Slot-Wise Maximum Likelihood Estimation of the Tag Population Size in FSA Protocols

B. Knerr, M. Holzer, C.Angerer, *Student Member IEEE*,
and M. Rupp, *Senior Member IEEE*

***Abstract*—Framed Slotted Aloha (FSA) is a popular anti-collision technique in state-of-the-art RF-ID systems, as in ISO/IEC CD 18000-6 for 900MHz or the EPCglobal HF Gen 2 draft for 13.56MHz. In many applications the number of tags entering and leaving the detection range of the reader is subject to a strong fluctuation and usually unknown. The current number of tags in the field is a crucial parameter to operate the FSA anti-collision in an optimal manner. Therefore, a lot of effort is spent on the estimation of this parameter and a range of different estimation techniques exist. The contributions of this paper are: 1) a closed formula for the probability of any observed event defined by the number of *empty*, *singleton*, and *collision* slots in the observed frame is developed and empirically verified. 2) This formula is then modified to compute the probability for partly observed frames as well which is of great interest as the referred standards allow for the in-frame adjustment of the frame size without quitting the interrogation round. 3) Then, a maximum likelihood estimator is formulated to yield the estimated number of tags on a slot-wise basis. 4) Its superior estimation performance is compared to the known best estimators over the complete parameter set. While its performance is strongly superior compared to Schoute's estimate, compared to Vogt's MSE estimator only marginally improvement is obtained[1].**

***Index Terms*—RF-ID, framed slotted aloha, maximum likelihood estimator, anti-collision protocol.**

I. Introduction

ANY information processing system performs one or more of the tasks of acquiring, interpreting, retaining, and distributing data. Whenever the communication channel is shared, the tasks acquisition and distribution face the risk of possible contention of simultaneously transmitted signals and thus degrading the communication and lowering throughput. A special case is the 1-to-n communication, in which the entity that serves as master of the communication link encounters a yet unknown number of n unidentified clients. Under these circumstances stochastic anti-collision techniques can be utilised that enable a time dispersion of responding clients (n-to-1) in a controlled manner. Radio Frequency Identification (RF-ID) systems are exemplary for this kind of communication scenario and their passive category gained recently enormous interest as it is generally considered to have dramatic effect on economical, social, and intellectual life [2].

For RF-ID systems two stochastic anti-collision methods are commonly used in state-of-the-art standards: binary tree search (e.g. the ISO/IEC 15693, ISO/IEC CD 18000-3, ISO/IEC CD 18000-6 Type B), and Framed Slotted ALOHA in Draft EPC Global HF Generation 2 and Type C of ISO/IEC 18000-6 for UHF.

In Framed Slotted ALOHA any client that tries to answer a request submitted by the communication master with a data packet chooses at random a time slot of a frame [3]. The frame length, i.e. the number of available slots, is the parameter that determines the achievable throughput of the system. In RF-ID systems the number of clients (or tags) in the powering field is usually unknown and is often subject to a strong fluctuation. Consider a harbour cargo gate through which trucks loaded with tagged objects of differing size and number enter and leave. In such an environment the number of tags in the field may vary from a few dozen in one instant (e.g. one second) to a few thousands in the next. The transmission control strategy has to estimate and adjust the frame size, which determines the broadcast probability of the tags, to the optimal value.

The contribution of this paper is twofold. Although a number of algorithms and techniques exist that estimate the number of tags in the field and dynamically adjust the frame size to the current scenario, a maximum likelihood (ML) estimator has not yet been formulated. In this work we close this gap in literature and derive such an ML estimator. Secondly, the in-frame adjustment of the frame size has only been considered recently by Flörkemeier [4] with a Bayesian belief estimator. All other existing approaches estimate the number of tags depending on the past, completely observed, frames, then set the frame size anew and process the current frame accordingly. However, the EPC Global HF Class Gen 2 and the ISO/IEC CD 18000-6 standards explicitly favour an in-frame adjustment of the frame size with the `QUERYADJUST` command. Thus, it is possible to observe only a small part of a frame and to react immediately, if the number of collisions or empty slots exceeds a certain value. With this feature it is not necessary to quit the current interrogation round and to retransmit the request. The maximum likelihood formulation is hence modified towards partly observed frames on a slot-

Paper approved by A. Zanella, the Editor for Wireless Systems of the IEEE Communications Society. Manuscript received November 1, 2008; revised April 20, 2009.

The authors are with the Institute of Communications and Radio-Frequency Engineering at the Vienna University of Technology. (e-mail: mrupp@nt.tuwien.ac.at).

This work has been funded by the Christian Doppler Laboratory for Design Methodology of Signal Processing Algorithms.

Digital Object Identifier 10.1109/TCOMM.2010.080571

[1]Some parts of this journal version have been published in [1], B.Knerr, M.Holzer, C.Angerer, M.Rupp, "Slot-by-slot Maximum Likelihood Estimation of Tag Populations in Framed Slotted Aloha Protocols," International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2008), Edinburgh, UK, pp. 303–308, June 2008.

0090-6778/09$25.00 © 2010 IEEE

Name	Explanation
F	number of slots
N	inspected number of slots, $N \leq F$
n	number of tags
$\widehat{n}$	estimated number of tags
r	fill level of r tags in a given slot, occupancy number
m_r	number of slots with fill level r
m_0	number of empty slots
m_1	number of singleton slots
m_c	$= m_{\geq 2}$ number of slots with collision, i.e., $r \geq 2$
$E(\chi)$	expectation value of a random variable χ

TABLE I
TERMS AND VARIABLES.

by-slot basis.

