (FA) based algorithms [8-16] and Interval based algorithms [7]. Among these, only FA based algorithms are widely used in RFID communication standards [2-4] for their simplicity and robustness.

In different literature, there are variations on the working mechanisms of FA based RFID systems, but not all of them are used in real applications. Up to now, the most popular RFID system is the one defined in the EPCglobal standards [2]. We summarize its anti-collision mechanism as follows:

1) The reader starts a frame by broadcasting a special command 'QueryAdjust' with a parameter L[1]. Each tag chooses a random value from 0 to $L-1$ as its transmission delay. Those generating '0' contend the channel immediately.
2) The reader uses the 'QueryRep' command to ask tags to decrement their counters by 1. Tags contend the channel when their counters reach 0.
3) When contending the channel, the tag only sends a short packet containing its temporary ID. In EPCglobal standards, this 'ID' is named 'RN16' (random number 16 bits). If only one tag replies, the reader can receive this short packet successfully.
4) The reader has a set of *operation* commands: reading data, writing new data, changing password, etc. After receiving a temporary ID, the reader can select a particular tag by including its ID in the operation commands.
5) The reader can send a *silence* command[2] to a selected tag. Silenced tags will not contend the channel in future frames.
6) The reader can use the 'QueryAdjust' command again even before all tags' counters reach 0. When hearing this command, unsilenced tags regenerate their counter values according to the new frame size. Those generating '0' contend the channel immediately.

RFID systems satisfying the above properties are called **Gen-2 RFID systems.** One feature of this system is the 3-way handshaking mechanism. (The reader sends a *probe* (QueryAdjust or QueryRep); tags reply their temporary ID; the reader sends the temporary ID back in operation commands.) Conventionally the time from the point that the reader sends out a *probe* to the point that the tags finish replying their temporary IDs is called a *contention time slot*, or just *slot* for short. Since the communication after the 3-way handshaking is collision-free, the time involved does not depend on the anti-collision strategy. Therefore, the performance of anti-collision algorithms is conventionally compared by their average contention time measured by the number of contention slots.

Another feature is frame cancelation. The reader can initiate a new frame using the *'QueryAdjust'* command whenever the current frame size is found unsuitable. So a Gen-2 algorithm should specify how to choose frame size and when to cancel a running frame.

The goal of algorithm design is to minimize the average contention time T. Unfortunately, T cannot be expressed explicitly as a function of L. So previous research [8-15] uses the expected instantaneous throughput U as the optimization objective instead. In Random Access theory, a classical formula to calculate the throughput U with terminal population N and frame size L is given in [10] as:

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

In (1), U can be optimized by setting the frame size equal to the terminal number, or $L = N$. Based on this, previous algorithms [8-15] either directly set the frame size L equal to

[1] In EPCglobal standards, the frame sizes are limited to 2^Q, where $Q = 0, 1, \ldots, 15$. In this paper, we consider both this special case and the general case where $L \in \mathbb{Z}_{++}$.

[2] In some literature, this command is also referred to as *Kill* command. In EPCglobe standards, it corresponds to the *Select* command, which have other uses besides silencing a tag.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

the expected tag population $E[N]$ or try to find an optimal L to maximize U for a given probability distribution of N. However, these approaches are not suitable because maximizing U in every frame will not necessarily minimize the average contention time. As we will show later, optimizing L based on (1) will only yield a frame-local optimal result, and the concatenation of locally optimal results are usually far from the globally optimal one.

In this paper, we model the reading process as a Markov Chain and derive the optimal reading strategy through first-passage-time analysis. To the best of our knowledge, it is the first time that the performance bound of the Gen-2 RFID system is rigorously derived. We show that the optimal strategy can be easily incorporated into the EPCglobal standards to give significant performance improvement with minimum increase of the system complexity.

In section II, we give a survey of the traditional strategies of FA algorithms and point out an unjustified assumption used in previous attempts of reading strategy optimization. In section III, we derive the optimal reading strategy for the precise estimation case. In section IV, we generalize the strategy to the imprecise estimation case and show the simple use of it in EPCglobal standards.

II. A Survey of Previous Works

In this section, we give a survey of the previous works and show why a new theory is needed.

In real applications, the number of tags are unknown before identification. So a proper FA algorithm always contains two parts: **Population Estimation** part and **Reading Strategy Determination** part. The first part is for estimating the tag population based on tags' replies while the second part is for adjusting the parameter, such as the frame size, using the estimation. Based on the different Estimation method, algorithms can be divided into: the max-likelihood approach and the probability distribution approach.

A. The max-likelihood approach

Schoute [10] noticed that when N is large and L suitably chosen (say $L \approx N$), the number of tags attempting each slot has a Poisson distribution with mean 1. So in the *Population Estimation* part, his algorithm uses $\hat{N} = \text{round}(2.39 s_c)$, where s_c is the number of collided slots in the last frame. Based on this, in the *Reading Strategy Determination* part, the frame size is set as

$$L = \hat{N}. \tag{2}$$

This choice is based on (1). It tries to maximize the instantaneous throughput by setting the frame size equal to the expected terminal number.

Vogt [11] improved the *Population Estimation* strategy of Schoute's algorithm by using the statistics of empty slots s_e and singleton slots s_s in addition. Tag population is estimated to be the value $\hat{N}$ that minimizes the error between the observed values of s_e, s_s, s_c and their expected values using $\hat{N}$. In the *Reading Strategy Determination* part, it also uses (2).

