On the Optimal Configuration of Grouping–Based Framed Slotted ALOHA

Young-Beom KIM∗†, Member

SUMMARY In this Letter, we consider several optimization problems associated with the configuration of grouping–based framed slotted ALOHA protocols. Closed-form formulas for determining the optimal values of system parameters such as the process termination time and confidence levels for partitioned groups are presented. Further, we address the maximum group size required for meaningful grouping gain and the effectiveness of the grouping technique in light of signaling overhead.

key words: Anti-collision algorithms, RFID, Framed Slotted ALOHA

1. Introduction

Collisions resulting from simultaneous tag responses are among the key problems affecting the performance of radio frequency identification (RFID) systems. Among the numerous anti-collision protocols proposed thus far, the framed slotted ALOHA (FSA) protocol is the most widely used because of its superior performance and simple implementation [1]. FSA has two variants, i.e., basic framed slotted ALOHA (BFSA) and dynamic framed slotted ALOHA (DFSA). In BFSA, the frame size is fixed throughout the tag identification (ID) processes, while in DFSA, the frame size is adjusted in real time according to the estimated number of tags to keep the system efficiency optimal. Currently, the ISO/IEC 18000-6 Type A and 13.56 MHz ISM band EPC Class 1 Generation 2 (Gen2) standards use the DFSA protocol based on the Q-algorithm [2].

The FSA protocol has two options: FSA-no-muting and FSA-muting [1]. In FSA-no-muting, the tags are not informed by the reader about the outcome of each reading frame. Therefore, each tag transmits its ID once every frame regardless of whether or not the previous transmissions were successful. On the other hand, in FSA-muting, the number of tags decreases after each read frame, since the tags are silenced after ID. When a read frame ends collision free, the reader concludes that all the tags have been identified successfully.

The FSA protocol can be described as follows [1] [3] (specifically, we consider the FSA with no muting). Suppose N passive tags are present in the reader’s interrogation zone (IZ). At the beginning of each FSA frame, the frame size, i.e., the number of time slots (TS), L, for the upcoming FSA frame, is broadcast to the tags. During the reading frame, every tag responds to the reader by randomly selecting a slot to send its ID number. Collision occurs when more than one tag responds to the same slot. The tag reading procedure is repeated frame-to-frame in the same manner until the reader terminates the process.

Basically, RFID readers are required to identify multiple objects as quickly and reliably as possible, with minimal power consumption and computation [3]. These performance requirements can be quantified in terms of the ID delay and confidence level, respectively. The system performance can be greatly improved by properly configuring system parameters such as the frame size and the required number of read frames for the confidence level. Refs. [3] [4] [5] [6] discuss methods to optimally set the system parameters for tag ID processes adopting the FSA protocol. Specially, [5] presents closed-form formulas for the aforementioned parameters under the independence assumption while [6] derives asymptotic results on the optimal frame size for large number of tags.

The tag grouping technique is employed mainly for the following reasons. First, when the number of RFID tags is large, the number of responding tags needs to be limited to maintain system efficiency due to maximum frame size constraints (e.g., no greater than 256, in the Gen2 standard), because the system performance degrades rapidly once the tag population exceeds the frame size [2]. Second, even if the number of tags is not large, the system performance can be enhanced by dividing the tag set into multiple smaller groups and performing the FSA protocol sequentially on a group-by-group basis, utilizing the (collision group) downsizing gain, i.e., the system throughput increases as the group size decreases. Related works on this topic include [7] and [8], where the DFSA-muting protocol is considered.

In this work, we discuss the optimal configuration of grouping–based ID processes adopting the FSA-no-muting. Specifically, the following parameters need to be configured properly: the minimum number of groups (or, equivalently, the maximum group size) required for meaningful grouping gain and the termination time and confidence levels for partitioned groups.

In reality, the number of tags N is unknown to the reader and the reader should estimate it using an appropriate tag estimation function, utilizing the collected statistics at the end of each frame. The tag estimate, namely N, is again utilized in computing the appropriate frame sizes to be used for the following frames. In this scenario, frame size L is adjusted according to N at appropriate time instances.