Relevant related work will be addressed in Section II. The following Section III of this article introduces necessary terms and definitions used for Framed Slotted ALOHA scenarios and the classical formulation of the ALOHA anti-collision as *occupancy problem*. Section III also deals with the probability equations and derives a closed formula for the probability of an event for a partly observed frame. In Section IV the formula is analysed and the maximum likelihood estimator is formulated. Section V compares the performance of the new ML estimator with the minimum squared error (MSE) estimator of Vogt [5], which has been modified by us to be applicable to partly observed frames, before the paper will be concluded by Section VI.

II. Related Work

Under the assumption that the frame size is chosen such, that in case the number of clients that transmit in each frame is Poisson distributed with unit mean, Schoute developed a backlog estimation technique for FSA [3]. The estimated number of clients after observing a complete frame is then estimated to be $\widehat{n} = 2.39 m_c$, where m_c is the number of observed slots with collisions. The most important parameters in this context are listed in Table I for convenience of the reader. Evidently, the observed values for empty slots, m_0, and singleton slots, m_1, are neglected in this scheme, as well as the assumption of the Poisson distribution may not be true in all applications. In this paper its modification to partly observed frames (N out of F slots) leads to the following estimate of the number of clients

$$\widehat{n}(N) = 2.39\, m_c\, \frac{F}{N}, \tag{1}$$

which is a straightforward extension to Schoute's formula [3] and will be compared to in Section V.

Vogt introduced an MSE estimator and analysed its behaviour for completely observed frames. As it has been shown by us [6], [1], his approach can be easily modified towards partly observed frames and is, as such, a promising candidate for modern RF-ID standards. Moreover, Vogt's work clarifies the superior performance of the MSE estimator in comparison with the Q algorithm defined in ISO/IEC 18000-6 C. This algorithm keeps a representation of its current frame size $F_{cur} = 2^{\lfloor C \rceil}, C \in \mathbb{R}^+$, since the standard requires frame sizes in powers of two. With any slot being observed, the exponent C is increased by a constant β, when a collision occurred, or diminished by β for an empty slot. A singleton slot does not alter the exponent. After any slot the algorithm checks whether the modified C_{mod} translates into a new frame size $F_{new} = 2^{\lfloor C_{mod} \rceil} \neq F_{cur}$. In this case the new frame size is immediately broadcasted to any client still in the current interrogation round forcing them recalculate the slot for their response accordingly. The standard does not state how to compute a suitable β, but at least comments on a reasonable range $1.07 \leq 2^{\beta} \leq 1.41$.

Most recently, Lee et al. [7] published an improved Q$^+$-algorithm with optimised parameters β_c for detected collisions and β_e for detected empty slots, with $2^{\beta_c} = 0.35$ and $2^{\beta_e} = (e-2)2^{\beta_c} = 0.25$ as optimal ratio. In a recent paper the minor performance of the Q$^+$-algorithm compared to the slot-by-slot MSE estimator has been shown [6].

Krohn et al. [8] recently proposed an approach which presents a fast technique to estimate the number of tags present in the read range. They state explicitly to optimise rather for fast detection than for accuracy of the estimate and base their work on the assumption that there are empty and occupied slots only. From this perspective, our approach exhibits a contrary objective with the identification of the achievable lower bound for the estimation error.

Recently, Flörkemeier [4] published a Bayesian slot-by-slot estimator for the tag population. This approach utilises exponential generating functions to create a probability formula for all distinct events $\langle m_0, m_1, m_c \rangle$ under the parameters N observed slots out of F slots in a complete frame and n tags being distributed in them. Unfortunately, he could not find a closed form for the coefficient formula of the exponential generating functions, which avoids to *count* all permutations that generate a given event $\langle m_0, m_1, m_c \rangle$ with n tags. Hence, his approach is only applicable for scenarios, in which the number of tags n and the number of collisions m_c is very low, as the number of required multiplications in this counting problem grows exponentially with both of them.

III. Probabilities in FSA

As RF-ID systems serve as example application due to their outstanding importance at present, we will use expressions typically used in such systems. However, the following considerations are of course true for any system applying FSA.

A request issuing master is named *interrogator* or just the *reader* as shown in Figure 1. The clients responding to the reader's request are called *tags*. After issuing a request including the current frame size, the reader waits for a specified time interval, i.e. the length of the frame, which is divided into a number of F slots, separated by additional signalling issued by the reader. At the beginning of the frame any tag in the field randomly generates a number between 1 and the frame size F that determines the slot in which it tries to respond. Whenever more than one tag use the same slot for their responses, a collision occurs and typically the data sent by the involved tags are corrupted. In general, it is not possible to deduce the exact number of tags that caused the collision. The number of present tags in the field, the tag population, is denoted by n.

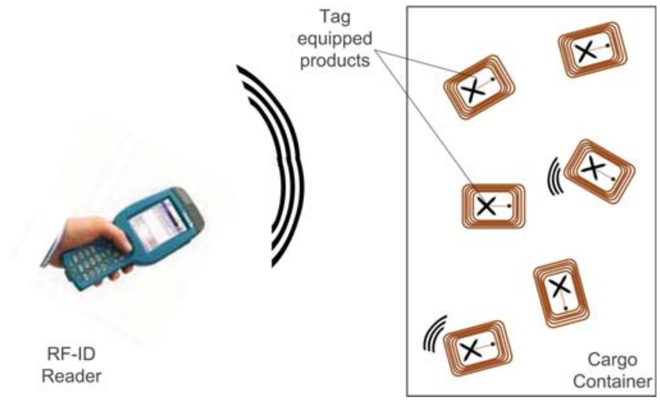

Fig. 1. RF-ID Reader and tagged products in a cargo container.