Kodialam [16] proposed a new *Population Estimation* strategy based on the Central Limit Theorem. That is when the number of contending tags is large enough, the number of collision slots and empty slots in a frame should obey the Normal distribution. Thus using his method, one may obtain the estimation accuracy bound as well as the max-likelihood tag population. But after deriving $\hat{N}$, it also sets $L = \hat{N}$.

Another example is the Q algorithm in EPCglobal standards [2]. The RFID reader maintains a floating-point variable Q_{fp}. It decreases a typical value C when no tag replies, increases C when multiple tags reply and stays unchanged when only 1 tag replies.[3] The tag population is estimated as $\text{round}(2^{Q_{fp}})$ while the frame size is set to 2^Q, where $Q = \text{round}(Q_{fp})$. In [13][14], the efficiency of the Q algorithm was obtained with different choices of C and Q_{fp} and some methods to improve the estimation strategy were proposed.

In summary, algorithms of this type compute the maximum-likelihood tag population $\hat{N}$ based on the reading results and set $L = \hat{N}$ as the frame size. The advantage is simplicity. In Q algorithm, the reader only needs to perform the '*add*' operation once every time slot.

B. The Distribution approach

Floerkemeier [12][15] assumes that a rough estimation of the target group size is always available in the form of a distribution $\Pr\{N = n\}$. As a new *Population Estimation* strategy, it updates the population distribution by Bayesian method based on tags' replies. In the *Reading Strategy Determination* part, the frame size is chosen as

$$L^* = \left\{ L : \max_{L \in \Upsilon} \sum_{n=0}^{N_{max}} U(N = n, L) \Pr\{N = n\} \right\}, \tag{3}$$

where Υ is the set of possible frame sizes while $U(N, L)$ is calculated by (1). A running frame will be canceled when the updated distribution prefers another value of L^* according to (3).

This approach can track the value of N more accurately. The tradeoff is complexity. In every time slot, the reader needs to do Bayesian update N times and optimize L according to (3).

C. The need for a new model

As we can see from this review, previous work focused on the *Population Estimation*, i.e., proposing different ways to estimate N. On the other hand, the *Reading Strategy Determination* part is underdeveloped. All the algorithms in literature use (1) for calculating throughput. As we mentioned before, (1) is obtained from the theory of Random Access (RA) systems. Since a terminal in an RA system would still attempts the channel after a successful transmission, the 'contending group' can be assumed unchange during a long enough period. The long-term throughput of a RA system is therefore equal to the expected instantaneous throughput U calculated by (1). However, in RFID systems, identified tags are silenced by the reader, leading to tag population decrease during the reading process. When the frames are not identical,

[3]In EPCglobe standards, it is recommended that $0.2 \le C \le 0.5$ and the initial $Q_{fp} = 4$

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

a concatenation of locally optimal solutions is not globally optimal. As an example, suppose the target group contains exactly 3 tags. From (1), the suitable frame size should be $L = 3$ and the efficiency is $U = 0.44$. However, if we choose $L = 2$ and follow the strategy in Figure 1, the efficiency can achieve 0.6.

In this paper, we focus on the *Reading Strategy Determination* part. Based on a given $\hat{N}$, no matter which estimation method is used[4], we derive the *Optimal Reading Strategy* that can minimize the expected contention time. In the following, we first derive the optimal reading strategy for the precise estimation case. Using the results, we generalize the analysis to the general cases.

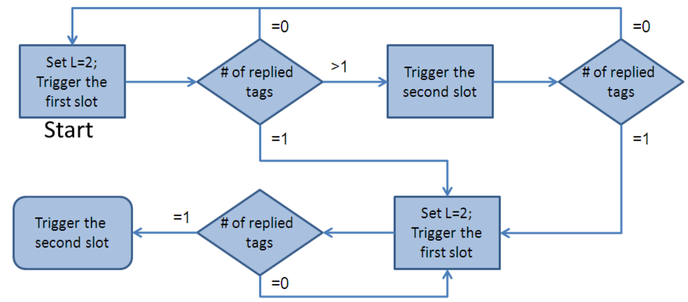

Fig. 1. The optimal strategy to read three tags

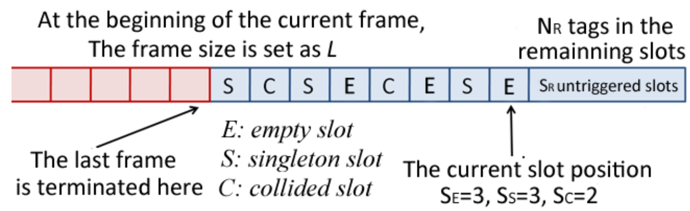

Fig. 2. An intermediate step of the reading process

III. Optimal Reading Strategy with Precise Population Estimation

To derive the optimal reading strategy, we first assume the estimated tag population $\hat{N}$ is precise, or $\hat{N} = N$. This assumption will be removed in section IV. In subsection A, we derive the **System State** for the optimal reading strategy. In subsection B, we model the reading process by a Markov Chain and derive the **System Equation** which establishes the functional relationship between the System State and the expected contention time. In subsection C and D, the System Equation is solved in different cases to yield the optimal reading strategies. For ease of referencing, we list all major variables used in analysis as follows:

L: the current frame size
N: the tag population at the beginning of the current frame
S_E: the number of empty slots in the current frame
S_S: the number of singleton slots in the current frame
S_C: the number of collided slots in the current frame
S_R: the number of remaining slots in the current frame
N_R: the number of tags in the remaining slots
N_U: the number of unidentified tags
r_e: the probability that the next slot is empty
r_s: the probability that the next slot is singleton
r_c: the probability that the next slot is collided

A. System State

Figure 2 shows an intermediate step of the reading process. Let L denote the frame size and N denote the unidentified tag population at the beginning of this frame. In this section, we assume N is precisely available from the estimation. With this assumption, we can focus on the reading strategy part. Further let S_E, S_S and S_C denote the number of empty slots, singleton slots and collided slots up to the current slot in the current frame. To illustrate, the current slot position in Figure 2 is 8 and the current reading result is $(S_E = 3,\ S_S = 3,\ S_C = 2)$.

Suppose $L > 8$. Then the reader has two options: 1) continue this frame and trigger the next slot, or 2) terminate this frame and start a new one.[5] The decision is made according to the **Cancelation Rule** of an algorithm.

Let S_R be the number of Remaining (untriggered) slots in the current frame and N_R be the number of tags in these slots. Ideally, the reader should choose option 1 when $S_R \approx N_R$ and choose option 2 otherwise. Although S_R can be obtained as $S_R = L - S_E - S_S - S_C$, the precise value of N_R is usually unavailable when some slots are collided. Let $\mathrm{B}el(N_R)$ denote the *belief* of N_R, or the conditional distribution of N_R based on all the information we know [18]. Due to the memoryless property of passive tags, $Bel(N_R)$ is independent of all the previous frames given L and N. Thus we have

$$Bel(N_R) = \Pr\{N_R | S_S,\, S_C,\, S_R,\, N,\, L\}.$$

Lemma 1: $Bel(N_R)$ is a function of N_U, S_R and S_C only, where $N_U = N - S_S$ is the unidentified tag population.

Proof: By definition, $Bel(N_R) = \Pr\{N_R|S_S, S_C, S_E, N, L\}$. Since $S_R = L - S_S - S_C - S_E$, S_E can be replaced by S_R as a condition. With this substitution, we use Bayes rule to obtain:

$$\begin{aligned} Bel(N_R) = \Pr\{N_R \,|\, S_S,\, S_C,\, S_R,\, N,\, L\} = \\ \Pr\{S_S,\, S_C \,|\, N_R,\, S_R,\, N,\, L\} \Pr\{N_R \,|S_R,\, N,\, L\} Z^{-1}, \end{aligned} \quad (4)$$

where Z is the normalization constant. The first term in (4) can be derived by drawing analogy to the urn problem [17]. Specifically, when putting $N - N_R$ balls into $L - S_R$ urns, the probability that S_S urns contain exactly 1 ball, S_C urns

[4]The choice of population estimation method depends on hardware capability of the reader. Often, very elaborate statistical estimation methods are not suitable due to real-time requirement.

[5]If the current frame is terminated, the tag population N will be updated as $N^{(new)} = N - S_S$. The variables S_E, S_S and S_C will be reset to track the reading results of the new frame.

contain more than 1 balls and the others are empty is:

$$\Pr\{S_S = s_s,\, S_C = s_c \,|\, N_R = n_r,\, S_R = s_r,\, N = n,\, L = l\} = \binom{l - s_r}{s_s, s_c, l - s_r - s_s - s_c} \frac{(n-n_r)!}{(n-n_r-s_s)!(l-s_r)^{n-n_r}} * \sum_{\substack{m_1, m_2, \ldots, m_{sc} \geq 2, \\ m_1+m_2+\ldots+m_{sc}=n-n_r-s_s}} \binom{n-n_r-s_s}{m_1, m_2, \ldots, m_{sc}}, \tag{5}$$

where $m_1, m_2, \ldots, m_{sc}$ denote the numbers of tags in the s_c collided slots. Similarly, the second term in (4) is the probability that the last S_R urns contain N_R balls when N balls are randomly put into L urns, or

$$\Pr\{N_R = n_r \,|\, S_R = s_r, N = n, L = l\} = \binom{n}{n_r} \left(\frac{s_r}{l}\right)^{n_r} \left(1 - \frac{s_r}{l}\right)^{n-n_r}. \tag{6}$$

Substituting (5) and (6) into (4), we obtain:

$$\begin{aligned} & Bel(N_R) \\ = \; & \Pr\{N_R = n_r | S_S = s_s, S_C = s_c, S_R = s_r, N = n, L = l\} \\ = \; & \frac{1}{Z} \frac{s_r^{n_r}}{n_r!(n_u - n_r)!} \\ & * \sum_{\substack{m_1, m_2, \ldots, m_{sc} \geq 2, \\ m_1+m_2+\ldots+m_{sc}=n_u-n_r}} \binom{n_u - n_r}{m_1, m_2, \ldots, m_{sc}}, \end{aligned} \tag{7}$$

where $n_u = n - s_s$ is the number of unidentified tags.