Manuscript received April 27, 2018.
Manuscript revised July 1, 2018.
† The author is with Dept. of Electronics Eng., Konkuk Univ., Seoul, Korea.
a) E-mail: ybkim@konkuk.ac.kr
DOI: 10.1587/trans.E0.??.1

Copyright © 200x The Institute of Electronics, Information and Communication Engineers
(e.g., on a frame-by-frame basis) during the identification process. Consequently, the DFSA protocol is more realistic than FSA. However, in this work, we assume that the frame size is fixed throughout the ID process in order to facilitate the discussion, because our interest is focused mainly on the optimal configuration of system parameters for a given $N$ (i.e., we consider the situation after $N$ or $\bar{N}$ is determined). Therefore, it is noteworthy that the results herein are not restricted only to the FSA protocol.

### 2. Effects of tag population on system throughput

Consider an FSA process with $N$ tags and frame size $L$. The probability that a tag successfully transmits its ID during a frame, denoted by $\rho(N, L)$, is given by

$$\rho(N, L) = (1 - 1/L)^{N-1}.$$  \hspace{1cm} (1)

Note that $\rho(N, L)$ also represents the normalized throughput, defined as the average number of successful transmissions per time slot (TS) during each frame. Clearly, $\rho(N, L)$ decreases as $N$ increases with $L$ fixed, while $\rho(N, L)$ should increase as $L$ increases with $N$ fixed. We are now interested in the behavior of $\rho(N, L)$ if both $N$ and $L$ increase while $\lambda \triangleq L/N$ remains constant. The system performance strongly depends on proper selection of $\lambda$. In fact, the optimal value of $\lambda$ depends on the muting options, i.e., $\lambda_{opt} = 1$ and $1/\ln 2$ for DFSA-muting and DFSA-no-muting, respectively [2] [6]. Note that $L$ should actually be set to $\lceil LN \rceil$ ($\lfloor x \rfloor$ represents the integer closest to $x$). However, we retain the expression $L = \lambda N$ for simplicity, and assume $\lambda \geq 1$ throughout. With this problem setting, $\rho(N, L)$ is a function of $N$ only; thus, we simply write $\rho(N)$ (obviously, $\rho(1) = 1$ for $L = 1, 2, \ldots$). Setting $L = \lambda N$ in (1), we have

$$\rho(N) = (1 - 1/(\lambda N))^{N-1} \text{ for } N = 2, 3, \ldots,$$

with $\rho(1) = 1$. The quantity $\rho(N)$ has the following properties:

**Proposition 1:** For $N = 1, 2, \ldots$,

- (P1) $\rho(N) > \rho(N + 1)$;
- (P2) $\rho(N)/\rho(N + 1) > \rho(N + 1)/\rho(N + 2)$;
- (P3) $\rho(N) - \rho(N + 1) > \rho(N + 1) - \rho(N + 2)$.

Proof: (P1) Because $\rho(1) = \rho(2) = 1 - 1/(2.1)$, it suffices to show that the function $\rho(x)$ is strictly decreasing for $x \geq 2$. We first find that $\ln \rho(x)^{x'} = \rho'(x)/\rho(x) = \ln(1 - 1/Lx) + x - 1/(Lx - 1)$. Noting that $0 < 1 - 1/Lx < 1$, from the well known inequality $\ln \theta < \theta - 1$ for $0 < \theta < 1$, we obtain $\rho'(x)/\rho(x) < -1/Lx + (x - 1)/(Lx(x - 1)) = -1/(Lx(x - 1)) < 0$. Because $\rho(x) > 0$, we can conclude that $\rho'(x) = \rho(x)(\ln \rho(x))' < 0$.

(P2) Similarly, it suffices to show that the function $f(x) \triangleq \rho(x)/\rho(x + 1)$ is strictly decreasing, i.e., $(\ln f(x))' < 0$, for $x \geq 1$ (the detailed proof is omitted).

(P3) From (P1) and (P2), $\rho(N) - \rho(N + 1) = \rho(N + 1)(\rho(N)/\rho(N + 1) - 1) = \rho(N + 1)(\rho(N + 1)/\rho(N + 2) - 1) = \rho(N + 1)(\rho(N + 1)/\rho(N + 2) - 1) = \rho(N + 1) - \rho(N + 2)$.

- (P1) describes the (collision group) down-sizing gain, i.e., it indicates that $\rho(N)$ increases as $N$ decreases (with $\lambda$ fixed). On the other hand, (P2) and (P3) imply that the gain diminishes as $N$ increases (i.e., $\rho(N)/\rho(N + 1) \to 1$ and $\rho(N) - \rho(N + 1) \to 0$ as $N \to \infty$).

