

Estimating Collision Set Size in Framed Slotted Aloha Wireless Networks and RFID Systems

Andrea Zanella, *Member, IEEE*

Abstract—In wireless networks and RFID systems, a collision set is a group of nodes such that the simultaneous transmission of two or more nodes generates destructive interference at the receiver and, consequently, the loss of all transmitted packets. Knowing the number of nodes in the collision set, i.e., the *collision set size*, it is possible to design effective channel access strategies that minimize the time required to read all the tags or, more generally, collect the packets generated by each node in the set. However, the collision set size is often unknown and, then, needs to be estimated. In this paper we propose a novel estimation method for Framed Slotted Aloha systems that applies a maximum likelihood argument on a Poisson approximation of the packet arrivals process to the transmission channel, yielding an estimate with minimal mean error and mean square error, with respect to the best-performing estimate algorithms in the literature having similar complexity.

Index Terms—Framed Aloha, RFID, conflict, estimate.

I. INTRODUCTION

IN wireless networks and RFID systems, a *collision set* is a group of nodes such that the concurrent transmission of two or more nodes in the set yields a *packet collision*, that is to say, the loss of all transmitted packets due to destructive radio signal interference at the receiver. The design of channel access strategies to reduce the rate of collisions and maximize the number of successful transmissions per unit time is, then, a fundamental problem that has been deeply studied in the past (see, e.g., [1]–[7] and references therein), and whose interest has been recently reawakened by the massive diffusion of RFID systems.

The Framed Slotted Aloha (FSA) scheme is at the basis of many contention-arbitration algorithms [5]–[9] and RFID standards (e.g., ISO 1800-6 TYPE A, ISO 14443-3, ISO/IEC 15693). Basically, FSA divides the time in slots, which are grouped in frames. In each frame, nodes transmit their packet in a random slot, chosen with uniform probability. Slots are said to be *idle*, *successful* or *collided* according to whether they are interested by zero, one or more than one transmission, respectively. Packets sent over successful slots are correctly delivered to the destination, whereas those sent on collided slots are lost and have to be retransmitted in the following frame. The performance of FSA-based protocols is maximized by adapting the frame length w to the cardinality n of the collision set, i.e., the so-called *collision set size* [8]. However, the actual value of n is often unknown and needs to be estimated. This estimation problem is interesting by its own and of particular relevance for the emerging technology of RFID systems.

In this letter, we address the node-counting problem in a FSA access system, and propose a novel *Collision set size*

Estimator (CE) that, compared against the best-performing CEs in the literature with similar complexity [5], [10]–[14], achieves the minimal mean estimate error and mean square error over the largest region of values of n , for any given frame length w . The low complexity and good performance of the proposed method make it particularly attractive for the important field of RFID applications.

II. SYSTEM MODEL

Let n and w denote the collision set size to be estimated and the frame length, respectively. The number K_j of nodes that transmit on the j th slot of the frame is a binomial random variable with parameters n and $1/w$. Note that the random variables $\{K_j\}_{j=1,\dots,w}$ are identically distributed, but not independent, since it must hold $\sum_{j=1}^w K_j = n$. The j th slot is collided, successful, or idle according to whether $K_j \geq 2$, $K_j = 1$, or $K_j = 0$, respectively. Let $\mathbf{V} = \langle C, S, I \rangle$ denote the random vector representing the number of collided (C), successful (S), and idle (I) slots in the frame. The statistical mean of C , S , and I given n are then equal to [12]

$$\begin{aligned} m_C(n) &= w \left[1 - \frac{n}{w} (1 - 1/w)^{n-1} - (1 - 1/w)^n \right] \\ m_S(n) &= n(1 - 1/w)^{n-1}, \quad m_I(n) = w(1 - 1/w)^n. \end{aligned} \quad (1)$$