From (7), it is clear that $Bel(N_R)$ depends only on N_U, S_R and S_C. □

When the RFID system is treated as an intelligent system [18], The initial information (N, L) and the evidence (S_E, S_S, S_C) [6] together cover all the information in the system and are sufficient for the optimal decision. But some information is redundant, a smaller sufficient set can be obtained.

Let $V = (N_U, S_C, S_R)$ and $\mathscr{V}_N$ be a set of all the possible combinations of (N_U, S_C, S_R) beginning from the tag population N, or

$$\mathscr{V}_N = \Big\{(N_U, S_C, S_R) \,\Big|\, N_U, S_C, S_R \in \mathbb{Z}_+, \; N_U \leq N,\; S_C \leq \frac{N_U}{2},\; S_C + S_R \leq L_N^*\Big\},$$

where L_N^* is the optimal frame size for N tags. Since the value of L_N^* is around N, or $L_N^* \sim \mathcal{O}(N)$, we have $|\mathscr{V}_N| \sim \mathcal{O}(N^3)$.

Theorem 1: When the tag population is precisely estimated, the optimal Cancelation Rule depends only on $V = (N_U, S_C, S_R)$.

Proof: Let $\mathcal{T}(N)$ denote the expected contention time for N tags using the optimal strategy when this N is precisely estimated. Although the value of $\mathcal{T}(N)$ is not available yet, it should be exact and depends only on N. We prove the theorem by mathematical induction as follows.

The first case is $V \in \mathcal{V}_1$, where $\mathcal{V}_1 = \{(N_U, S_C, S_R) \in \mathscr{V}_N : S_R = 1\}$, or there is only 1 remaining slot in the current frame. If the current frame is canceled, the expected finishing time for the N_U unidentified tags should be $T_1 = \mathcal{T}(N_U)$. Otherwise, if the reader triggers the last slot, the expected finishing time can be obtained by averaging the different outcomes of the last slot as

$$\begin{aligned} T_2 &= 1 + \Pr\{\text{The last slot is singleton}\}\, \mathcal{T}(N_U - 1) \\ &\quad + \Pr\{\text{The last slot is empty or collided}\}\, \mathcal{T}(N_U) \\ &= 1 + r_s\, \mathcal{T}(N_U - 1) + (1 - r_s)\, \mathcal{T}(N_U). \end{aligned}$$

In this case, $r_s = Bel(N_R = 1)$, because there is only 1 slot left. By Lemma 1, $Bel(N_R)$ is a function of N_U, S_C and S_R. Thus T_2 depends only on V. Obviously, the frame will be canceled when $T_1 < T_2$. So the cancelation rule depends only on V. Without loss of generality, we let $\mathcal{T}_F^{(1)}(V) = \min\{T_1, T_2\}$ denote the mapping from V to the expected finishing time when $V \in \mathcal{V}_1$.

We assume the theorem still holds when $V \in \mathcal{V}_k$, where $\mathcal{V}_k = \{(N_U, S_C, S_R) \in \mathscr{V}_N : S_R = k\}$ $(k \geq 1)$, or the expected finishing time for V can be obtained from $T = \mathcal{T}_F^{(k)}(V)$. Then for the case that $V \in \mathcal{V}_{k+1}$, where $\mathcal{V}_{k+1}$ is similarly defined, we have

1) If the current frame is canceled, the expected finishing time is $T_1 = \mathcal{T}(N_U)$.

2) If the next slot is triggered, only k slots left. Let v_e, v_s and v_c denote the triples when the triggered slot is empty, singleton and collided respectively. We have $v_e,\, v_s,\, v_c \in \mathcal{V}_k$. Then the expected contention time is

$$T_2 = 1 + r_e\, \mathcal{T}_F^{(k)}(v_e) + r_s\, \mathcal{T}_F^{(k)}(v_s) + r_c\, \mathcal{T}_F^{(k)}(v_c),$$

where r_e, r_s and r_c denote the probability that the next slot contains 0, 1 and multiple tags respectively, which can be obtained from $Bel(N_R)$ as

$$r_e = \sum_{i=0}^{N_U} \left(1 - \frac{1}{S_R}\right)^i Bel(N_R = i); \tag{8}$$

$$r_s = \sum_{i=1}^{N_U} \frac{i}{S_R} \left(1 - \frac{1}{S_R}\right)^{i-1} Bel(N_R = i); \tag{9}$$

$$r_c = 1 - r_e - r_s. \tag{10}$$

Similar to Case 1, T_2 depends only on V. As the frame will be canceled when $T_1 < T_2$, the theorem still holds.

By mathematical induction, we can claim for any $V \in \mathscr{V}_N$, the optimal cancelation rule and the expected finishing time depends only on V. □

From Theorem 1, V is the system state which determines the reading strategy. Let $\mathcal{C}(V)$ be the optimal *Cancelation Rule* defined on $V \in \mathscr{V}_N$, i.e. $\mathcal{C}(V) = 1$ when the current frame should be canceled and $\mathcal{C}(V) = 0$ otherwise.

B. The Markov Chain

Let $V_j = (N_U, S_C, S_R)$ denote the state of the reading process in time slot j. For N tags to identify, let $V_0 = (N,\, 0,\, L_N^*)$ be the initial state and $\mathcal{U}_T = \{(0, 0, S_R) \in \mathscr{V}_N : S_R \in \mathbb{Z}_+\}$ be the terminal states.