### 3. System parameter configuration for FSA-no-muting ID processes

We assume that each tag in the reader’s IZ has been assigned a sequential ID number, say, from 1 to $N$. Let $E_i$ and $E_i'$ denote the events in which all the tags are identified through the end of the $i^{th}$ reading frame and the tag with a sequential number $i$ is identified throughout the $t$ reading frames, respectively. Because $E_i = \bigcap_{j=1}^{N} E_i'$, the probability of successful ID through $t$ consecutive reading frames, i.e., $P_t$, is given by

$$P_t \triangleq P[E_t] = P \left[ \bigcap_{j=1}^{N} E_i' \right].$$  \hspace{1cm} (2)

The process terminates when the reliability constraint

$$P_t \geq \alpha$$  \hspace{1cm} (3)

can be met for a preassigned confidence level $\alpha$, obeying the requirement that an RFID reader should basically identify all the tags in its range.

The minimum value of $t$ satisfying (3) is referred to as the termination time (denoted by $T$); thus, $T$ can be expressed as

$$T \triangleq \min\{t = 1, 2, \ldots: P_t \geq \alpha\}.$$  \hspace{1cm} (4)

If we assume that any additional delay components (e.g., signaling overhead to maintain the TDMA frame structure) are negligible, the ID delay $\tau$ can be simply obtained by $\tau = LN \times T$.

On the other hand, no closed-form expression for $P_t$ exists. If we assume that $\{E_i', i = 1, 2, \ldots, N_k\}$ are mutually independent in (2), then

$$P_t^{*} \triangleq \prod_{i=1}^{N} P[E_i'] = (1 - \rho(N))^N.$$  \hspace{1cm} (5)

### 4. Grouping-based ID procedure Design

Consider the case of partitioning a tag set into $K$ groups, namely, $G_1, G_2, \ldots, G_K$, with group sizes $|G_k| = N_k$, $k = 1, 2, \ldots, K$. For this purpose, at the beginning of each frame, the reader needs to broadcast $K$ (or alternatively the group size) for grouping, e.g., using a SELECT message in the Gen2 standard. A tag can find its own group via modulo $K$ operation with either a tag ID or random number. Considering the fact that a considerable number of empty groups can exist after partitioning (on average, about 36.8% of groups are empty for large $K = N$), we reconstruct the ID process as follows. The process involves two consecutive
stages, namely, a probing stage and an ID stage. In the probing stage, empty or singleton groups (i.e., \(|G_k| = 0 \text{ or } 1\)) are identified, using a trial frame of size \(K\). Each tag in \(G_k\) transmits its ID in the \(k^{th}\) TS within the frame. Empty and singleton groups are skipped in the next phase of the ID process (note that tags in singleton groups need no further transmission). In the ID stage, the ID process continues sequentially on a group-by-group basis, skipping over empty and singleton groups.

In the ID stage, for each group, say, \(G_k\), it is necessary to determine the optimal frame size, termination time, and confidence level such that the ID delay can be minimized under constraint (3). While \(L_k\) can be readily determined using \(L_k = \lfloor A_{\text{opt}} \cdot N_k \rfloor\) from the estimated number of tags in \(G_k\), say, \(N_k\), it is not straightforward to find proper values of \(T_k\) and \(\alpha_k\).

First, the exact value of \(T_k\) can be obtained using the Markov chain (MC) method presented in [3]. However, the computation therein is extremely complicated and sometimes intractable for large number of tags due to computational overflows. To circumvent this difficulty, we use \(\mathcal{P}_1^i\) in (5), instead of \(\mathcal{P}_1\) in (2). Replacing \(\mathcal{P}_1\) in (4) with \(\mathcal{P}_1^i\) (also \(N\) with \(N_k\) and \(\alpha\) with \(\alpha_k\)) yields

\[
T_k = \lfloor \ln(1 - \alpha_k^{1/N_k})/\ln(1 - \rho(N_k)) \rfloor, \tag{6}
\]

where the ceiling operator \(\lceil x \rceil\) represents the smallest integer that is no smaller than \(x\).