The allocation of tags to time slots within a frame can be formulated as an occupancy problem [9] that can be found in a broad range of applications [10]. In these problems balls are randomly allocated to a number of bins. Balls correspond to tags as bins correspond to slots.

Having F slots available and n tags in the field, the fill level of r tags in a given slot is described by a binomial distribution:

$$B_{n,\frac{1}{F}}(r) = \binom{n}{r}\left(\frac{1}{F}\right)^r\left(1-\frac{1}{F}\right)^{n-r}. \tag{2}$$

Let $\mathcal{X}_r$ be a random variable indicating the number of slots with fill level r. The expected number of slots $E\left(\mathcal{X}_1 = m_1\right)$ with just a single tag response, i.e. fill level $r = 1$, is of major interest to measure the throughput $T = \frac{E(\mathcal{X}_1)}{F}$ of the system. By simple calculus it has been shown by Schoute [3] that T is maximised if the frame length F equals the tag population size n.

The number r of tags in a particular slot is called its *occupancy number*. Since the distribution (2) applies for all slots of a frame, the expected number of slots $E(\mathcal{X}_r)$ with *occupancy number* r is given by:

$$\begin{aligned} E\left(\mathcal{X}_r\right) &= NB_{n,\frac{1}{F}}(r) \\ &= N\binom{n}{r}\left(\frac{1}{F}\right)^r\left(1-\frac{1}{F}\right)^{n-r}, \end{aligned} \tag{3}$$

where N is the number of observed slots ($N = F$ for a completely observed frame). As already indicated it is only possible for any observed slot, whether it features the fill levels $r = 0, 1, \geq 2$, thus having for N out of F observed slots: m_0 *empty* slots, m_1 *singleton* slots, and $m_{\geq 2} = m_c$ *collision* slots.

The probability to observe exactly $\mathcal{X}_r = m_r$ slots with the fill level r, given the frame size F and the tag number n when observing all $N = F$ slots is:

$$\begin{aligned} P(\mathcal{X}_r = m_r; F, n) = {} & \\ & \underbrace{\binom{F}{m_r}}_{A}\left[\underbrace{\prod_{l=0}^{m_r-1}\binom{n-lr}{r}\left(\frac{1}{F}\right)^r}_{B}\right] \times \\ & \sum_{k=0}^{S}(-1)^k\underbrace{\binom{F-m_r}{k}}_{C} \\ & \times\left[\underbrace{\prod_{j=0}^{k-1}\binom{n-rm_r-jr}{r}\left(\frac{1}{F}\right)^r}_{D}\right] \\ & \times\underbrace{\left(\frac{F-m_r-k}{F}\right)^{n-rm_r-kr}}_{E} \end{aligned} \tag{4}$$

This closed formula (4) has been published by Vogt [5] for the random variable $\mathcal{X}_r$. Herein, the distribution of $\mathcal{X}_r$ depends on the probability in (4) with $S = \min\binom{F-m_r,}{\lfloor (n-rm_r)/r \rfloor}$, as it is explained in the following paragraph.
The underbraces denote the subterms to facilitate later references. It is necessary to thoroughly explain the subtleties of this formula to comprehend the new considerations. In the experiment for (4), n tags shall hit arbitrary m_r out of F slots exactly r times, and the remaining $F - m_r$ slots **not** r times.

- Term A chooses m_r out of F slots arbitrarily.
- Term B distributes in any of these slots rm_r arbitrary tags out of n candidates. The following large sum is concerned with the distribution of the remaining $n - rm_r$ tags in the S slots.
- The alternating sum represents the *inclusion-exclusion* principle that can be applied here: For $k = 0$ terms C and D are 1 and term E gives the probability of all remaining $n - m_r r$ tags being distributed arbitrarily into the remaining $F - m_r$ slots. Then we have inevitably considered also all possibilities with **at least** one slot having fill level r, which is forbidden as none of the $F - m_r$ should contain exactly r tags.
- And even more important, term E considers these forbidden events with different permutations of the $n - rm_r$ tags. To successively diminish this error, the forbidden combinations have to be subtracted.
- For $k = 1$ term C chooses one slot, which is then filled with r tags (term D), and all remaining $n - rm_r - r$ tags are again distributed arbitrarily into the remaining $F - m_r - 1$ slots. Then a part of the former erratically considered possibilities has been corrected (those having exactly one forbidden slot with fill level r).
- Now the arbitrary distribution of the remaining $n - rm_r - r$ considers too many possibilities for the forbidden slots ≥ 2. For $k = 2$ these are added again. And so forth for $k = 3 \ldots S$ until no tags are available anymore ($S = \lfloor (n - rm_r)/r \rfloor$) or all remaining slots have been considered ($S = F - m_r$). In both cases the *problematic* term E does not introduce an error anymore because it equals 0 when $S = F - m_r$ or 1 when $S = \lfloor (n - m_r)/r \rfloor$.

