Medium Access Control Analytical Modeling and Optimization in Unslotted IEEE 802.15.4 Wireless Sensor Networks

C. Fischione, S. Coleri Ergen, P. Park, K. H. Johansson, A. Sangiovanni-Vincentelli

Abstract—Accurate analytical expressions of delay and packet reception probabilities, and energy consumption of duty-cycled wireless sensor networks with random medium access control (MAC) are instrumental for the efficient design and optimization of these resource-constrained networks. Given a clustered network topology with unslotted IEEE 802.15.4 and preamble sampling MAC, a novel approach to the modeling of the delay, reliability, and energy consumption is proposed. The challenging part in such a modeling is the random MAC and sleep policy of the receivers, which prevents to establish the exact time of data packet transmission. The analysis gives expressions as function of sleep time, listening time, traffic rate and MAC parameters. The analytical results are then used to optimize the duty cycle of the nodes and MAC protocol parameters. The approach provides a significant reduction of the energy consumption compared to existing solutions in the literature. Monte Carlo simulations by ns2 assess the validity of the analysis.

Index Terms—Wireless Sensor Networks, MAC, Duty Cycle, Optimization.

I. INTRODUCTION

Energy-efficient wireless sensor networks (WSNs) are providing new and affordable services for a huge variety of applications. Given the lack of cabling that characterizes these networks, WSNs allow to embed sensing, communication, control, and actuation everywhere in the real world, creating the so-called physical internet. Ensuring energy efficiency of WSNs is a difficult task for applications where WSN must provide sensed information for real time action, because reliable and timely data transmission may threat energy consumption.

In this paper, we consider the design of an energy efficient duty cycling based on the IEEE 802.15.4 communication standard [1]. Idle listening of sensors is one of the major components of the energy budget. Asynchronous duty cycling medium access control (MAC) protocols such as B-MAC [2] and X-MAC [3] have been proposed as an effective mechanism to reduce idle listening in random access networks. These duty cycling protocols work on top of IEEE 802.15.4 and do not require any modification of the standard. In these methods, the receiver wakes up periodically to check whether there is a transmission, and the sender transmits preambles to check if the receiver is awake. The main advantage of these protocols is that there is no complex control mechanisms, as in time division multiple access (TDMA) schemes, for discovering the network topology, keeping the nodes synchronized [4] and running the schedules efficiently [5] or in synchronized duty cycling networks, for negotiating a schedule among the neighboring nodes to specify when the nodes are awake and asleep [6, 7]. However, the intrinsic simplicity of the mechanism has the drawback of smaller energy saving potential as compared to the more complex solutions listed above, unless optimization of sleep and wake times is adapted to changing data traffic conditions.

Energy modeling and the usage of this energy model in sleep and wake time optimization was considered in B-MAC [2] and X-MAC [3]. However, when using these protocols with IEEE 802.15.4, the formulation does not take into account the effect of random access, which is a function of the data traffic, MAC parameters and topology. This is a crucial aspect, since the duration of the random access is much larger than the actual packet transmission: In IEEE 802.15.4 [1] radios with default parameter settings, the maximum backoff before packet transmission is 27.4 ms whereas the transmission time of a 56 byte packet is 1.79 ms at 250 kbps. Moreover, in [2, 3] and references therein, no delay or reliability constraint on packet delivery is included. MAC protocols for sensor networks have certain latency and reliability requirements in addition to low energy consumption. Since many applications, e.g. security monitoring, require guaranteed arrival of sensor data to the collection center and others require a certain degree of reliability in delivering sensor data (e.g., control and automation applications), latency and reliability must be considered in MAC design.

The original contributions of this paper are two: First, we provide accurate expressions of delay, reliability, and energy consumption as a function of sleep time, listening time, traffic rate and MAC parameters, the validity of which is demonstrated by both analysis and Monte Carlo simulations in ns2. Second, we demonstrate the usage of these formulations in optimization of duty cycle of the nodes by minimizing energy consumption under latency constraints and reliability. To the best of our knowledge, this is the first work to provide such an accurate analytical modeling and optimization of duty cycled networks with latency and reliability constraints.
The rest of the paper is organized as follows: Section II describes the preamble based protocol. Sections III, V and VI provide the analytical expressions for delay probability, reliability, and energy consumption of preamble sampling MAC on top of unslotted IEEE 802.15.4 networks, respectively. In Section VII we illustrate the advantage of such formulations in the optimization of duty cycle parameters. In Section VII we summarize the main results and future work.

II. Protocol Description

We assume that the nodes of the WSN are organized into clusters as shown in Fig. 1. Clustered network topology is supported in large networks that require energy efficiency, since transmitting data directly to the base station may consume more than routing through intermediate nodes [8]. In a clustered topology, nodes organize themselves into clusters with a node acting as cluster head. All non-cluster head nodes transmit their data directly to the cluster head, while the cluster head receives data from all cluster members and transmits them to a remote base station. Throughout this paper we consider applications where nodes asynchronously generate packets with rate $\lambda$ packets per second (see Table I for a list of main symbols used in the paper). The protocol we are investigating is referred to low data rate applications, namely we assume that $\lambda \leq 1$.