Theorem 2: When the initial tag population is known, the states $V_0 V_1 V_2 \ldots$ following the optimal reading strategy form a Markov Chain.

[6]It can be proved that the position of the empty slots, singleton slots and collided slots does not matter. Only their respective sums affect the reading decision.

Proof: From Theorem 1, the transition probability from a particular state $V_j = (n_u, s_c, s_r)$ to V_{j+1} is obtained as follows:

$$\Pr\{V_{j+1} \Big| V_j V_{j-1} \dots V_0\}$$
$$= \begin{cases} r_e \mathcal{C}(n_u, s_c, s_r - 1) & , \quad V_{j+1} = (n_u, 0, L^*_{n_u}) \\ r_s \mathcal{C}(n_u - 1, s_c, s_r - 1) & , \quad V_{j+1} = (n_u - 1, 0, L^*_{n_u - 1}) \\ r_c \mathcal{C}(n_u, s_c + 1, s_r - 1) & , \quad V_{j+1} = (n_u, 0, L^*_{n_u}) \\ r_e \left[1 - \mathcal{C}(n_u, s_c, s_r - 1)\right] & , \quad V_{j+1} = (n_u, s_c, s_r - 1) \\ r_s \left[1 - \mathcal{C}(n_u - 1, s_c, s_r - 1)\right] & , \quad V_{j+1} = (n_u - 1, s_c, s_r - 1) \\ r_c \left[1 - \mathcal{C}(n_u, s_c + 1, s_r - 1)\right] & , \quad V_{j+1} = (n_u, s_c + 1, s_r - 1) \\ 0 & , \quad \text{others} \end{cases}$$

From (8) $\sim$ (10), r_e, r_s and r_c depends on V_j only. Hence $\Pr\{V_{j+1} \,|\, V_j V_{j-1} \dots V_0\} = \Pr\{V_{j+1} \,|\, V_j\}$, and the evolution of V_j following the optimal reading strategy is a Markov Chain. □

For all $V \in \{(N_U, S_C, S_R) \in \mathscr{V}_N : N_U > 0\}$, let $\mathcal{T}_F(V)$ denote the expected first passage time from V_j to $\mathcal{U}_T$. Then we have the following equations from Markov Chain theory as

$$\begin{aligned} \mathcal{T}_F(V_j) &= 1 + \sum_{V_{j+1} \in \mathscr{V}_N} \Pr\{V_{j+1} | V_j\} \, \mathcal{T}_F(V_{j+1}) \qquad (11) \\ &= 1 + r_e \mathcal{T}_F\left(V^{(e)}_{j+1}\right) + r_s \mathcal{T}_F\left(V^{(s)}_{j+1}\right) + r_c \mathcal{T}_F\left(V^{(c)}_{j+1}\right), \end{aligned}$$

where where

$$V^{(e)}_{j+1} = \Big\{ V : \min_V \mathcal{T}_F(V),$$
$$V \in \Big\{ (n_u, s_c, s_r - 1), \, (n_u, 0, L^*_{n_u}) \Big\} \Big\},$$
$$V^{(s)}_{j+1} = \Big\{ V : \min_V \mathcal{T}_F(V),$$
$$V \in \Big\{ (n_u - 1, s_c, s_r - 1), \; (n_u - 1, 0, L^*_{n_u - 1}) \Big\} \Big\},$$
$$V^{(c)}_{j+1} = \Big\{ V : \min_V \mathcal{T}_F(V),$$
$$V \in \Big\{ (n_u, s_c + 1, s_r - 1), \; (n_u, 0, L^*_{n_u}) \Big\} \Big\}.$$

For the special cases $V_j = (n, 0, L^*_n)$ (the beginning of a frame), since the optimal frame size L^*_n is not available yet, (11) becomes

$$\mathcal{T}_F(V_j) = \mathcal{T}(n) = \qquad (12)$$
$$\min_L \left\{ 1 + r_e \mathcal{T}_F\left(V^{(e)}_{j+1}\right) + r_s \mathcal{T}_F\left(V^{(s)}_{j+1}\right) + r_c \mathcal{T}_F\left(V^{(c)}_{j+1}\right) \right\},$$

where

$$\begin{aligned} V^{(e)}_{j+1} &= \Big\{ V : \min_V \mathcal{T}_F(V), \; V \in \Big\{ (n, 0, L - 1), \, (n, 0, L^*_n) \Big\} \Big\}, \\ V^{(s)}_{j+1} &= \Big\{ V : \min_V \mathcal{T}_F(V), \\ &\qquad V \in \Big\{ (n - 1, 0, L - 1), \; (n - 1, 0, L^*_{n-1}) \Big\} \Big\}, \\ V^{(c)}_{j+1} &= \Big\{ V : \min_V \mathcal{T}_F(V), \; V \in \Big\{ (n, 1, L - 1), \; (n, 0, L^*_n) \Big\} \Big\}. \end{aligned}$$

(11) and (12) are the **System Equations** of the optimal strategy for Gen-2 RFID systems, which establishes the functional relationship between the expected contention time and system state. The optimal frame size L^* and the cancelation rule $\mathcal{C}(V)$ can be obtained by solving the system Equations.