The explicit forms for fill levels $r = 0$ and $r = 1$ are straightforward, even if we observe only N out of F slots. The probabilities for the fill levels $r = 0$ and $r = 1$ are given

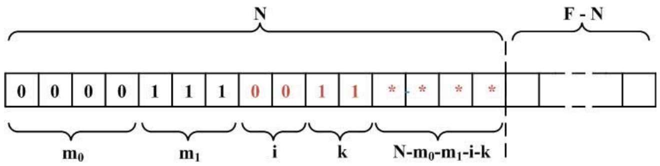

Fig. 2. Slots and their fill levels $r = 0, 1, \geq 2$ depicting indices of (7).

respectively:

$$P(\mathcal{X}_0 = m_0; N, F, n) = \binom{N}{m_0} \times \sum_{i=0}^{\min\binom{N-m_0,}{n}} (-1)^i \binom{N-m_0}{i} \underbrace{\left(\frac{F-m_0-i}{F}\right)^n}_{E_{m_0}} , \quad (5)$$

$$P(\mathcal{X}_1 = m_1; N, F, n) = \underbrace{\binom{N}{m_1}}_{A_{m_1}} \left[\underbrace{\frac{n!}{(n-m_1)!} \left(\frac{1}{F}\right)^{m_1}}_{B_{m_1}} \right] \times \sum_{i=0}^{\min\binom{N-m_1,}{n-m_1}} (-1)^i \underbrace{\binom{N-m_1}{i}}_{C_{m_1}} \left[\underbrace{\frac{(n-m_1)!}{(n-m_1-i)!} \left(\frac{1}{F}\right)^i}_{D_{m_1}} \right] \times \underbrace{\left(\frac{F-m_1-i}{F}\right)^{n-m_1-i}}_{E_{m_1}} . \quad (6)$$

This rather complex equation yields for any alleged number of slots with a certain fill level r, how probable the outcome m_r is under a given parameter scenario of n tags being distributed in F slots. We are interested in an event with more information contained. We are looking for the formula yielding the probability $P(\mathcal{X}; N, F, n)$ for the event $\mathcal{X} = \langle m_0, m_1, m_c \rangle$ characterised by "m_0 empty **and** m_1 singleton **and** m_c collision slots" in $N = m_0 + m_1 + m_c$ observed slots with the parameter scenario of n tags being distributed in F slots. Apparently, this event is characterised by the two versions of (4) for $r = 0$ and $r = 1$, with $r \geq 2$ being implicit, since $m_c = N - m_0 - m_1$ is hence predetermined. Equation (5) represents the case $r = 0$ and (6) shows the case $r = 1$. Until here, only minor modifications have been performed to consider only N out of F observed slots.

$$P(\mathcal{X} = \langle m_0, m_1, m_c \rangle; N, F, n) = \binom{N}{m_0} \sum_{i=0}^{\min\binom{N-m_0-m_1,}{n-m_1}} (-1)^i \binom{N-m_0}{i} \underbrace{\left(\frac{F-m_0-m_1-i}{F}\right)^{n-m_1}}_{\widehat{E}_{m_0}} \times \underbrace{\binom{N-m_0-i}{m_1}}_{\widehat{A}_{m_1}} \underbrace{\left[\frac{n!}{(n-m_1)!} \left(\frac{1}{F}\right)^{m_1}\right]}_{\widehat{B}_{m_1}} \times I(\langle m_0, m_1, m_c \rangle; N, F, n, i). \quad (7)$$

$$I(\langle m_0, m_1, m_c \rangle; N, F, n, i) = \sum_{k=0}^{\min\binom{N-m_0-i-m_1,}{n-m_1}} (-1)^k \underbrace{\binom{N-m_0-i-m_1}{k}}_{\widehat{C}_{m_1}} \times \left[\underbrace{\frac{(n-m_1)!}{(n-m_1-k)!} \left(\frac{1}{F-m_0-i-m_1}\right)^k}_{\widehat{D}_{m_1}} \right] \underbrace{\left(\frac{F-m_0-i-m_1-k}{F-m_0-i-m_1}\right)^{n-m_1-k}}_{\widehat{E}_{m_1}} . \quad (8)$$

With (5) and (6) it is possible to develop a new closed formula for the composed event of "m_0 and m_1 and (implicitly) m_c slots" presented in (7) and (8). Note, the splitting in *two* functions has just been done for clarity of the presentation. Consider the term E_{m_0} in (5) that is concerned with the arbitrary and iteratively erroneous distribution of the remaining n tags. With precisely these tags the event m_1 described in (6) has to be composed. Hence, we replace the term E_{m0} with term $\widehat{E}_{m_0}$ and insert thereafter (6) leading to the Equation (7) and its inner sum (8).

For the index transformations and further necessary modifications, consider Figure 2, in which the scenario for a given event has been captured. For clarity the slots with equal fill levels are grouped together in this figure. Instead of distributing all n tags into the remaining $F - m_0 - i$ slots in term E_{m_0} in (5), we have to put aside m_1 tags to be filled into m_1 slots to complete the event. Thus, term $\widehat{E}_{m_0}$ fills only $n - m_1$ tags into $F - m_0 - m_1 - i$ slots. Term $\widehat{A}_{m_1}$ chooses those m_1 slots from the remaining $N - m_0$ slots and fills in each one exactly one tag via term $\widehat{B}_{m_1}$. Here, we enter the inner sum (8) originating from (6). This sum distributes precisely one tag into any of k slots out of the now remaining $N - m_0 - i - m_1$ slots (term $\widehat{C}_{m_1}$ and $\widehat{D}_{m_1}$). The last term $\widehat{E}_{m_1}$ is concerned with all remaining tags distributed arbitrarily among the remaining $F - m_0 - i - m_1 - k$ slots. In this large formula there exist two terms that miscount the partial possibilities, $\widehat{E}_{m_0}$ and $\widehat{E}_{m_1}$, hence leading to the cascaded sums in the same manner as described for (4). The correctness of this formula has then been verified empirically. In opposition the Flörkemeier's approach [4], which requires the evaluation of a