Furthermore, let $\mathbf{v} = \langle c, s, i \rangle$ denote one observation of the vector \mathbf{V} at the end of the frame, and $\Omega_w = \{\mathbf{v} : s + c + i = w\}$ be the set of all admissible vectors \mathbf{v} . The conditional probability of observing $\mathbf{v} \in \Omega_w$, given n , can be expressed as [15]¹

$$\begin{aligned} P_n(\mathbf{v}) &= \Pr[\mathbf{V} = \mathbf{v} | n] \\ &= \binom{w}{s} \binom{w-s}{c} \sum_{j=0}^c \sum_{\ell=0}^{c-j} \binom{c}{j} \binom{c-j}{\ell} \frac{(-1)^{c-j} n! j^{n-\ell-s}}{w^n (n-\ell-s)!}. \end{aligned} \quad (2)$$

This probability mass distribution (PMD) serves as the basis for sophisticated Bayesian refinement methods (see, for instance, [16]–[18]). However, the computation of the binomial terms in (2) may give rise to numerical stability problems, which can be avoided by resorting to the recursive expression provided in [10]. On the other hand, the recursive method involves a number of operations that increases geometrically with n so that, in general, the evaluation of (2) has large computational cost.

A CE is a deterministic function $\mathcal{H}(\cdot)$ that maps each observation vector $\mathbf{v} \in \Omega_w$ into a non-negative integer that is the estimate of n . We define the computational complexity of a CE as

$$K(w) = \frac{\sum_{\mathbf{v} \in \Omega_w} k(\mathbf{v})}{|\Omega_w|} = \frac{2 \sum_{\mathbf{v} \in \Omega_w} k(\mathbf{v})}{(w+1)(w+2)} \quad (3)$$

Manuscript received October 3, 2011. The associate editor coordinating the review of this letter and approving it for publication was A. Sekercioglu.

The author is with the Department of Information Engineering, University of Padova, Padova, 35131, Italy (e-mail: zanella@dei.unipd.it).

Digital Object Identifier 10.1109/LCOMM.2012.011312.112067

¹Incidentally, we observe that some papers make use of an approximate expression of $P_n(\mathbf{V})$ (see, for instance, [16], [17]), though this approximation is not always recognized or explicitly stated.

where $|\Omega_w| = (w+1)(w+2)/2$ is the cardinality of Ω_w , whereas $k(\mathbf{v})$ denotes the number of iterations required to compute $\mathcal{H}(\mathbf{v})$, each iteration consisting in a fixed number of numerical and logical operations.

III. RELATED WORK

For space constraints, we only provide a quick overview of the CEs for FSA systems considered in this work for comparison purposes. More extensive surveys of the subject can be found, for instance, in [6], [7], [13], [19].

The simplest CE is due to Vogt [11] that proposed a lower bound estimate given by $\mathcal{H}_{Vog1}(\mathbf{v}) = s + 2c$. The rationale is that a successful slot counts for a single node, whereas any collided slot counts for *at least* two nodes.

Schoute [10] and, successively, Cha and Kim [14], proposed the estimate $\mathcal{H}_{Sch}(\mathbf{v}) = s + 2.39c$, where 2.39 is the *asymptotic* mean number of transmissions per collided slot as the collision set size n grows to infinity, and assuming optimal frame length $w = n$.

In [12], Vogt proposes a more complex estimate technique that minimizes the distance between the observed vector \mathbf{v} and the conditional expectation of \mathbf{V} , given n and w , i.e.,

$$\begin{aligned} \mathcal{H}_{Vog2}(\mathbf{v}) &= \arg \min_n \|\mathbf{v} - \mathbf{E}[\mathbf{V}|n, w]\| \\ &= \arg \min_n \{ (m_S(n) - s)^2 + (m_C(n) - c)^2 + (m_I(n) - i)^2 \}. \end{aligned} \quad (4)$$

This estimate however requires the numerical minimization of a composite function [17].

Khandelwat *et al.* [5] apply the minimum distance principle to the number of idle slots only, obtaining

$$\mathcal{H}_{Kha}(\mathbf{v}) = \arg \min_n |m_I - i| = \left\lceil \frac{\log\left(\frac{w}{i}\right)}{\log\left(\frac{w}{w-1}\right)} \right\rceil, \quad (5)$$

where $\lceil x \rceil$ denotes the integer rounding of x .