C. Solution of the System Equation for $\Upsilon = \mathbb{Z}_{++}$

The system equations include $|\mathscr{V}_n|$ nonlinear equations. Limited by space, we skip the analytical derivation of the solution and only provide an iterative program for computing the expected contention time. Starting from $\mathcal{T}(1) = 1$, $\mathcal{T}_F(V)$ for $V \in \mathscr{V}_N$ $(N \geq 2)$ can be obtained by running an iterative program as follows:

For n=2:N
  1) Let $\mathcal{T}(n) = \mathcal{T}(n-1) + 2.71818$ as the initial value.
  2) For $s_r = 1 : \lceil 1.5n \rceil$
      For $s_c = 0 : \min(2n - s_r, \lfloor n/2 \rfloor)$
        Update $\mathcal{T}_F(n, s_c, s_r)$ from (11);
      End For
    End For
  3) Update $\mathcal{T}(n)$ using (12). Repeat step 2 if the difference of old and new values of $\mathcal{T}(n)$ exceeds a threshold.
End For

This iterative algorithm is guaranteed to converge as (11) and (12) satisfy the contraction mapping conditions [19]. In our experiment, the difference converges to 0.001 within several loops. The results are obtained as follows:

The optimal frame size: As shown in Figure 3, for n ($n > 2$) tags to be read, the optimal frame size L^*_n is found to be a little less than n. This serves as a correction to the results in [11][12][15]. This difference is apparently caused by the introduction of frame cancelation in Gen-2 RFID systems.

The expected contention time: For $N \leq 40$, the performance of the optimal strategy is shown in Figure 4. We can see that the average contention time per tag of the Gen-2 RFID system is always below the bound of RA systems [10].

The cancelation rule: Figure 5 shows the optimal cancelation rule for $N_U = 40$. The current frame should be terminated whenever the state wanders outside the permitted region marked by $\times$. For other values of N_U, similar permitted regions can be found.

D Special Solutions for $\Upsilon = \{2^i \,|\, i \in \mathbb{Z}_+\}$

In this special case, the system equation can be solved similarly. The result is obtained by running the iterative program with limited choice of L:

1) The optimal frame size is shown in Figure 3 by marks $'+'$. We see that these are just quantized values of the previous case.
2) The performance is shown in Figure 4 by marks $'+'$. We observe that although the choices of frame size are severely limited, the performance loss is very small.
3) The cancelation rule for $N_U = 40$ is shown in Figure 6. Compared with the case in subsection C, the permitted region is larger. As a result, frame cancelation is significantly less than that of the previous case.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