For the network topology and applications we are considering, the asynchronous cycle-MAC protocol based on preamble sampling with acknowledgment called X-MAC [3] offers good performance. In preamble sampling protocols, the receiver wakes up periodically for a short time to sample the medium. Such a time is defined as the listening time. When a sender has data, it transmits a series of short preamble packets, each containing the ID of the target node, until it either receives an acknowledgement packet (ACK) from the receiver or a maximum time is exceeded (see Fig. 2). We assume that such a maximum time is given by the sleep plus listening time of the receiver. Following the transmission of each preamble packet, the transmitter node goes in a listening state having a maximum timeout duration $T_{TX_{out}}$. If the receiver is the target, it sends an acknowledgement (ACK) during the pause between the preamble packets. When the receiver node sends an ACK, it waits for data packets for a duration of at least $T_{out}$ even after the end of the wake-up time. Consequently, the maximum listening time is $R_1 + T_{out}$. The extension of $T_{out}$ to the regular listening time allows for the reception of the data packets whose ACK was sent near the expiration of $R_1$. Upon reception of the ACK, the sender transmits the data packet to the receiver. However, the transmission of such a packet occurs after sensing the channel idle. If the channel is busy, data transmission may be delayed too much. The transmitter gives up the transmission of the data packet if the delay from the first attempt to transmit a preamble is larger than $R_s + R_1$.

It is natural that preambles and acknowledgements in preamble sampling protocols are sent by using a random access to avoid collisions, as allowed by IEEE 802.15.4 [1]. Although this is not considered in [3], it gives obvious benefits in terms of delay and reliability, because acknowledgments and preambles may collide with any other packet. The amount of random access, which depends on the data traffic, network topology and the parameters of the MAC protocol should be included in the energy minimization problem, because random access determines the time interval between the transmissions of two consecutive preamble packets. It determines wake time, since the receiver node should receive at least one preamble packet during the wake time. Furthermore, the amount of random access is affected by sleep time, since increasing sleep time increases the number of preambles.

The variables established by our protocol are the listening time $R_1$ and sleep time $R_s$ of the receiver node given the channel condition, data traffic, topology of the network, and number of nodes. These variables can be obtained as solution of optimization problems that consider the total energy consumption, delay and reliability in the packet delivery, as we will see in Section VI.

A. IEEE 802.15.4 unslotted CSMA/CA Mechanism

In the unslotted IEEE 802.15.4 carrier sense multiple access with collision avoidance (CSMA/CA) mechanism, each device in the network has two variables: $NB$ and $BE$. $NB$ is the number of times the CSMA/CA algorithm is required to backoff while attempting the current transmission. $NB$ is initialized to 0 before every new transmission. $BE$ is the backoff exponent, which is related to how many backoff periods a device must wait before it attempts to access the channel. The algorithm is implemented using units of time called backoff.
periods, which are given by $aUnitBackoffPeriod$ symbols. The parameters that affect the random backoff are $BE_{\text{min}}$, $BE_{\text{max}}$ and $NB_{\text{max}}$, which correspond to the minimum and maximum of $BE$ and the maximum of $NB$, respectively.

The unslotted CSMA/CA mechanism works as follows. $NB$ and $BE$ are initialized to 0 and $BE_{\text{min}}$ respectively (Step 1). The MAC layer delays for a random number of complete backoff periods in the range 0 to $3^{BE} - 1$ (step 2) and then requests PHY to perform a CCA (clear channel assessment) (step 3). If the channel is assessed to be busy (step 4), the MAC sub-layer increments both $NB$ and $BE$ by one, ensuring that $BE$ is not more than $BE_{\text{max}}$. If the value of $NB$ is less than or equal to $NB_{\text{max}}$, the CSMA/CA must return to Step 2, else the CSMA/CA must terminate with a Channel-Access-Failure status. If the channel is assessed to be idle (Step 5), the MAC layer starts transmission immediately.

The expected number of times random backoff is repeated is a function of the probability of sensing the channel busy, which depends on the channel traffic. Channel traffic depends on data traffic, network topology and duty cycle parameters, i.e. sleep and wake time, since they determine the expected number of preamble packets. This complex interdependence is investigated in the following sections.

III. DELAY MODELING

In this section, we model the delay needed for the successful packet transmission from a transmitter node to the receiver. First, we restrict our attention to a transmitter-receiver pair, and then in Section IV we generalize the analysis to the case of many transmitters. Let us denote by TX the transmitting node, and by RX the receiver node. In the modeling of the delay that we propose in this section, we assume that the time is counted from the moment in which the TX node has a packet to send.

The delay to transmit a packet successfully is a function of three random components:
- $T_1$: random delay spent by the TX node to complete the transmission of a preamble packet. It includes also the processing time and the transmission time of the preamble.
- $T_2$: random delay spent by the TX node until the receiver node is in the active state and an acknowledgment packet reaches the TX node;
- $T_3$: random delay spent by the TX node from the instant of acknowledgement reception until the transmission of a data packet. It includes also the processing time and the transmission time of the data packet.

Hence, the delay to transmit successfully a data packet is given as $T_p = T_2 + T_3$. In the following, we characterize the three delay components $T_1$, $T_2$ and $T_3$.

A. Modeling of $T_1$

In this subsection, we provide the exact expressions of the average and variance of $T_1$. Then, we approximate the distribution of $T_1$ by a normal distribution, whose average and variance are obtained through a moment matching approach. Such an approximation is motivated by that a closed form expression for the distribution of $T_1$ cannot be achieved, as we will discuss later. We will show that the approximation is quite accurate.

The mechanism to transmit a preamble packet is the same as the one for data packets, for we are using to assume IEEE 802.15.4. If the channel is busy, a random back off is spent before a further trial. Let $NB_{\text{max}} \leq NB \leq R_s/S_c$ be the maximum number of back-off of a preamble, namely the number of times that the TX node attempts to access the channel before giving up the transmission of a preamble, where $S_c$ is the sensing time. By denoting with $S_{p,j}$ the random back-off time at the $j$-th trial, it follows that $S_{p,j}$ has a uniform distribution in the interval $[0, 2^{(j)} - 1]$, for $j = 1, \ldots, Nb$, where $r(j) = \min(\text{rem}(j, NB_{\text{max}}) + BE_{\text{min}} - 1, BE_{\text{max}})$, with $\text{rem}(\cdot, \cdot)$ being the remainder of the division of the first by the second argument.