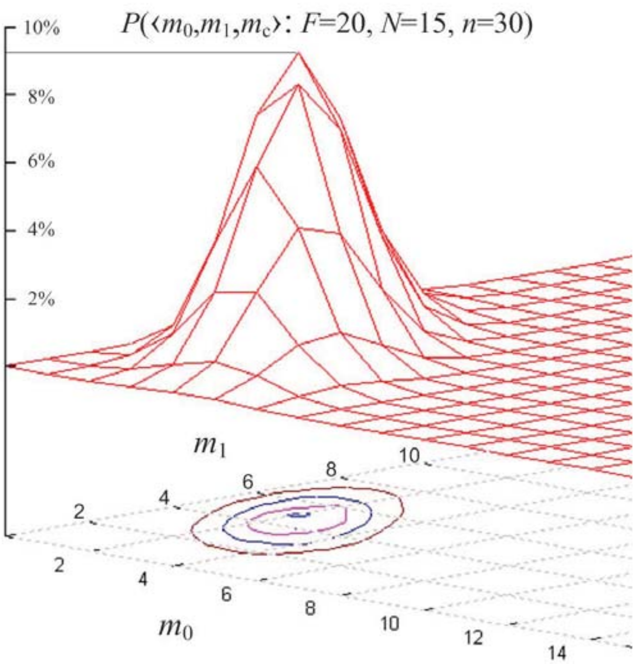

Fig. 3. Probabilities over event space (m_0, m_1) for parameters $N = 15$, $F = 20$, $n = 30$.

number of terms that increases exponentially with the number of assumed tags n, this equation requires only $\frac{m_c}{2}(m_c+1)$ terms to be computed. In other words, the number of terms to be calculated grows quadratically with the number of observed collisions.

IV. Maximum Likelihood Estimator for FSA

In Figure 3 an example for the developed probability function (7) is depicted over the event plane (m_0, m_1) for the parameters $F = 20, N = 15$ and a given tag population $n = 30$. For any permitted parameter set $(m_0 \neq N \wedge m_c \neq N \wedge n \geq m_1 + 2m_c \wedge F \geq N \geq 1 \wedge m_0 + m_1 + m_c = N)$ this function always reveals a *sugar loaf*-like shape. Since in our scenario the parameter n is not known, but the outcome of this random experiment has been observed, given as event $\mathcal{X} = \langle m_0, m_1, m_c\rangle$, we may assume that the unknown parameter n is that particular n, which makes the observed event most likely. Hence, the probability function for the observed event is analysed over different values of n, looking for that value n, which yields maximum probability. The plots in Figure 4 and 5 show the outline of the probability for given events over the parameter n. The plots in Figure 4 are subject to three different events $\mathcal{X} \in \{\langle 2,2,1\rangle, \langle 0,3,2\rangle, \langle 1,2,2\rangle\}$ for a low number of $N = 5$ observed slots out of $F = 32$ slots. The plots in Figure 5 outline the probabilities for a higher number of observed slots $N = 20$ out of $F = 32$ slots for three different events $\mathcal{X} \in \{\langle 4,6,10\rangle, \langle 8,4,8\rangle, \langle 1,9,10\rangle\}$. It is important to notice the well-behaved shape of this function. Except from the border case $m_c = N$ for which the maximum lies with $n = \infty$, a single well defined maximum can be identified (for $m_0 = N$ the maximum lies at $n = 0$). This

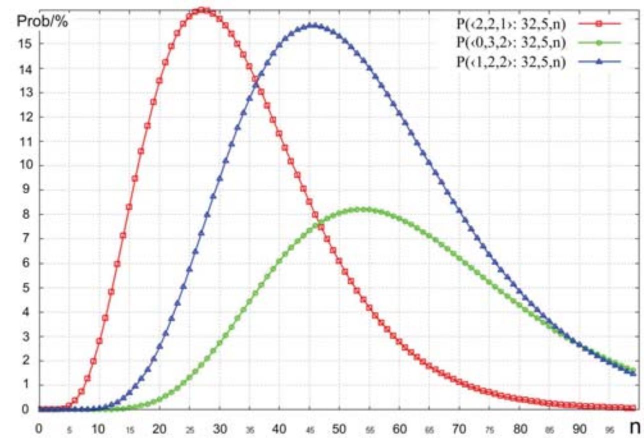

Fig. 4. Probability for given events over possible values for the tag population n, $N = 5$, $F = 32$.

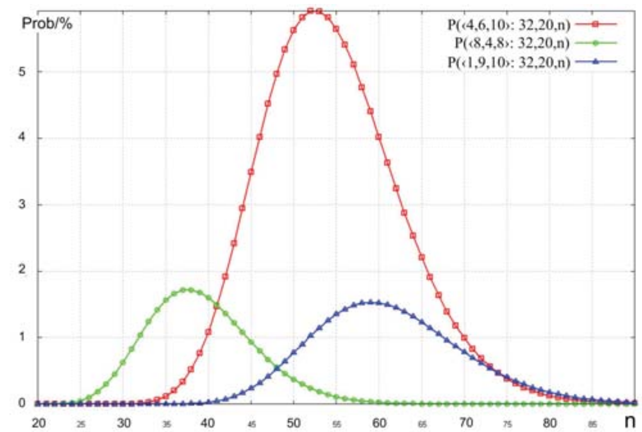

Fig. 5. Probability for given events over possible values for tag population n, $N = 20$, $F = 32$.

maximum indicates, as stated before, that value n for which the observed event is most likely. Hence, we develop a slot-by-slot tag population estimator by performing a maximum search. Unfortunately, the derivation $\frac{\partial P(\mathcal{X};N,F,n)}{\partial n}$ exhibits an even more complex structure. Thus, in order to obtain the maximum analytically, an approximation technique would have to be utilised to find the zero-crossing $\frac{\partial P(\mathcal{X};N,F,n)}{\partial n} = 0$. In comparison with iterative techniques, for instance Newton's method, a gradient search for the maximum is then still less computationally expensive.