Similarly, Cha and Kim consider the number of collided slots only, obtaining

$$\mathcal{H}_{CK}(\mathbf{v}) = \arg \min_n |m_C - c|. \quad (6)$$

It is easy to realize that (6) admits a single solution that, however, needs to be searched through numerical methods.

In [13], Kodialam and Nandagopal apply the same concept as (5) and (6) under the simplifying assumption that the number of transmissions in each slot are independent Poisson-distributed random variables with (unknown) parameter $\mu = \frac{n}{w}$. In this way, the authors obtain the following idle-slot and collided-slot CEs:

$$\mathcal{H}_{KNI}(\mathbf{v}) = \left\lceil w \log\left(\frac{w}{i}\right) \right\rceil \quad (7)$$

$$\mathcal{H}_{KNC}(\mathbf{v}) = \lceil w\mu_0 \rceil \quad (8)$$

where μ_0 is the only solution of the transcendental equation

$$(1 + \mu)e^{-\mu} = \frac{w - c}{w} \quad (9)$$

in the unknown μ .

We observe that the estimators \mathcal{H}_{Vog} , \mathcal{H}_{Sch} , \mathcal{H}_{Kha} , and \mathcal{H}_{KNI} have unit computational complexity, requiring a single operation to find the estimate. Conversely, \mathcal{H}_{Vog2} , \mathcal{H}_{CK} , \mathcal{H}_{KC} need to minimize a cost function by means of iterative

numerical methods in order to find the estimate. Assuming that the cost function is well behaved, and considering that the cardinality of the output space of the CEs cannot exceed that of the input space $|\Omega_w| \simeq w^2$, a binary search method would take, on average, order of $\log_2 w^2$ iterations to find the estimate. Then, the computational complexity of these CEs is roughly logarithmic in w , i.e., $K(w) \simeq A \log_2(w)$ for some $A > 0$, as confirmed by the experimental results presented later on in the paper.

IV. MAXIMUM LIKELIHOOD ESTIMATORS

In this section we describe a novel estimation method that exhibits excellent performance with limited computational complexity. To begin with, we consider the classical maximum likelihood estimator (MLE), which returns the value n of the collision set size that maximizes the conditional probability of observing the vector $\mathbf{v} = \langle c, s, i \rangle$ given that n nodes transmit, i.e.,

$$\mathcal{H}_{MLE}(\mathbf{v}) = \arg \max_n P_n(\mathbf{v}). \quad (10)$$

Unfortunately, the computation of $P_n(\mathbf{v})$ requires the evaluation of approximately $(c+1)^2$ terms, so that even without considering the iterative search for the maximum of $P_n(\mathbf{v})$, we have $k_{MLE}(\mathbf{v}) \geq (c+1)^2$ and $K_{MLE}(w) \geq w$. This fact, together with the numerical instability problems in the evaluation of $P_n(\mathbf{v})$, make the MLE unsuitable for simple low-end devices.

A simpler version of the MLE can be obtained by using the approximation suggest in [13] for the number of transmissions in each slot, which are considered as independent Poisson random variables of (unknown) mean $\mu = n/w$. Under this approximation, the conditional probability of observing \mathbf{v} , given μ , is easily found to be

$$\hat{P}_\mu(\mathbf{v}) = \mu^s e^{-\mu w} (e^\mu - 1 - \mu)^c. \quad (11)$$

Applying the MLE argument to this PDF we thus obtain

$$\mathcal{H}_{Zan}(\mathbf{v}) = \lceil w\hat{\mu} \rceil, \quad (12)$$

where $\hat{\mu}$ maximizes (11). To find this value we set to zero the derivative of (11) in μ , which gives

$$\frac{\mu w - s}{c} = \frac{\mu(e^\mu - 1)}{e^\mu - 1 - \mu}. \quad (13)$$

A simple functional analysis reveals that (13) is well behaved and admits only one non-negative solution $\hat{\mu}$, which can be determined by bisection search.