Denote by $A_k$ the event occurring when the channel is busy for $k-1$ times, and then is free at the $k$-th time. The probability of such an event is

$$\Pr[A_k] = e^{k-1}(1 - c),$$

where $c$ is the probability to sense the channel busy. We assume that this probability is independent at each attempt. Such an approximation is very accurate for saturated traffic in [2], and has been widely adopted in the literature also for unsaturated traffic (see, e.g., [10], [11], [12] and references therein). In Sections III IV and V we show by Monte Carlo simulations that this approximation is quite accurate within the operational region of WSNs. Consider the attempt of transmission of the $i$-th preamble. Then, random delay $T_1$ spent by the TX node before transmitting a preamble packet within $N_b$ attempts can be described as

$$T_1 = \begin{cases} S_{p,1} + S_c + T_{hr}, & \text{if } A_1 \mid A; \\ S_{p,1} + S_c + S_{p,2} + S_c + T_{hr}, & \text{if } A_2 \mid A; \\ \vdots \\ \sum_{j=1}^{N_b}(S_{p,j} + S_c) + T_{hr}, & \text{if } A_{N_b} \mid A. \end{cases}$$

where $T_{hr}$ is the time employed by the hardware platform to process the packets and transmit them, and $A$ is the event that a preamble is transmitted with at maximum $N_b$ preambles:

$$\Pr[A] = \Pr\left[\sum_{j=1}^{N_b} A_j = \sum_{j=1}^{N_b} \Pr[A_j]\right],$$

where previous inequality comes from that the events $A_j$, $j = 1, \ldots, N_b$ are mutually exclusive. It holds

$$\Pr[A_k \mid A] = \frac{\Pr\left[\sum_{j=1}^{N_b} A_j\right]}{\Pr[A]} = \frac{\Pr[A_k]}{\sum_{j=1}^{N_b} \Pr[A_j]},$$

We can rewrite $T_1$ as

$$T_1 = \sum_{k=1}^{N_b} \left( \sum_{j=1}^{k} (S_{p,j} + S_c) + T_{hr} \right) \mathbb{I}_{A_k \mid A} = \sum_{j=1}^{N_b} \sum_{k=1}^{k} \mathbb{I}_{A_k \mid A},$$

(1)
where $\mathbb{1}_{(.)}$ it the indicator function (its value is 1 if the argument is true, and 0 otherwise) and

$$\Sigma_k = \sum_{j=1}^{k} (S_{p,j} + S_c) + T_{hr}.$$  

From previous equation, $\Sigma_k$ is given by the sum of independently uniformly distributed random variables plus a constant. The expectation $\Sigma_k$ can be computed by recalling the distribution of $S_{p,j}$, whose average is

$$\mu_{S_{p,j}} = \frac{(2^{r(j)} - 1) S_b}{2},$$  

where $S_b = aUnitBackoffPeriod$. Hence

$$\mu_{\Sigma_k} = \mathbb{E}[\Sigma_k] = \sum_{j=1}^{k} [\mu_{S_{p,j}} + S_c] + T_{hr}. \quad (2)$$

The variance of $\Sigma_k$ is given by the sum of the variances of $S_{p,j}$, whose variance is

$$\sigma^2_{S_{p,j}} = \frac{(2^{2r(j)} - 1) S_b^2}{12},$$

hence

$$\sigma^2_{\Sigma_k} = \mathbb{E}[(\Sigma_k - \mathbb{E}[\Sigma_k])^2] = \sum_{j=1}^{k} \sigma^2_{S_{p,j}}. \quad (3)$$

Using (2) and (3) it is possible to compute the exact expression of the average value and the correlation of $T_1$ as

$$\mu_{T_1} = \mathbb{E}[T_1] = \sum_{k=1}^{N_b} \frac{\mu_{\Sigma_k} e^{k}}{\sum_{j=1}^{N_b} e^{j}} e^{-\rho_{\Sigma_k}},$$

$$\rho_{T_1} = \mathbb{E}[T_1^2] = \sum_{k=1}^{N_b} \frac{\rho_{\Sigma_k} e^{k}}{\sum_{j=1}^{N_b} e^{j}} e^{-\rho_{\Sigma_k}},$$

where $\rho_{\Sigma_k} = \frac{\sigma^2_{\Sigma_k}}{\mu_{\Sigma_k}}$. From these moments, the variance of $T_1$ follows

$$\sigma^2_{T_1} \triangleq \rho_{T_1} - \mu_{T_1}.$$  

Since $T_1$ is the weighted sum of uniform random variables having different mean and variance, no closed form expression is available for the probability mass function (PMF). However, we resort to a normal distribution to approximate the PMF of $T_1$, namely, we assume that

$$f_{T_1}(x) \sim \frac{1}{\sigma_{T_1} \sqrt{2\pi}} \exp \left( -\frac{(x - \mu_{T_1})^2}{2\sigma_{T_1}^2} \right). \quad (4)$$

In Subsection III.C we show that this approximation matches well the real one obtained via Monte Carlo simulations. Finally, from (4) we can compute the cumulative distribution function (CDF) of $T_1$ by the error function.

**B. Modeling of $T_2$**

In this section we model $T_2$, the random delay the TX node waits until an ACK is sent by the RX node and reaches the TX node that sent the preamble. With this goal in mind, we need to define three random variables, $T_a$, $T_1$ and $N_p$, which we present next.