A two staged gradient search initially searches for the first occurrence of a negative gradient with step width $n(i) = (m_1 + 2m_c)\, 2^i, i = 1, 2, \ldots$. When the first negative slope is detected at i, the maximum lies between $n(i-1)$ and $n(i)$. Within this interval a binary search implements the second stage of the gradient search. It analyses the gradient in the middle of the nested interval and follows the positive slope to identify the next interval of half the size. As any binary search this procedure requires only $\log_2(n(i) - n(i-1))$ computations of the probability function in the worst case. The identified maximum yields the estimated tag population size of the maximum likelihood estimator:

$$\widehat{n}_{\text{ML}} = \arg\max_{n} P\left(\mathcal{X} = \langle m_0, m_1, m_c \rangle; N, F, n\right). \quad (9)$$

It has to be mentioned that the referred standards momentarily only allow for a frame size adjustment in powers of two to facilitate the tags' computation of random values. In that case the first stage of the gradient search begins with $n(0) = 2^{\lfloor \log_2 2m_c + m_1 \rfloor}$, which is the next larger power of two to the minimally required number of tags to obtain the observed event with m_c collisions and m_1 singleton slots. The second stage, i.e. the binary search, is apparently not necessary at all, since the decision for the frame size has then to be made between $n(i)$ and $n(i-1)$ only. For this decision a simple lookup table can be computed beforehand, which ensures a frame size in powers of two that optimises the throughput [6].

V. Comparison

The best applicable estimation technique is the minimum squared error (MSE) estimator by Vogt [5]. Vogt's approach can easily be modified to partly observed frames (N out of F slots seen) and also discloses the path to exploit the in-frame adaptability of the frame size enabled by current standards, when the evidence of the currently processed frame suggests that. The MSE estimator is formulated as follows. Chebychev's inequality states that the result of a random experiment with random variable $\mathcal{X}$ is most likely near the expected value of that particular event $E(\mathcal{X}_r^{N,F,n})$ of fill level r when distributing n tags into N observed slots out of F slots in a frame:

$$E_{r,N} = E(\mathcal{X}_r^{N,F,n}) = N B_{n,\frac{1}{F}}(r). \quad (10)$$

Hence, a squared error function (11) further ahead is defined using the distance between the expected events $E_{0,N}, E_{1,N}$ and $E_{\geq 2,N} = E_{c,N}$ out of (10) for a given parameter n and the observed values m_0, m_1 and m_c in $N = m_0 + m_1 + m_c$ out of F slots. Via this function that particular n is determined for which the distance function $D(E(\mathcal{X}); N, F, n)$ is minimised.

$$\widehat{n}_{\text{MSE}} = \arg\min_{n} \underbrace{\left| \begin{pmatrix} E_{0,N} \\ E_{1,N} \\ E_{c,N} \end{pmatrix} - \begin{pmatrix} m_0 \\ m_1 \\ m_c \end{pmatrix} \right|^2}_{D(E(\mathcal{X});N,F,n)} \quad (11)$$

As $m_c = N - m_0 - m_1$ and with some simple calculus this is equivalent to

$$\hat{n}_{\text{MSE}} = \arg\min_{n} \left((E_{0,N} - m_0)^2 + (E_{1,N} - m_1)^2 + (E_{0,N} - m_0)(E_{1,N} - m_1) \right). \quad (12)$$

The minimisation of this function can be performed analogously to the maximum search for ML estimator in the preceding section, since, except from the border cases of the parameter set, a single global minimum is formed. Again, although this error function is simpler than (7), an analytical calculation of the minimum is not possible due to the inner terms $\left(1 - \frac{1}{F}\right)^n$ and $n\left(1 - \frac{1}{F}\right)^{n-1}$ in $E_{0,N}$ and $E_{1,N}$, respectively.

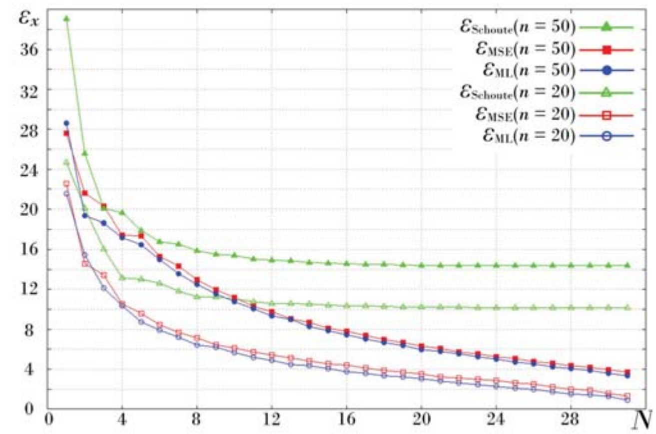

Fig. 6. Cumulated error ε for the MSE, ML, and the Schoute estimator.

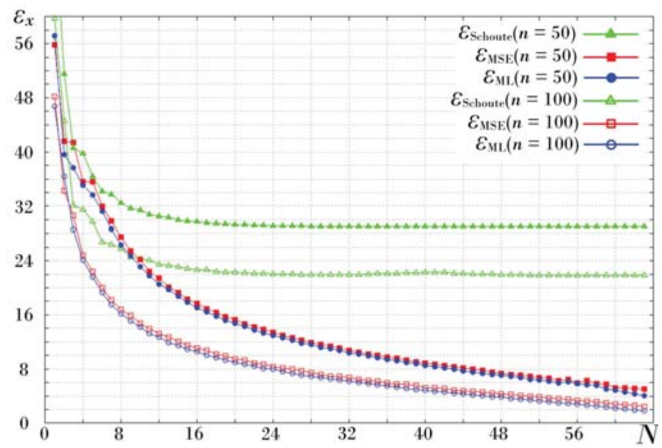

Fig. 7. Cumulated error ε for the MSE, ML, and the Schoute estimator.