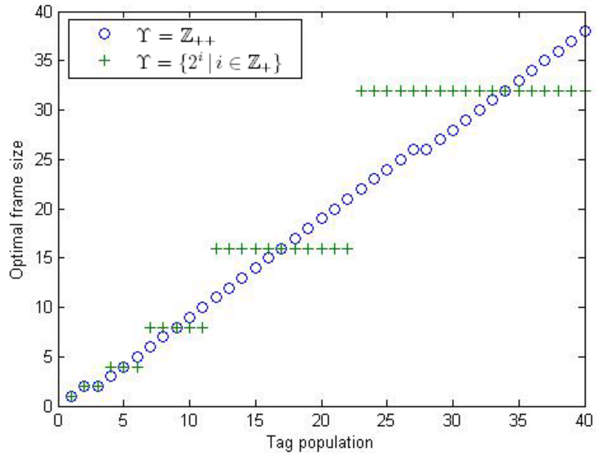

Fig. 3. The optimal frame size

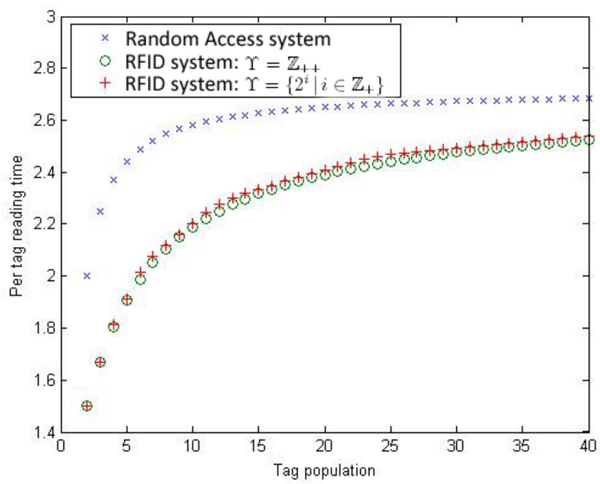

Fig. 4. The average contention time per tag using the optimal reading strategy

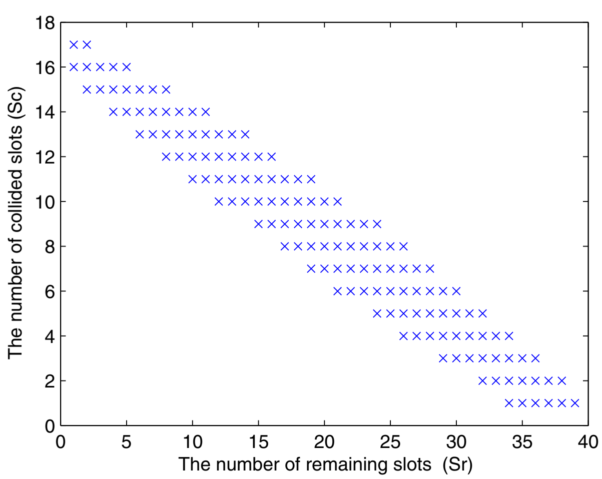

Fig. 5. The cancelation rule for $N_U = 40$, $\Upsilon = \mathbb{Z}_{++}$

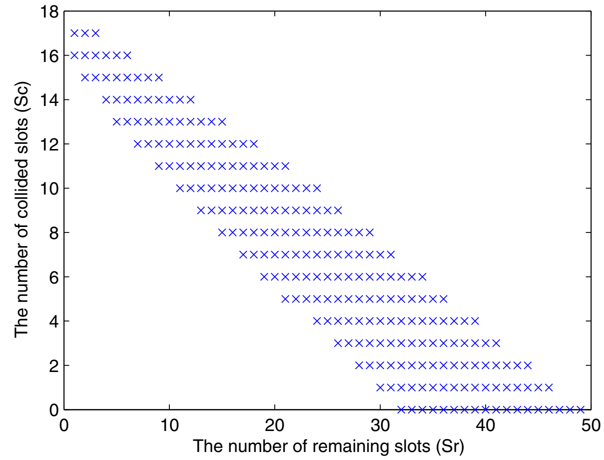

Fig. 6. The cancelation rule for $N_U = 40$, $\Upsilon = \{2^i \,|\, i \in \mathbb{Z}_+\}$

IV. Generalization to Imprecise Population Estimation

In this section, we remove the assumption that the estimated tag population is precise. As mentioned in Section III, most FA algorithms adopt the maximum-likelihood approach for simplicity concern. We show significant improvement can be obtained when the optimal reading strategy is used in maximum-likelihood estimation algorithms.

Now consider the case where the estimated tag population $\hat{N}$ may not equal to N. We do not care about how this $\hat{N}$ is obtained, i.e. it can be obtained by Schoute's method [10], or Vogt's method [11], or something else. When the precise value of N is replaced by $\hat{N}$, two new problems arise:

1) What is the optimal frame size for $\hat{N}$?

Previous maximum-likelihood algorithms choose frame size $L = \hat{N}$. This is reasonable, as $\hat{N}$ is the only knowledge using maximum-likelihood estimation. When a suitable estimation method is used, the estimation becomes more and more accurate as reading proceeds. In Q algorithm, $\hat{N}$ is very close to N for most of the time. Thus choosing the frame size as $L^*_{\hat{N}}$ is the best we can do.

2) Should the current frame be canceled when $\hat{N}$ changes?

In EPCglobal standards, the Q algorithm cancels the current frame whenever round(Q_{fp}) changes. This action is improper as although the frame size is wrongly chosen, the number of remaining slots may still be suitable for the remaining tags. Our analysis show that the cancelation decision is only determined by $V = (N_U, S_C, S_R)$ instead of the frame size. For an estimated tag population $\hat{N}$, the whole theory can be similarly proved. Therefore we can simply replace N_U by $\hat{N}$ and check whether V satisfies the cancelation rule.

We now show how to adapt the optimal reading strategy into the Q algorithm to obtained the Improved Q algorithm, or IQ algorithm for short.

Q algorithm: (From the EPCglobal standards [2])

1) Set the initial value for Q_{fp} and C.

2) Set the frame size $L = 2^Q$, where $Q = \text{round}(Q_{fp})$.
3) During the reading process, maintain a variable S_R to track the number of remaining slots and another float-point variable Q_{fp} as

$$Q_{fp} = \begin{cases} Q_{fp} + C, & \text{collided reply} \\ Q_{fp} + 0, & \text{singleton reply} \\ Q_{fp} - C, & \text{no reply} \end{cases}$$

4) Cancel the current frame when $Q = \text{round}(Q_{fp})$ changes and set the new frame size as $L = 2^Q$.

IQ algorithm:

1) Set the initial value for Q_{fp} and C.
2) Let $\hat{N} = \text{round}(2^{Q_{fp}})$ and set the frame size as $L = L^*_{\hat{N}}$.
3) During the reading process, track variables S_R and S_C and maintain another float-point variable Q_{fp} as

$$Q_{fp} = \begin{cases} Q_{fp} + C, & \text{collided reply} \\ \log_2\left(2^{Q_{fp}} - 1\right), & \text{singleton reply} \\ Q_{fp} - C, & \text{no reply} \end{cases}$$

4) Cancel the current frame when the state $(\hat{N}, S_C, S_R)$ satisfies the cancelation rule and set the new frame size as $L = L^*_{\hat{N}}$.

The optimal frame size $L^*_{\hat{N}}$ and the cancelation rule can be precalculated and stored in the reader. Therefore the IQ algorithm only requires one more table-lookup in every time slot, which barely increases the system complexity. Further, the IQ algorithm does not change the working mechanism of the Q algorithm. Thus the good properties, such as robustness, are still preserved.

Figure 7 shows the simulation results of different algorithms.[7] We observe,

1) Schoute's algorithm needs the longest reading time; but it is also the simplest one, which does not need the frame cancelation.

2) The Q algorithm gives around 10% improvement by introducing the frame cancelation.

3) The IQ algorithm is around 8% better than the Q algorithm with minimum increase of system complexity.