First, let us denote by $T_a$ the random time to wait from the beginning of the transmissions until the start of the listen time. The time $T_a$ can be modeled by the following random variable:

$$T_a = \begin{cases} 0, & \text{if } \bar{S}; \\ T_s, & \text{if } S. \end{cases} \quad (5)$$

where $T_s$ is the random time to wait that the receiver wakes up. This time can be modeled as a uniform distribution in the range $[0, R_s]$, since such a time is computed from the beginning of the transmission of the TX node, which may uniformly fall in the interval $[0, R_s]$ (see Fig.2). The event $S$ occurs when the RX node is sleeping. Since a node sleeps for $R_s$ seconds and is awake for $R_t$ seconds, it follows that

$$\Pr[S] = \frac{R_s}{R_s + R_t}, \quad \Pr[\bar{S}] = 1 - \Pr[S].$$

From the definition (5) we rewrite $T_a$ as

$$\Pr[T_a = 0] = \Pr[T_s | S] = \Pr[T_s] \frac{R_s}{R_s + R_t}.$$  

We define by $T_l$ the time duration from the moment wherein the preamble packet is received in the listening time of the receiver, until the listening time expires. By following the same approach as the one used for the characterization of $T_s$, it follows that $T_l$ has a uniform distribution in the interval $[0, R_t]$.

Consider $N_p$, the random number of preambles that should be sent before one falls in the active time of the receiver and the acknowledgment is sent back by the RX node. Denote with $T_{ack}$ the random time to complete the transmission of an acknowledgement sent by the RX node after a preamble packet is received. Notice that its statistical distribution is the same as $T_1$, since an acknowledgement is transmitted by following the same mechanism of a preamble, the only difference being that $N_b$ must be replaced with $NB_{max}$. Furthermore, let us define $B_k$ as the event that a preamble has to be sent $k$ times before being received in the active time of the RX node and the corresponding acknowledgement is sent by the RX node and received before the time out of the TX node. Clearly, the event $B_k$ is conditioned on the random active time of $T_a$ of the RX and on the random remaining listening time $T_l$ of the RX. These times are random from the point of view of the transmitter, which does not know when the receiver wakes up and when it will go to sleep.

We are now in the position of defining the delay $T_2$. The following equation gives $T_2$:

$$T_2 = \begin{cases} T_{1,1} + T_{ack}, & \text{if } B_1 \mid B; \\ T_{1,1} + T_{TX, out} + T_{1,2} + T_{ack}, & \text{if } B_2 \mid B; \\ \vdots & \vdots \\ \sum_{j=1}^{N_p} T_{1,j} + (N_p - 1)T_{TX, out} + T_{ack}, & \text{if } B_{N_p} \mid B. \end{cases}$$

where $T_{1,j}$ is the delay for the transmission of $j$-th preamble. The distribution of $T_{1,j}$ is given by (1). $B$ is the probability
that the TX node receives an ACK within $N_p$ preambles:
\[
\Pr[B] = \Pr \left[ \sum_{k=1}^{N_p} B_k \right] = \sum_{k=1}^{N_p} \Pr[B_k],
\]
where previous inequality comes from the fact that the events $B_j, j = 1, \ldots, N_p$ are mutually exclusive. It holds
\[
\Pr[B_i|B] = \frac{\Pr[B_i \sum_{k=1}^{N_p} B_k]}{\Pr[B]} = \frac{\Pr[B_i]}{\sum_{j=1}^{N_p} \Pr[B_j]}.
\]
We describe $B_k$ next. First, let $b$ be the collision probability of preambles or ACK. Such a probability is different from the probability of collision of data packets, which we denote by $p$, because the size of preambles and ACK is much smaller than data packets. We assume that these probabilities are independent at each attempt. Such an approximation has been widely adopted in the literature (see, e.g., [9], [10], [11], [12] and references therein). In Sections III TV and V we show by Monte Carlo simulations that this approximation is quite accurate within the operational region of WSNs.

**Proposition 1** ([13]) Let $B_k$, with $k \in \mathbb{N}$, the event occurring when $k-1$ preambles are sent before the $k$-th is received in the active time of the RX node, and the acknowledgement is sent back by the RX node and received before the time out of the TX node. Let $\Omega$ be the certain event. Then
\[
B_k = [C_k + D_{k-1}E_{k-1} + b + D_{k-1}E_{k-1}(1-b)F_{k-1} + D_{k-1}E_{k-1}(1-b)F_{k-1}b]D_{k}E_{k}F_{k}(1-b)^2,
\]
where
\[
C_k = [(k-1)I_{k-1=0}T_1 + (k-2)I_{k-2=0}T_{TX,\text{out}} \leq T_a],
\]
\[
D_k = [kT_1 + (k-1)I_{k-1=0}T_{TX,\text{out}} > T_a],
\]
\[
E_k = [kT_1 + (k-1)I_{k-1=0}T_{TX,\text{out}} \leq T_a + T_1],
\]
\[
F_{k-1} = [T_{\text{ack}} > T_{\text{TX, out}}|D_{k-1}],
\]
\[
F_k = [T_{\text{ack}} \leq T_{\text{TX, out}}|D_k],
\]
\[
F_0 = [T_{\text{ack}} > T_{\text{TX, out}}],
\]
\[
D_0 = \Omega.
\]

**Proposition 2** ([13])
\[
\Pr[B_k] = \Pr[C_kE_k] - \Pr[D_k] \Pr[T_{\text{ack}} \leq T_{\text{TX, out}}]
\times (1-b)^2 + (\Pr[E_k] - \Pr[C_kE_k])
\times \Pr[T_{\text{ack}} \leq T_{\text{TX, out}}]b(1-b)^2
+ (\Pr[E_k] - \Pr[C_kE_k]) (1 - \Pr[T_{\text{ack}} \leq T_{\text{TX, out}}])
\times \Pr[T_{\text{ack}} \leq T_{\text{TX, out}}](1-b)^3
+ (\Pr[E_k] - \Pr[C_kE_k])
\times (\Pr[T_{\text{ack}} \leq T_{\text{TX, out}}])^2 b(1-b)^3,
\]
where
\[
\Pr[C_k] = P_1 \left( \frac{T_a - (k-2)T_{TX,\text{out}}}{k-1} \right),
\]
\[
\Pr[D_k] = P_1 \left( \frac{T_a - (k-1)T_{TX,\text{out}}}{k} \right),
\]
\[
\Pr[E_k] = P_1 \left( \frac{T_a + T_1 - (k-1)T_{TX,\text{out}}}{k} \right),
\]
\[
\Pr[C_kE_k] = \Pr[C_k]\Pr[T_1 \leq T_a - T_{TX,\text{out}}] + \Pr[E_k](1 - \Pr[T_1 \leq T_a - T_{TX,\text{out}}]).
\]