To obtain a clear comparison between the tag population estimation techniques, the evaluation function for the absolute expected error of the estimation techniques ($\hat{n}_x = \hat{n}_{\text{ML}}, \hat{n}_{\text{MSE}}, \hat{n}_{\text{Schoute}}$) testify to their quality. For the absolute error function ε we sum up the errors of the estimation functions over the possible event space for a given parameter set and weight the summands by their occurrence probability as it has been proposed by Vogt [5]:

$$\varepsilon_x = \sum_{\langle m_0, m_1, m_c \rangle} |\hat{n}_x - n| \; P(\langle m_0, m_1, m_c \rangle). \quad (13)$$

In Figure 6 the cumulated errors ε_{MSE}, $\varepsilon_{\text{Schoute}}$, and ε_{ML} have been plotted over the observed $N = 1 \ldots F$ slots for the random experiment with $F = 32$, and $n = 20, 50$. Similarly, Figure 7 plots a different scenario with $F = 64$, $n = 50, 100$. Apparently, for all three estimators the quality of the result improves, the more of the current frame has been observed ($N \rightarrow F$). Note, that for very low values of N, the event $m_c = N$ for which both estimators cannot apply their gradient search, is very likely. For both a fallback mechanism has been implemented for this case $N = m_c$ to estimate the number of tags in the current frame to be $\hat{n} = 2.39 m_c \frac{F}{N}$, which has been shown by Schoute [3] for the case $N = F$ to provide a reasonable estimate, although with a much lower quality than the MSE estimator when compared over the complete event space [11]. Hence, the behaviour depicted for $N < 5$ may deviate from this general observation. But more importantly, the ML estimator reveals a better quality for the full event space for any observed number of slots during frame processing. A more accurate estimation of the tag population is directly related to a higher throughput, since the frame size can be adjusted more precisely to the true size of the tag population [6].

Admittedly, the improvement compared to the MSE estimator does not disclose an analogue boost in estimation performance as the MSE reveals compared with Schoute's estimator. This may or may not be surprising since before the knowledge of the ML estimator it remained unclear how much performance improvement compared to the MSE was possible. Moreover, the authors would like to emphasise that, as long as there is no additional knowledge of the tag population available, which could affect the distribution probability of $P(\mathcal{X} = \langle m_0, m_1, m_c \rangle)$, the ML estimator provides a lower bound for (13) to any other stochastic estimation technique.

A disadvantage of the ML estimator lies apparently in its computational overhead. Although the gradient search can be applied in a binary search form, some of the inner terms of (7) are to the power of n or factorial of n and divisions between values of very different size occur, thus necessitating a rather complex computation algorithm to prevent the introduction of overflow, underflow or fractional errors. In fact, the successive slot-by-slot computation of (7), in which only the parameters N and one out of m_0, m_1 or m_c are incremented, offers a lot of potential to diminish this overhead. Similarly, the gradient search can be optimised as well, as only its parameter n is varied from iteration to iteration. Additionally, it seems to be possible to exploit the precise knowledge of the probability function to precociously restart frames, when the momentary observation indicates that the current event is extraordinarily beneficial (considerably more singleton slots than expected in the observed part of the frame) [12].

VI. Conclusion

In this work the maximum likelihood estimator for the tag population of Framed Slotted ALOHA protocols has been developed. The exact probability distribution of the observable event space in FSA systems has been determined, thus enabling the ML formulation of a slot-by-slot tag population estimator. As expected, its estimation performance reveals a better quality over the complete event space compared to Schoute's estimator and most notably compared to Vogt's MSE estimator for completely and partly observed frames.

This result may serve for future algorithms to close in to the theoretical bound of the achievable throughput in FSA protocols. Especially, for state-of-the-art protocols that allow for an in-frame adjustment of the frame size, as EPC Global UHF/HF Class 1 Generation 2, this method can be applied for an immediate update of the frame size according to the probability level of the current slot-by-slot estimate. This issue will be subject of upcoming research. Future work will also focus on its implementation in an RF-ID testbed [13], [14], [15].

References

[1] B. Knerr, M. Holzer, C. Angerer, and M. Rupp, "Slot-by-slot maximum likelihood estimation of tag populations in Framed Slotted ALOHA protocols," in *2008 Int. Symp. Performance Evaluation Computer Telecommun. Syst. SPECTS*, Edinburgh, UK, June 2008, pp. 303-308.

[2] V. Standford, "Pervasive computing goes the last hundred feet with RFID systems," *IEEE Pervasive Computing*, Apr. 2003.

[3] F. C. Schoute, "Dynamic frame length ALOHA," *IEEE Transactions on Communications*, vol. 4, no. 31, pp. 565-568, Apr. 1983.

[4] C. Floerkemeier, "Bayesian transmission strategy for framed ALOHA based RFID protocols," in *Proc. IEEE Int. Conf. RFID 2007*, Grapevine, TX, USA, Mar. 2007.

[5] H. Vogt, "Efficient object identification with passive RFID tags," in *Proc. 1st Int. Conf. Pervasive Computing*. London, UK: Springer, 2002, pp. 98-113.

