On the Performance of Correlation-Based Packet Detection Techniques

Estefanía Recayte, Andrea Munari[a]

[a]*Institute of Communications and Navigation of DLR, German Aerospace Center, Wessling, Germany*

Abstract

This work studies the performance of packet detection techniques in wireless systems where nodes transmit in a sporadic and asynchronous manner. Modelling a transmitter as a M/D/1 queue and focusing on a correlation-based approach, we derive exact expressions for the detection probability and false alarm rate at the receiver under AWGN conditions for any packet generation rate. Very tight approximations are also provided, and the performance of the correlator is compared to that of a likelihood-ratio-test solution. A significant aspect that has been frequently overlooked in literature is the necessity of addressing false alarms that may be caused by the data portion of sent messages. This factor has emerged in this work as a key consideration.

1. Introduction

An increasing number of applications in the context of wireless communications are characterized by the transmission of possibly short packets in a sporadic and asynchronous fashion. Relevant examples are for instance offered by low-power wide-area networks (LWPAN), where low-complexity devices access the channel with very low duty cycle and no scheduling coordination with the receiver. Many commercial solutions systems employ asynchronous medium access policies based on ALOHA where devices transmit without a predefined schedule (LoRa, 2023; Sigfox, 2023; Abramson, 1977). In such settings, the ability to correctly identify the arrival instant of incoming messages at the receiver is crucial to avoid both the loss of data and

Email address: `{estefania.recayte, andrea.munari}@dlr.de` (Estefanía Recayte, Andrea Munari)

do unnecessary signal processing. To this aim, the most common approach is to prepend a known sequence of bits (known as preamble or sync-word) to each transmitted message. This enables the receiver to detect the message by using techniques such as correlation.

In view of its importance, packet detection has been extensively studied in the literature. Among the key results a pioneering role is played by Massey's contribution (Massey, 1972), which identifies the optimal rule for locating a periodical sync-word in continuous data streams under white Gaussian noise (AWGN) conditions. In the same setting, (Robertson, 1995) considers the marker concept with maximum selection, processing a large window and storing the detection value for each position to determine the position of the sync word (or preamble) start. By selecting the position with the highest correlation value, this technique can improve receiver performance by reducing false alarms at the cost of an increased design and processing complexity. The problem was further explored by Chiani et al. in (Chiani et al., 2007) where analytical bounds for coherent detection and known constant-length frames were derived, while bounds for the optimum frame synchronization in AWGN channels were presented in (Chiani et al., 2005). Preamble detection was also studied in (Ziabari, 2013) and (Vangelista, 2022) where a rough search for the preamble is carried out with a certain degree of error allowance, i.e. a tolerated time windows is defined. This is then followed by a precise synchronization algorithm to fine-tune the margin of error.

In (Nagaraj et al., 2006) the correlator performance in the presence of Poisson traffic was considered, performance approximations in the low signal-to-noise ratio (SNR) regime were presented. Stimulated by the increased interest for LPWAN applications, a number of recent works have further addressed the problem. In particular, the study of optimal detectors was tackled in (Wuerll et al., 2018, 2017) where authors focus on detecting the preamble in an AWGN channel in presence of frequency offsets. However, the surrounding data symbols have always been neglected in such works, creating possible inaccuracies in the overall system performance. More recently, optimum burst detectors with and without knowledge of the noise variance in SIMO systems were also investigated in (Neumüller et al., 2022). The optimized thresholds for preamble detection in commercial systems were re-examined by overlooking irrelevant false alarm events in (Kang, 2023). These fundamental works, however, either assume periodic traffic or only consider false alarms resulting from noise, and thus not fully captures some relevant aspects of many wireless systems. In several practical wireless systems with sporadic transmissions, the erroneous detections induced by the data portion of incoming packets becomes a critical factor.