**Proposition 3**: ([13]) The maximum number of preambles to send is
\[
N_p = 2 + \frac{R_s}{T_{\text{TX, out}}}.
\]
From Propositions 1 and 2 the average and variance of $T_2$ is
\[
\mu_{T_2} = \sum_{k=1}^{N_p} \left[ k\mu T_1 + (k-1)T_{TX,\text{out}} + \mu T_{\text{ack}} \right] \Pr[B_k|B],
\]
\[
\sigma_{T_2}^2 = \sum_{k=1}^{N_p} \sigma_{T_2,k}^2 \Pr[B_k|B],
\]
where $\sigma_{T_2,k}$ is the variance of $\sum_{j=1}^{k} T_{1,j} + (k-1)T_{TX,\text{out}} + T_{\text{ack}}$. Since $T_2$ is given by the weighted sum of variables that we approximated in Subsection III-A as normal distributed, it follows that the PMF of $T_2$ can be approximated by a normal random variable, namely
\[
f_{T_2}(x) \sim \frac{1}{\sigma_{T_2} \sqrt{2\pi}} \exp \left( -\frac{(x - \mu_{T_2})^2}{2\sigma_{T_2}^2} \right).
\]
In Subsection III-C we show that this approximation is quite accurate.

**C. Delay Probability**

The delay to send successfully a data packet is given by $T_p = T_2 + T_3$, given $T_a$ and $T_1$. In particular, $T_2$ is a function of $T_a$ and $T_1$ (see Subsection III-B). Looking at Fig. 2 it is straightforward to see that $T_3$ can be characterized as $T_{\text{ack}}$ except for a higher constant transmission time within $T_{hr}$, so that $T_3$ is approximated by a normal random variable. It follows that $T_p$ is approximated by a normal distribution as well, namely
\[
f_{T_p}(x) \sim \frac{1}{\sigma_{T_p} \sqrt{2\pi}} \exp \left( -\frac{(x - \mu_{T_p})^2}{2\sigma_{T_p}^2} \right),
\]
where $\mu_{T_p} = \mu_{T_2} + \mu_{T_3}$, $\sigma_{T_p}^2 = \sigma_{T_2}^2 + \sigma_{T_3}^2$.

The distribution of the delay we have modeled so far is conditioned on the active time of the receiver $T_a$ and the time duration from preamble reception in the listening time of the receiver until the listening time expires $T_1$. The probability that a packet is delayed some $t_{\text{max}}$ seconds and falls in the listening time of the RX node is therefore given by
\[
E_{T_a} E_{T_1} \Pr(T_p \leq t_{\text{max}}) \triangleq D_{\text{max}}(R_t, R_s, c, b, t_{\text{max}}),
\]

where $E_{T_a}$ and $E_{T_l}$ denote the statistical average with respect to the distribution of $T_a$ and $T_l$, respectively. Since the CDF of (9) is given by the error function, it is a highly nonlinear function of the random variables $T_a$ and $T_l$. Therefore, the averages $E_{T_a}$ and $E_{T_l}$ are obtained by replacing $T_a$ and $T_l$ with their respective expectations, as proposed in [14] pag. 428. This is equivalent to replace $\mu_{T_a}$ with $E_{T_a}$, $\mu_{T_l}$ and $\sigma_{T_l}^2$ with $E_{T_a}$ and $\sigma_{T_l}^2$ in (9). Furthermore, notice that in (10) we evidenced the dependency on the active period and sleep period of the receiver, as well as on the busy channel probability $c$.

We validated the analysis of the delay by comparing the expectation and variance of (10) to extensive Monte Carlo simulations obtained by an ns2 simulator. The simulator reproduced the system depicted in Fig. 1 where transmitter nodes send packets according to the preamble-based MAC. All the numerical values set for the simulations are taken coherently with the IEEE 802.15.4 standard [11] and the Tmote wireless sensors [15]. Although the expectation and variance of (10) have been derived for a single transmitter-receiver pair, we consider it as an approximation for the general case of several transmitters. This is motivated by the fact that the analysis considers the collision and busy channel probabilities, which accounts in some manner for the case of multiple transmitters. As a matter of fact, we observed an almost perfect matching analysis-simulations for all the cases of practical interest. Specifically, in Fig. 3 we reported the average delay of $T_p$ for $N = 8$ transmitter nodes, whereas in Fig. 4 we reported the average variance of $T_p$ for several cases of traffic period $\lambda$ and sleep time for $N = 8$ transmitter nodes. We chose a traffic period larger than 10 s since higher traffic rates would exhibit packet losses probabilities larger than 50%, which is of no-interest. It can be shown that the behavior of Fig. 3 and Fig. 4 is observed for any network size that gives a packet reception rate larger than 50%, which is attained for $R_s \geq 6$ms.

We conclude that the analysis of the delay is quite accurate both for the single transmitter-receiver pair, and for multiple transmitters.

IV. RELIABILITY ANALYSIS

In this Section, we analyze the reliability, or probability that a data packet is successfully received.