In terms of computational complexity, we observe that, for $c = 0$, we immediately have $\mathcal{H}_{Zan}(\mathbf{v}) = s$, whereas for $c = w$ the estimate is infinite. We then have $k(\mathbf{v}) = 1$ for all $\mathbf{v} = \langle 0, s, w - s \rangle$ and $\mathbf{v} = \langle w, 0, 0 \rangle$. When $c \in \{1, \dots, w-1\}$, instead, the estimate requires the solution of (13) by a binary search method. Now, the right-hand side of (13), denoted by $R(\mu)$, is a monotonic increasing function of the only unknown μ that starts from $R(0) = 2$ and rapidly converges to the straight line $y(\mu) = \mu$ as μ grows to infinity.² The left-hand side of (13), denoted by $L(\mu, \mathbf{v})$, is instead

²Since $R(\mu)$ does not depend on the parameters w , s , c , and i , it is possible to avoid the online computation of the function and use in its place a pre-computed table of sampled values.

a simple linear function of μ , with slope w/c and abscissa intercept s/c . The solution $\hat{\mu}$ of (13) thus always falls into the interval $\left(\frac{s}{w-c}, \frac{s+2c}{w-c}\right)$, delimited by the intersection of $L(\mu; \mathbf{v})$ with the lower bound $y(\mu)$ and the upper bound $y(\mu) + 2$, respectively. The length of the interval is $\frac{2c}{w-c}$, so that the estimate (12) shall be searched in a set of approximately $\frac{2cw}{w-c}$ integer values. A binary search method takes order of $k(\mathbf{v}) = 1 + \log_2\left(\frac{2cw}{w-c}\right)$ iterations to find the solution in this interval. Adding $k(\mathbf{v})$ over all the possible vectors $\mathbf{v} \in \Omega_w$, and dividing by $|\Omega_w|$, we get after some algebra

$$K_{Zan}(w) \simeq \frac{w^2 + 3w + 1 + 2 \sum_{r=1}^{w-1} \log_2\left(\frac{[2w(w-r)]^{r+1}}{r^{r+1}}\right)}{(w+1)(w+2)}. \quad (14)$$

Relaxing the sum in r to an integral, it is possible to prove that (14) asymptotically grows with w as quickly as $\log_2(w)$, so that the complexity of (12) is comparable with that of (6), (8), and (4).

V. PERFORMANCE ANALYSIS

To characterize the performance of a CE, we consider the random variable $\hat{n} = \mathcal{H}(\mathbf{V})$ obtained by applying the estimate function $\mathcal{H}(\cdot)$ to the random vector \mathbf{V} .

The normalized estimate error is defined as

$$\varepsilon(\mathbf{V}) = \frac{\hat{n} - n}{n} = \frac{\mathcal{H}(\mathbf{V}) - n}{n}, \quad (15)$$

and its conditional mean and power, given n , are respectively denoted by $m_\varepsilon(n, w) = \mathbb{E}[\varepsilon(\mathbf{V})|n, w]$ and $M_\varepsilon(n, w) = \mathbb{E}[\varepsilon(\mathbf{V})^2|n, w]$, where the expectation is taken with respect to PMD (2). Note that $M_\varepsilon(n, w)$ corresponds to the normalized mean square error (MSE) of the estimate.

The conditional PMD of \hat{n} , given the actual n , can be obtained as

$$\hat{P}_n(m) = \Pr[\hat{n} = m|n] = \frac{\sum_{\mathbf{v} \in H(m)} P_n(\mathbf{v})}{\sum_{\mathbf{v} \in \Omega_w} P_n(\mathbf{v})}$$

where $H(m) = \{\mathbf{v} : \mathcal{H}(\mathbf{v}) = m\}$ is the subset of observations that give estimate m . With respect to $\hat{P}_n(m)$, we define the *uncertainty interval*

$$U_r(n) = [m_{\frac{1-r}{2}}(n), m_{\frac{1+r}{2}}(n)] \quad (16)$$

where

$$m_k(n) = \arg \min_m \left\{ \sum_{h=0}^m \hat{P}_n(h) \leq k \right\}. \quad (17)$$

For a given n , the uncertainty interval $U_r(n)$ identifies the range of values where the estimate falls with probability r : the tighter the interval around n , the more reliable the estimate.