Now let us consider the problem of optimizing \(\alpha_k\) (denoted by \(\alpha_k^*\)). Note that \((\alpha_1^*, \ldots, \alpha_k^*)\) is an element of the set

\[
S = \{(x_1, \ldots, x_k) \in (\alpha, 1)^k : x_1x_2 \cdots x_k = \alpha\}. \tag{7}
\]

Suppose that \(K = 2\). Using (6), the ID delay \(\tau\) is given by

\[
\tau = \sum_{k=1}^{2} \lambda N_k \lfloor \ln(1 - \alpha_k^{1/N_k})/\ln(1 - \rho(N_k)) \rfloor. \tag{8}
\]

To find \(\alpha_k^*\) minimizing \(\tau\), we define a function \(d(\alpha_1)\) as

\[
d(\alpha_1) = \sum_{k=1}^{2} \lambda N_k \ln(1 - \alpha_k^{1/N_k})/\ln(1 - \rho(N_k))\]

by dropping the operator \(\lceil \cdot \rceil\). The value of \(\alpha_1\) minimizing \(d(\alpha_1)\), denoted by \(\alpha_{\text{min}}\), can be obtained numerically. Fig. 1 describes \(d(\alpha_1)\) for \((N, N_1) = (32, 8)\) and \((N, N_1) = (64, 8)\), respectively, when \(\alpha = 0.9\) and \(\lambda = 1.5\). The values of \((\alpha_{\text{min}}, \alpha_k^{1/N})\) are given by \((0.9751, 0.9740)\) and \((0.9877, 0.9869)\), respectively, indicating that \(\alpha_{\text{min}} \approx \alpha_k^{1/N}\). Through extensive calculations using MATLAB, the observations can be summarized as follows.

(O1) When dividing \(N\) tags into two groups of sizes \(N_1\) and \(N_2\), \(\alpha_k^* = \alpha_k^{N_k/N}\) for \(k = 1, 2\), in the sense that the ID delay is minimized.

We can observe from (7) that the ID delay indeed decreases by grouping according to (O1) because \(\alpha_k^{1/N_k} = \alpha_k^{N_k/N} = \alpha_k^{1/N}\) for \(\alpha_k = \alpha^{N_k/N}\) and \(\rho(N_k) > \rho(N)\) by (P1).

We can further extend the result (O1) to the cases: \(K > 2\). For \(K = 3\), we can obtain three groups, say, \(G_1\), \(G_2\), and \(G_3\) (of sizes \(N_1\), \(N_2\), and \(N_3\), respectively), through two consecutive partitioning steps; first, the initial tag group is divided into two groups \(G_1\) and \(G_{11}\) (of sizes \(N_1\) and \(N_2 + N_3\), respectively), and, next, \(G_{11}\) is divided into \(G_2\) and \(G_3\) (of sizes \(N_2\) and \(N_3\), respectively) again. Applying (O1) to each step, we get \(\alpha_{1}^* = \alpha_1^{N_1/N}, \alpha_{11}^* = \alpha_{11}^{(N_2+N_3)/N}, \alpha_{111}^* = \alpha_{111}^{N_2/(N_2+N_3)} = \alpha_2^{N_2/N}, \) and \(\alpha_{111}^* = \alpha_{111}^{N_3/(N_2+N_3)} = \alpha_3^{N_3/N}\). Similarly, continuing this procedure for \(K > 3\), (O1) can be rewritten as: when dividing \(N\) tags into \(K\) groups, \(\alpha_k^* = \alpha_k^{\gamma_k}\), where \(\gamma_k = N_k/N\) for \(k = 1, 2, \ldots, K\).

5. Numerical results and discussion

Fig. 3 compares the ID delays computed using the formula (6) with the exact ones (computed via the MC method), for \(\alpha_1 = n/N, n = 1, 2, \ldots, N\), where \(\alpha = 0.99, \lambda = 1.4427, \) and \(K = 2\) \((\alpha = 0.9, \lambda = 1.5), \) and \(\alpha = 0.0\) (‘no grouping’). The improvement is not significant (less than 1%) for average group size \(\bar{g} \leq 32\) \((\bar{g} = N/K)\), while the delay is decreased by 6.49% at \(\bar{g} = 8\).

This result equally holds for the other values of \(N\) aforementioned and indicates that \(K\) should be very large (at least 128
for } N = 1024 \text{) to achieve meaningful improvement. Considering the fact that signaling overhead for grouping usually grows in proportion to } K, \text{ the “grouping” technique can be very inefficient for dense tag population. This phenomenon can be explained in the context of (P2) and (P3) and appears in the FSA-muting protocol as well, but less severely.}

6. Conclusion

In this work, we considered grouping-based FSA protocol with no muting and established several formulas for opti-