[6] B. Knerr, M. Holzer, C. Angerer, and M. Rupp, "Slot-by-slot minimum squared error estimator for tag populations in FSA protocols," in *Proc. 2nd Int. EURASIP Workshop RFID Technol.*, Budapest, Hungary, July 2008.

[7] D. Lee, K. Kim, and W. Lee, "Q^+-Algorithm: An enhanced RFID tag collision arbitration algorithm," *Lecture Notes Comput. Sci., Ubiquitous Intelligence Comput.*, vol. 4611, no. 31, pp. 23-32, 2007.

[8] A. Krohn, T. Zimmer, M. Beigl, and C. Decker, "Collaborative sensing in a retail store using synchronous distributed jam signalling," in *Proc. 3rd Int. Conf. Pervasive Comput.*, Munich, Germany, May 2005.

[9] N. L. Johnson and S. Kotz, *Urn Models and Their Applications*. Wiley, 1977.

[10] F. S. Roberts, *Applied Combinatorics*. Prentice-Hall, 1984.

[11] C. Floerkemeier, "Transmission control scheme for fast RFID object identification," Pisa, Italy: IEEE Computer Society, Mar. 2006.

[12] M. Holzer, B. Knerr, C. Angerer, and M. Rupp, "Early frame restart in RFID systems," in *Proc. 2nd Int. EURASIP Workshop RFID Technol.*, Budapest, Hungary, July 2008.

[13] C. Angerer, B. Knerr, M. Holzer, A. Adalan, and M. Rupp, "Flexible simulation and prototyping for RFID designs," in *Proc. 1st Int. EURASIP Workshop RFID Technol.*, Vienna, Austria, Sept. 2007, pp. 51-54.

[14] C. Angerer, M. Holzer, B. Knerr, and M. Rupp, "A flexible dual frequency testbed for RFID," in *Proc. 4th Int. Conf. Testbeds Research Infrastructures Development Netw. Communities (TridentCOM08)*, Innsbruck, Mar. 2008.

[15] R. Langwieser, G. Lasser, C. Angerer, M. Rupp, and A. Scholtz, "A modular UHF-reader frontend for a flexible RFID testbed," in *Proc. 2nd Int. EURASIP Workshop RFID Technol.*, Budapest, Hungary, July 2008.

Bastian Knerr studied electrical engineering at the Saarland University in Saarbruecken, Germany, and at the University of Technology in Hamburg-Harburg, Germany. He finished his diploma thesis on the evaluation of algorithms for carrier frequency synchronisation of base stations in a cellular OFDM system and graduated in 2002. From 2002 to 2003 he worked in the field of image processing and 3D computer vision for the company UZR GmbH & Co. KG in Hamburg, Germany. In June 2003 he joined the Christian-Doppler Laboratory for Design Methodology of Signal Processing Algorithms at the Institute for Communciations and RF Engineering at Vienna University of Technology in Austria and finished his dissertation about heuristic optimisation methods for system partitioning in hardware/software co-design in 2008. Since then, he joined the company freiheit.com Technologies Gmbh in Hamburg, Germany, as software developer and is currently working on large-scale distributed web and database infrastructures.

Martin Holzer received his Dipl.-Ing. degree in electrical engineering from the Vienna University of Technology, Austria, in 1999. During his diploma studies, he worked on the hardware implementation of the LonTalk protocol for Motorola. From 1999 to 2001, he worked at Frequentis in the area of automated testing of TETRA systems and afterwards until 2002 at Infineon Technologies on ASIC design for UMTS mobiles. From 2002 on he had a research position at the Christian Doppler Laboratory for Design Methodology of Signal Processing Algorithms at the Vienna University of Technology where he received a Doctoral Degree in March 2008. Since October 2008 he is with Elektrobit in Vienna, Austria.

Christoph Angerer received his Dipl.-Ing. degree in electrical engineering from the Vienna University of Technology with distinction. For his master degree he collaborated with the Forschungszentrum Telekommunikation Wien (ftw). After he spent one and a half year in industry with embedded system design, he joined the Christian-Doppler Laboratory for Design Methodology of Signal Processing Algorithms at the Institute for Communciations and RF Engineering at Vienna University of Technology in Austria and started working towards his PhD. His research interests include radio frequency identification, embedded systems, wireless communications, digital signal processing and its implementation.

Markus Rupp received his Dipl.-Ing. degree in 1988 at the University of Saarbruecken, Germany and his Dr.-Ing. degree in 1993 at the Technische Universitaet Darmstadt, Germany, where he worked with Eberhardt Haensler on designing new algorithms for acoustical and electrical echo compensation. From November 1993 until July 1995, he had a postdoctoral position at the University of Santa Barbara, California with Sanjit Mitra where he worked with Ali H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 until August 2001 he was a member of Technical Staff in the Wireless Technology Research Department of Bell-Labs at Crawford Hill, NJ, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11 and UMTS. Since October 2001 he is a full professor for Digital Signal Processing in Mobile Communications at the Vienna University of Technology where he founded the Christian-Doppler Laboratory for Design Methodology of Signal Processing Algorithms in 2002 at the Institute for Communciations and RF Engineering. He served as Dean from 2005-2007. He was associate editor of IEEE Transactions on Signal Processing from 2002-2005, is currently associate editor of JASP EURASIP Journal of Advances in Signal Processing, JES EURASIP Journal on Embedded Systems, Research Letters in Signal Processing. He is elected AdCom member of EURASIP since 2004 and serving as president of EURASIP from 2009-2010. He authored and co-authored more than 300 papers and patents on adaptive filtering, wireless communications, and rapid prototyping, as well as automatic design methods.