Unless otherwise stated, the results presented in the following of this section have been obtained by setting the frame length $w = 128$, as typical in the related literature.³ In any case, the results remain qualitatively the same for different values of w , though the absolute values of the performance indexes clearly scale with w .

³To have meaningful results, we replaced the infinite estimate returned by each CE with the maximal finite estimate provided by the same CE, augmented by 2.39.

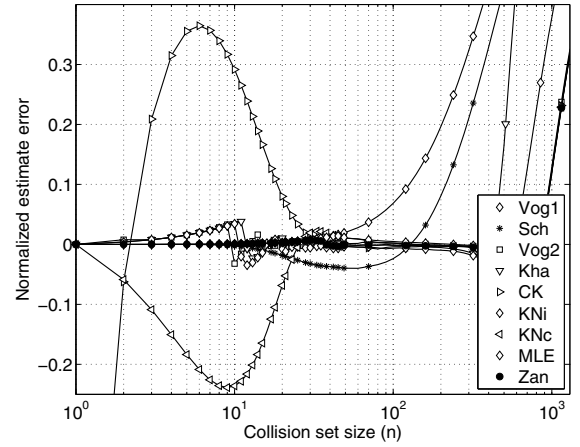


Fig. 1. Normalized mean estimate error $m_\varepsilon(n, w)$ vs collision set size n , with $w = 128$.

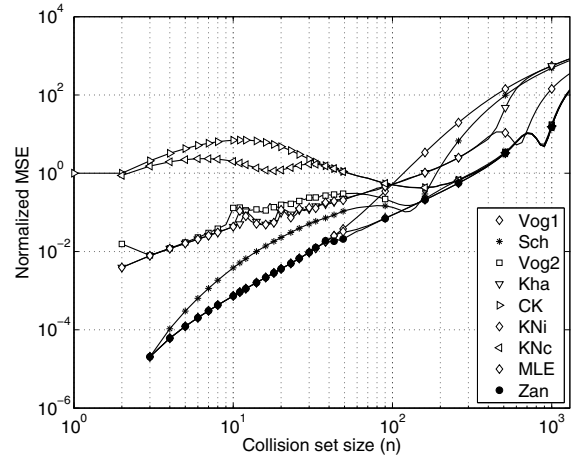


Fig. 2. Normalized mean square estimate error $M_\varepsilon(n, w)$ vs collision set size n , with $w = 128$.

Fig. 1 shows the normalized mean estimate error $m_\varepsilon(n, w)$ as a function of n . The graph is plotted in logarithmic scale in order to emphasize the results for low-medium values of n . We see that the estimate error of most CEs is non-zero even for small values of n . As expected, the simplest methods, \mathcal{H}_{Vog1} and \mathcal{H}_{Sch} , are very good for small values of n , but their performance quickly drops as n grows. Conversely, methods \mathcal{H}_{CK} and \mathcal{H}_{KNc} exhibit quite large normalized estimate error for small values of n , whereas their performance improves for medium to large values of n . The normalized estimate error of \mathcal{H}_{Vog2} , \mathcal{H}_{Kha} , and \mathcal{H}_{KNi} oscillates approximately between $\pm 4\%$ for small values of n . The error becomes negligible in the medium-value region, when $n \simeq w$, and increases again for larger values of n . Finally, \mathcal{H}_{MLE} and \mathcal{H}_{Zan} achieves the minimal mean normalized estimate error over the entire range of values of n .

Fig. 2 reports the normalized mean square estimate error for the different CEs when varying the collision set size n . We observe that, once again, \mathcal{H}_{MLE} and \mathcal{H}_{Zan} outperform the other estimate algorithms, in particular for small to medium values of n .