The failure of a data packet transmission is owed to three possibilities: 1) a preamble is not successfully received, 2) the ACK is not successfully received, and 3) the data packet is not successfully received. In the following, we characterize these events.

In Proposition 1 we defined $B_k$, with $k \in \mathbb{N}$, as the event occurring when $k-1$ preambles are sent before the $k$-th is received in the active time of the RX node, and the acknowledgement is sent back by the RX node and received before the time out of the TX node. For analytical tractability, $B_k$ was derived for a single transmitter-receiver pair. However, we assume to use it also for the derivation of the reliability in the general case of several transmitters. Such an assumption is reasonable, as we will show by extensive Monte Carlo simulations presented at the end of this Section.

Let the event $\mathcal{G}$ occur when a preamble is successfully transmitted during the active time of the receiver within $N_h$ trials (which occurs with probability $\Pr[|\mathcal{J}_p| = 1 - c^{N_h}]$) and the corresponding ACK is successfully sent within $NB_{\text{max}}$ trials (which occurs with probability $\Pr[|\mathcal{J}_a| = 1 - c^{NB_{\text{max}}}]$). Then

$$\mathcal{G} = \sum_{k=1}^{N_p} B_k.$$  

(11)

By observing that $B_k$ and $B_j$ are mutually exclusive if $i \neq j$, it follows

$$\Pr[\mathcal{G}] = (1 - c^{N_h})(1 - c^{NB_{\text{max}}}) \sum_{k=1}^{N_p} \Pr[B_k].$$  

(12)

Define the event $\mathcal{I} | \mathcal{G}$, which occurs when the TX sends successfully a data packet, provided that a preamble is successfully received and the ACK is also successfully received,
Fig. 5: Reliability as obtained by analysis (Eq. (14)) and simulations for a network with $N = 8$ nodes. The traffic period is 10 s, 30 s and 300 s. On the x axis, the ratio $R_s/R_l$ is reported.

then

$$\Pr[I|G] = (1 - e^{NB_{\text{max}}}) (1 - p).$$

Finally, by putting together (11) and (13), and averaging with respect to the distribution of $T_s$ and $T_l$, the reliability is given by

$$R_{\text{min}}(R_l/R_s, c, b) = \mathbb{E}_T_s \mathbb{E}_T_l \Pr[I] \Pr[I|G].$$

These expectations are computed as done in Eq. (10). We remark here that the reliability depends on the traffic through the busy channel probability $c$, and the collision probability of preambles or ACK $b$. However, we noticed that both analysis and simulations, discussed later, give a negligible dependency on the collision probability of data packet, because these packets are transmitted only upon the reception of an ACK.

We validated the analysis of the reliability by comparing Eq. (14) to extensive Monte Carlo simulations that were obtained by ns2, as described in Subsection III-C. In Fig. 5, we reported an example of such simulations for the case of 8 nodes. We can see that the analysis follows well the simulations results. The sharp fluctuations in the analytical results are due to the PMF of the random back-off, which is discontinuous with sudden jumps due to the discrete increase of the exponential back-off and the magnitude of the busy channel probability. The good behavior of the approximation observed in Fig. 5 is also obtained for other network sizes that give a packet reception rate larger than 50%. We observed that the difference analysis-simulations is always below 5%, so we conclude that Eq. (14) is a good approximation.

V. ENERGY CONSUMPTION

In this section we characterize the energy consumption of the network. The total normalized energy consumption over a listening-sleeping time $R_s + R_l$ is calculated considering the energy spent by a TX node to send a data packet ($\mathbb{E} E_{tx}$) and by the RX node to receive a data packet ($\mathbb{E} E_{rx}$):

$$\mathbb{E} E_{\text{tot}} = \frac{N d_{\text{TX}} \mathbb{E} E_{tx} + \mathbb{E} E_{rx}}{R_s + R_l},$$

where recall that $N$ is the number of TX nodes in a cluster, and $d_{\text{TX}}$ is the probability that a TX node has at least one data packet to send during the time $R_l + R_s$.

The total energy consumption was normalized by $R_s + R_l$ so that it is intended as an energy per time unit. The remaining energy components in (15) are characterized in the following.

A. Energy Consumption at the Transmitters

Denote with $P_{tx}$, $P_{rx}$, and $P_s$ the power required to transmit, receive and sleep, respectively. We have the following results:

**Proposition 4:** [13] Let $S_{p,k}^n$ be the $n$-th random backoff of the $k$-th preamble. The instantaneous transmit energy for a TX node is upper bounded as

$$E_{tx} \leq \sum_{i=1}^{N_p} \left[ \sum_{l,n=1}^{i} \left( E_{tx,T_i}^{(n)} + (i-1) E_{T_{\text{TX,out}}} + E_{tx,T_{\text{ack}}} + E_{tx,T_{\text{data}}} \right) \right] \mathbb{I}_B,$$

$$+ \left[ \sum_{n=1}^{N_p} E_{tx,T_i}^{(n)} + (N_p - 1) E_{T_{\text{TX,out}}} + E_{tx,T_{\text{ack}}} \right] \mathbb{I}_S,$$

where

$$E_{tx,T_i}^{(n)} = \sum_{j=1}^{N_b} \left[ \sum_{k=1}^{j} \left( P_{tx,S_{p,k}^n} + P_{rx,S_{c}^n} + P_{tx,S_d} \right) \right] \mathbb{I}_{A_j},$$

$$E_{T_{\text{TX,out}}} = T_{\text{TX,out}} P_{tx},$$

$$E_{tx,T_{\text{ack}}} = \sum_{j=1}^{N_b} \left[ \sum_{k=1}^{j} \left( P_{tx} S_{p,k} + S_a \right) \right] P_{tx} \mathbb{I}_{A_j},$$

$$E_{tx,T_{\text{data}}} = \sum_{j=1}^{N_b} \left[ \sum_{k=1}^{j} \left( P_{tx} S_{p,k} + P_{rx,S_c} + P_{tx} S_d \right) \right] \mathbb{I}_{A_j},$$

$$+ \left[ \sum_{k=1}^{N_b} \left( P_{tx} S_{p,k} + P_{tx} S_c \right) \right] \mathbb{I}_{A_d}.$$

**Remark 1:** We derived an upper bound for analytical tractability. Specifically, the upper bound in Eq. (16) is given by considering the worst case in the number of preambles to be transmitted when no idle channel is found. We see in Subsection V-C that such a bound is reasonable.