4) The performance of Floerkemeier's algorithm is similar to that of the IQ algorithm for N large. However, it needs to do Bayesian update N times and solving a nonlinear equation in every time slot.

The gap between the performance of IQ algorithm and the bound is caused by the population estimation error. Simulation result shows that a better choice of Q_{fp} and C can narrow this gap. In the best case, when $Q_{fp} = \log_2 N$ and $C \to 0$, IQ algorithm achieves this optimal performance. Optimizing C and Q_{fp} to derive a more accurate $\hat{N}$ belongs to the Population Estimation part.

VI. Summary

In this paper, we optimized reading strategy for Gen-2 RFID systems. The optimal frame length and cancelation rule can be obtained by running an iterative program and can easily be adapted in different RFID anti-collision algorithms. Simulation results show significant improvement.

[7] $\Upsilon = \{2^i \mid i \in Z_+\}$ for all algorithm. In Q and IQ algorithms, $C = 0.3$ while the initial $Q_{fp} = 4$. In Floerkemeier's algorithm, the initial distribution of tag population is uniform between 1 and 40.

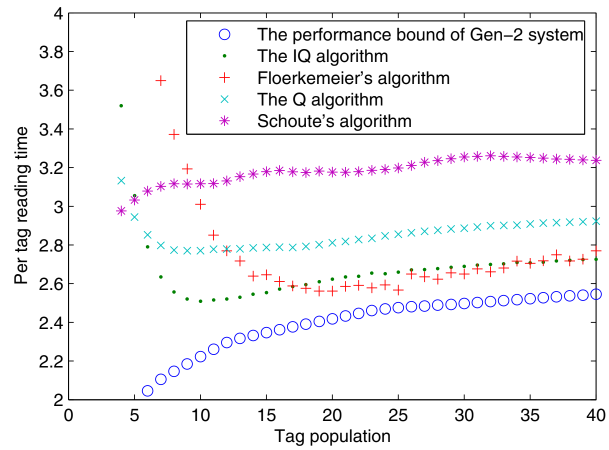

Fig. 7. The performance of different algorithms

References

[1] K. Finkenzeller. RFID handbook - Second Edition. JOHN WILEY & SONS, 2003.
[2] EPCglobal. EPCglobal Class 1 Generation 2 UHF Air Interface Protocol Standard Version 1.0.9, 2005. http://www.epcglobalinc.org/.
[3] International Organization for Standardization. Information technology - RFID for item management - Part 6: Parameters for air interface communications at 860 MHz to 960 MHz, 2004.
[4] Philips Semiconductor. I-CODE1 Label ICs Protocol Air Interface Datasheet, January 2005.
[5] J. Myung, W. Lee, J. Srivastatva, T.K. Shih. Tag-Splitting: Adaptive Collision Arbitration Protocols for RFID Tag Identification. *IEEE Transactions on Parallel and Dsitributed Systems*, Vol 18, No.6, June 2007.
[6] K.W. Chiang, C.Q. Hua and T.S.P. Yum. Prefix-Randomized Query-Tree Protocol for RFID Systems. *IEEE ICC*, 2006
[7] P. Popovski, F.H.P. Fitzek, R. Prasad. Batch Conflict Resolution Algorithm with Progressively Accurate Multiplicity Estimation. *ACM DIALM-POMC* Oct. 2004.
[8] J.Park, M.Y. Chung and T.J. Lee. Identification of RFID Tags in Framed-Slotted ALOHA with Tag Estimation and Binary Splitting, *Communications and Electronics*, 2006
[9] J. Mosely, P.A. Humblet. A Class of Efficient Contention Resolution Algorithms for Multiple Access Channels. *IEEE Transactions of communications, Vol. Com-33, No.2*, Feb. 1985.
[10] F.C.Schoute. Dynamic Frame Length ALOHA. *IEEE Transactions on Communications*, COM-31(4):565-568, Apr 1983.
[11] H. Vogt. Efficient Object Identification with Passive RFID Tags. *First International Conference, PERVASIVE 2002, volume 2414 of Lecture Notes in Computer Science (LNCS)*, pages 98-113, Zurich, Switzerland, August 2002. Springer-Verlag.
[12] C. Floerkemeier. Transmission control scheme for RFID object identification. *Proceedings of the Pervasive Wireless Networking Workshop at IEE PERCOM 2006*, Pisa, Italy, 2006
[13] B. Zhen, M. Kobayashi, M. Shimizu. Framed Aloha for Multiple RFID objects Identification. *IEICE Transactions of comminication. Vol.E88-B, No.3*, March 2005
[14] M. Buettner, D. Wetherall. An Empirical Study of UHF RFID Performance. *MobiCom*, San Francisco, California, USA, 2008
[15] C. Floerkemeier. Bayesian Transmission strategy for Framed ALOHA Based RFID Protocols. *IEEE International Conference on RFID*, Gaylord Texan Resort, Grapevine, TX, USA, March, 2007
[16] M. Kodialam, T. Nandagopal. Fast and Reliable Estimation Schemes in RFID Systems. *MobiCom*, Los Angeles, California, USA, Sepetember, 2006
[17] W. Feller. An Introduction to Probability Theory and Its Apllications, Second edition, Vol I, John Wiley.
[18] J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers.
[19] V. I. Istratescu. Fixed Point Theory, An Introduction. D.Reidel, Holland, 1981