The mean square estimate error is related to the statistical reliability of the estimate. To better appreciate this aspect, we

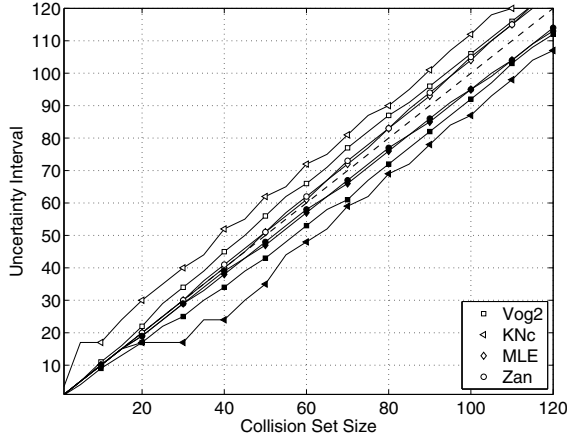


Fig. 3. δ -percentiles vs collision set size estimate \hat{n} , with $w = 128$.

report in Fig. 3 the uncertainty interval $U_{0.9}(n)$ for \mathcal{H}_{Vog2} , \mathcal{H}_{KNc} , \mathcal{H}_{MLE} , and \mathcal{H}_{Zan} . From Fig. 3, we see that \mathcal{H}_{MLE} and \mathcal{H}_{Zan} achieve the smallest uncertainty interval for small-medium values of n , though the performance gain vanishes for larger n . This property is particularly valuable when the CE is plugged into a frame-adaptation scheme that works on successive rounds, since the lower the mean square estimate error the fastest the convergence of the estimate to the actual value of the collision set size and, consequently, the more effective the frame-adaptation strategy. It is thus reasonable to expect that the performance gaps among the different CEs observed after one single frame will be magnified by any refinement strategies in successive frames.

Finally, in Fig. 4 we report the empirical computational complexity $K(w)$ of the best-performing CEs, when varying the frame length w .⁴ The dashed curve represents the theoretical computational complexity of \mathcal{H}_{Zan} , as given by (14). We can see that all the considered CEs have similar complexity, requiring a mean number of iterations that grows logarithmically with w . The computational complexity of \mathcal{H}_{MLE} , instead, is *at least* linear with w and has not been reported in the figure.

VI. CONCLUSIONS

In this paper, we proposed a novel method to estimate the number of transmitting nodes in a collision-prone FSA system. The estimate is obtained by applying the maximum likelihood argument to a Poisson approximation of the probability distribution of the number of successful, collided and idle slots in one frame. The estimate achieves the minimum mean error and mean square error over all the other methods with similar complexity in the literature, basically reaching the same performance as the exact MLE that, however, has much larger computational complexity. Conversely, the mean computational complexity of the proposed method is logarithmic in the frame length w and, hence, suitable for low-end devices.

⁴Note that the number of iterations taken by each CE may depend on the specific implementation of the search algorithm. For fair comparison, we used binary search in all the cases. The MATLAB code used to derive the results is available at <http://dgt.dei.unipd.it/download>.

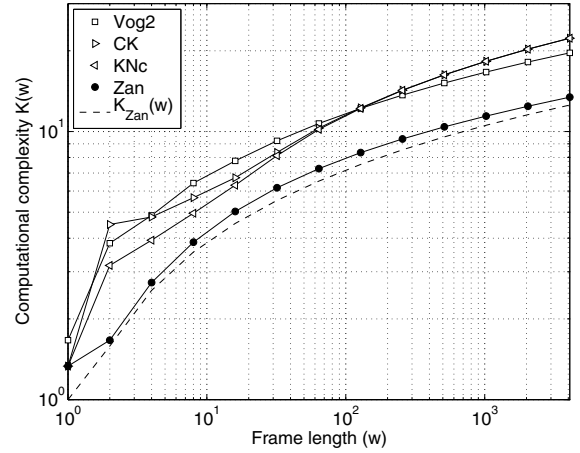


Fig. 4. Computational complexity $K(w)$ vs frame length w for different CEs.