Using previous Proposition, we can compute easily the average energy to transmit a data packet. Details are provided in our technical report [13].

B. Energy Consumption at the Receiver

The energy consumed at the receiver is upper bounded by

$$\mathbb{E} E_{rx} \leq R_s P_s + (R_l + T_{\text{out}}) \max(P_{tx}, P_{rx}).$$

where we considered the fact that the RX can be listening for a time $T_{\text{out}}$ after the end of the listening time if an acknowledgement was sent just before the end of the listening time. The upper bound for $\mathbb{E} E_{rx}$ is motivated by the fact that such an energy is given by the idle listening, sending an acknowledgement and receiving data packets. Since these events are highly cross correlated among them and among different sensors, it is difficult to provide a closed form expression.
and simulations. On the x axis, the sleep time $R_s$ is reported for a traffic rate of 300 s and 8 nodes in the network and for different listening time of 8, 15 and 30 ms. The traffic period is 300 s.

for the probabilities of these events. As a result, an accurate characterization would require modeling the probability that the RX node is busy with the reception of a data packet while some other node is trying to send another data packet, which is very difficult, if not impossible, to model. In the next Subsection, we will see that Eq. (21) is a satisfactory bound.

C. Average Energy Consumption

The dependency of $E_{\text{tot}}$ on the random variables $T_n$ and $T_i$ can be removed by taking the expectation with respect to $T_n$ and $T_i$. The total average energy consumption is denoted as

$$E(R_t, R_s, c, b, \lambda) \triangleq E_{T_n} E_{T_i} E_{E_{\text{tot}}}.$$ (22)

In this equation, we have evidenced that the average total energy consumption depends on the collision probability of preamble and ACK, the busy channel probability, the sleep time and the listen time. We noticed that the dependency on the collision probability of data packets is negligible numerically, because data packets are sent only upon the reception of an ACK. The expectations are computed as done in Eq. (10).

We validated the analysis of the average energy consumption by comparing Eq. (22) to extensive Monte Carlo simulations that were obtained by an ns2 simulator, as described in Subsection III-C. We observed a good matching for all cases of practical interest. In Fig. 6, we reported an example of the average energy consumption for several cases of sleep time and a fixed traffic rate. As expected, the analysis gives an upper bound, which is desirable for optimization purposes. The minimum of the energy consumption as given by Eq. (22) is very near the true minimum given by simulation results. It can be shown that the behavior of Fig. 6 is observed for other choices of listening time $R_l$ and for different network sizes. We conclude that our analysis of the average energy consumption gives a good upper bound.

VI. PARAMETER OPTIMIZATION

In the previous sections, we have modeled the distributions of the delay to send a data packet from the transmitter to the receiver, the reliability and the energy consumption. Given a set of collision and busy channel probabilities, these expressions can be used off-line to select the optimal values of the sleep time $R_s$ and listening time $R_l$ that minimize the energy consumption given the latency and reliability constraints. Specifically, the following optimization problem can be solved

$$\min_{R_l, R_s} E(R_l, R_s, c, b, \lambda)$$

subject to

$$D_{\text{max}}(R_l, R_s, c, b, t_{\text{max}}) \leq \tau_{\text{max}},$$

$$R_{\text{min}}(R_l, R_s, c, b) \geq \psi_{\text{min}},$$

where $\tau_{\text{max}}$ is the desired probability that the delay is less than $t_{\text{max}}$, and $\psi_{\text{min}}$ is the minimum desired probability with which a data packet should be received. Notice that the decision variables are the listening time $R_l$ and the sleep time $R_s$. In the optimization, we assume that $c$ and $b$ are feed-forward variables, whereas $\lambda$, $t_{\text{max}}$, $\tau_{\text{max}}$ and $\psi_{\text{min}}$ are application requirements.

Given $c$ and $b$, the optimal solution of problem (23), denoted by $R_l^*(c, b)$ and $R_s^*(c, b)$, can be loaded in a light look up table to be stored in the cluster-head node. The table can be thought of as a matrix with rows associated to the set of values of $c$ and columns associated to the values of $b$. The cluster-head node can easily do an estimation of the busy channel $\hat{c}$ and collision probabilities $\hat{b}$, and read from the look-up table the entries $R_l^*(c, b)$ and $R_s^*(c, b)$ at location $c, b$ closer to $\hat{c}$ and $\hat{b}$. For instance, if we consider 10 values for $\hat{c}$ and 10 for $b$, the table would have 100 entries. By assuming that each entry takes 1 byte, the table has the size of just 0.1 Kb.

We compared the minimization of (23) to the one provided by X-MAC [3]. Recall that such a protocol does not take into account random backoff, delay and reliability constraints. Therefore, for the sake of comparison of the protocol proposed in this paper and X-MAC, we pose $\tau_{\text{max}} = \infty$ and $\psi_{\text{min}} = 0$, which implies neglecting the delay and reliability requirements, i.e., the energy is minimized without constraints, as done in X-MAC.

Our optimal $R_s^*$ and $R_l^*$ outperforms X-MAC in all the scenarios of practical interest. Specifically, when the packet generation period is high (300 s) we achieve an energy consumption 5% less than X-MAC, but as the packet generation period decreases the improvement is substantial, more than 50%. The main reason for this large difference is that the nodes consume much less energy in packet transmission compared to the model in [3]. X-MAC is based on the assumption that the transmitter sends preamble packets back to back until the receiver wakes up, while actually there is random backoff before packet transmissions during which the transmitter puts its radio in sleep mode. Since the transmit energy dominates the receive energy much earlier according to the model in [3], the optimal wake time becomes considerably higher compared to the actual optimal wake time that we achieve.

As a final observation, we remark that our theoretical modeling can be used not only to minimize the energy consumption, but also to maximize the reliability under the constraint of a maximum latency and maximum energy expenditure, or to minimize the latency under the constraints of a minimum reliability and maximum energy expenditure, as we will show.
VII. CONCLUSIONS

In this paper we developed a novel analytical characterization of the delay and packet loss probability distribution, and energy consumption for a clustered network topology with unslotted IEEE 802.15.4 and preamble sampling MAC. The analysis was based on the statistical modeling of the preamble, acknowledgement and data packet transmission. Monte Carlo simulations in ns2 validated and illustrated our approach.

Our analysis can be used efficiently to provide a set of optimal sleep and listening times that minimize the energy consumption of the network while guaranteeing latency and reliability constraints. Compared to existing protocols that minimize only the energy consumption, as B-MAC and X-MAC, our optimization gives much better results. Thus our method can be effectively employed to ensure a longer lifetime of the network.

REFERENCES


<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX node</td>
<td>a transmitter node</td>
</tr>
<tr>
<td>RX node</td>
<td>a receiver node (cluster head)</td>
</tr>
<tr>
<td>$T_1$</td>
<td>random delay spent by the TX node before transmitting a preamble packet</td>
</tr>
<tr>
<td>$T_2$</td>
<td>random delay spent by the TX node from the beginning of a transmission until the reception of the acknowledgement</td>
</tr>
<tr>
<td>$T_3$</td>
<td>random delay spent by the TX node from the acknowledgement reception until the transmission of a data packet</td>
</tr>
<tr>
<td>$T_p$</td>
<td>random delay to wait before a data packet is successfully received</td>
</tr>
<tr>
<td>$T_s$</td>
<td>random sleep time of the RX as seen from the TX (it is uniformly distributed over $[0, R_s]$)</td>
</tr>
<tr>
<td>$T_l$</td>
<td>random listening time of the RX as computed upon the reception of a preamble (it is uniformly distributed over $[0, R_l]$)</td>
</tr>
<tr>
<td>$T_h$</td>
<td>time employed by the hardware platform to process the packets and transmit them</td>
</tr>
<tr>
<td>$T_{ack}$</td>
<td>random time before the RX node can access the channel and send an acknowledgement</td>
</tr>
<tr>
<td>$T_{RX, out}$</td>
<td>maximum time that a TX node waits for an ACK after having sent a preamble.</td>
</tr>
<tr>
<td>$T_{out}$</td>
<td>maximum time that a TX node waits from the moment of the reception of an ACK before giving up the data packet transmission.</td>
</tr>
<tr>
<td>$N_p$</td>
<td>maximum number of preambles that can be sent</td>
</tr>
<tr>
<td>$N_b$</td>
<td>number of nodes in a cluster</td>
</tr>
<tr>
<td>$c$</td>
<td>probability of busy channel</td>
</tr>
<tr>
<td>$b$</td>
<td>probability of preamble or ACK collision</td>
</tr>
<tr>
<td>$p$</td>
<td>probability of data packet collision</td>
</tr>
<tr>
<td>$\nu_{min}$</td>
<td>minimum probability of successful packet transmission (reliability requirement)</td>
</tr>
<tr>
<td>$\tau_{max}$</td>
<td>maximum probability of maximum delay (latency requirement)</td>
</tr>
<tr>
<td>$P_{max}$</td>
<td>maximum energy consumption per listening-sleeping cycle (energy requirement)</td>
</tr>
<tr>
<td>$S_{p,k}$</td>
<td>$k$-th random back-off time of a preamble</td>
</tr>
<tr>
<td>$\mu_{S_{p,k}}$</td>
<td>average of $S_{p,k}$</td>
</tr>
<tr>
<td>$S_{ch}$</td>
<td>time duration of channel sensing for clear channel assessment</td>
</tr>
<tr>
<td>$S_p$</td>
<td>time duration of a preamble packet</td>
</tr>
<tr>
<td>$S_{ch}$</td>
<td>time duration of an acknowledgement packet</td>
</tr>
<tr>
<td>$S_{data}$</td>
<td>time duration of a data packet</td>
</tr>
<tr>
<td>$S_k$</td>
<td>time duration of forming the basic time period used by the CSMACA algorithm</td>
</tr>
<tr>
<td>$P_x$</td>
<td>transmit power</td>
</tr>
<tr>
<td>$P_r$</td>
<td>receive power</td>
</tr>
<tr>
<td>$P_s$</td>
<td>sleep power</td>
</tr>
<tr>
<td>$R_s$</td>
<td>sleep time of the receiver node (cluster head)</td>
</tr>
<tr>
<td>$R_l$</td>
<td>active time of the receiver node (cluster head)</td>
</tr>
<tr>
<td>$A_k$</td>
<td>event occurring when the channel is busy for $k$ - 1 times</td>
</tr>
<tr>
<td>$B_k$</td>
<td>event occurring when a preamble has to be sent $k$ times before being received in the active time of the RX node and the corresponding acknowledgement is sent by the RX node and received before the time out of the TX node</td>
</tr>
</tbody>
</table>

TABLE I: Main symbols used in the paper.