REFERENCES

- [1] J. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 505–515, Sep. 1979.
- [2] R. Gallager, "Conflict resolution in random access broadcast networks," in *Proc. 1978 AFOSR Workshop Commun. Theory Appl.*, pp. 74–76.
- [3] M. Molle and G. Polyzos, "Conflict resolution algorithms and their performance analysis," Department of Computer Science and Engineering, UCSD, East Lansing, Michigan, Tech. Rep. CS93-300, July 1993.
- [4] C. Floerkemeier, "Transmission control scheme for fast RFID object identification," in *Proc. 2006 IEEE International Conference on Pervasive Computing and Communications Workshops*, pp. 462–467.
- [5] G. Khandelwal, A. Yener, K. Lee, and S. Serbetli, "ASAP: a MAC protocol for dense and time constrained RFID systems," in *Proc. 2006 IEEE International Conference on Communications*, vol. 9, pp. 4028–4033.
- [6] T. La Porta, G. Maselli, and C. Petrioli, "Anti-collision protocols for single-reader RFID systems: temporal analysis and optimization," *IEEE Trans. Mobile Comput.*, vol. PP, no. 99, pp. 1–1, 2010.
- [7] D. Klair, K.-W. Chin, and R. Raad, "A survey and tutorial of RFID anti-collision protocols," *IEEE Commun. Surveys Tutorials*, vol. 12, no. 3, pp. 400–421, 3rd quarter 2010.
- [8] D. K. Klair, K.-W. Chin, and R. Raad, "On the suitability of framed slotted Aloha based RFID anti-collision protocols for use in RFID-enhanced WSNs," in *Proc. 2007 IEEE International Conference on Computer Communications and Networks*, pp. 583–590.
- [9] Z. G. Prodanoff, "Optimal frame size analysis for framed slotted Aloha based RFID networks," *Computer Commun.*, 2010.
- [10] F. C. Schoute, "Dynamic frame length ALOHA," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 565–568, Apr. 1983.
- [11] H. Vogt, "Multiple object identification with passive RFID tags," in *2002 IEEE International Conference on Systems, Man and Cybernetics*.
- [12] —, "Efficient object identification with passive RFID tags," in *Proc. 2002 International Conference on Pervasive Computing*, pp. 98–113.
- [13] M. Kodialam and T. Nandagopal, "Fast and reliable estimation schemes in RFID systems," in *Proc. 2006 Annual International Conference on Mobile Computing and Networking*, pp. 322–333.
- [14] J.-R. Cha and J.-H. Kim, "Novel anti-collision algorithms for fast object identification in RFID system," in *Proc. 2005 International Conference on Parallel and Distributed Systems*, vol. 2, pp. 63–67.
- [15] J. Riordan, *An Introduction to Combinatorial Analysis*. Princeton University Press, 1978.
- [16] W.-T. Chen, "An accurate tag estimate method for improving the performance of an RFID anticollision algorithm based on dynamic frame length ALOHA," *IEEE Trans. Automation Sci. and Engineer.*, vol. 6, no. 1, pp. 9–15, 2009.
- [17] H. Wu and Y. Zeng, "Bayesian tag estimate and optimal frame length for anti-collision Aloha RFID system," *IEEE Trans. Automation Sci. and Engineer.*, vol. 7, no. 4, pp. 963–969, 2010.
- [18] Q. Tong, X. Zou, and H. Tong, "Dynamic framed slotted Aloha algorithm based on Bayesian estimation in RFID system," in *Proc. 2009 World Congress on Computer Science and Information Engineering*, vol. 1, pp. 384–388.
- [19] C. Floerkemeier and M. Wille, "Comparison of transmission schemes for framed Aloha based RFID protocols," in *Proc. 2006 International Symposium on Applications and the Internet Workshops*, pp. 97–100.