This manuscripts takes the initial steps to bridge the gap in the field of grant-free random access, i.e. when no coordination takes place. We consider a setting in which transmissions are modeled as an M/D/1 queue, packets are generated and transmitted to an unique receiver over an AWGN channel. In this setup, we derive exact expressions for the detection performance of a correlator as a function of the traffic generation rate, accounting for all possible sources of false alarms. Tight approximations for detection probability and false alarm rate are also derived, offering a useful system design tool. The behavior of the correlator is also compared to the performance of a bound represented by the likelihood ratio test (LRT) approach. The analysis highlights how neglecting the actual traffic profile when tuning the detection threshold can severely impact performance. Obtained results are representative for systems in which nodes operate at low duty cycle to preserve battery or due to channel access regulations, e.g. due to operations in the ISM band (LoRa, 2023). However, the present work can have practical applications in serveral scenarios where grant-free protocol are implemented. Examples of applications include, but are not limited to, tracking systems and monitoring systems. For instance, in maritime communications transmitter can be represented by vessels with containers which have to send their GPS position in an uncoordinater manner to a unique receiver, such as a satellite. In parallel, our work aims to stimulate further research on the topic, considering the behavior of detection techniques for in the presence of different traffic profiles as well as of multi-access interference.

The remaining of the paper is organize as follows. Section II describes the system model considered while Section III illustrates some useful preliminaries. Section IV and Section V address the correlator and the likelihood ratio test, respectively. Section VI presents the results while conclusions are drawn in Section VII.

Notation: We use serif (bold) capital letters for random variables (vectors) and their lowercase counterparts for realizations. The probability density function (pdf) of a random vector $\mathbf{X}$ is denoted as $f(\mathbf{x})$, the probability mass function (pmf) of a discrete rv X is denoted as $p_{\mathsf{X}}(\mathsf{x})$. The notation iid stands for rv independent and identically distributed.

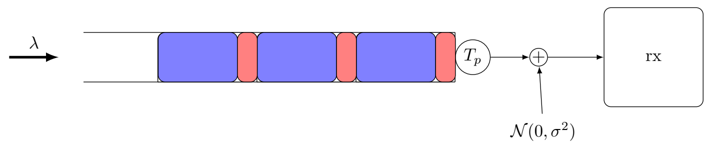

Figure 1: Queue system model: the arrival Poisson rate λ and service time T_p.

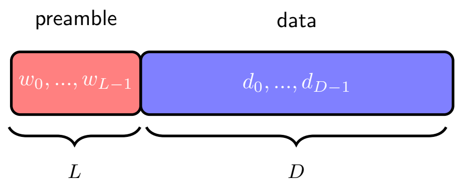

Figure 2: Packet structure of lenght P: L preamble symbols and D data symbols, $P = L + D$.

2. System Model

We focus on a source transmitting fixed-length packets towards a destination. Traffic generation follows a Poisson distribution of parameter λ [pkt/s], and physical layer parameters are set so that each transmission has duration T_p seconds, as shown in Figure 1. The behavior of the source is then appropriately captured by a M/D/1 queue model. We further consider an unbounded buffer size, and the system to operate in stable conditions, i.e., $\lambda T_p < 1$. Accordingly, we denote the stationary probability that the system is empty with π_0, and the probability to have a single packet (in service) as π_1, obtaining (Shortle et al., 1974)

$$\pi_0 = 1 - \lambda T_p, \quad \pi_1 = (1 - \lambda T_p)(e^{\lambda T_p} - 1). \tag{1}$$

As illustrated in Fig. 2, each packet is composed in total by P bits (or symbols). L bits are part of the preamble known to the receiver and employed for detection purposes and is followed by $D > L$ payload bits. Note that $D > L$ and $P = L + D$. BPSK modulation (Chiani et al., 2007) and propagation over an AWGN channel are considered such that in the presence of a transmission. The incoming signal at the receiver at discrete symbol time $\mu \in \mathbb{N}$ can be modeled as

$$\mathsf{Y}_\mu = \mathsf{X}_\mu + \mathsf{N}_\mu$$

with $\mathsf{X}_\mu \in \{-1, 1\}$ and in the absence of a transmission $\mathsf{Y}_\mu = \mathsf{N}_\mu$, and $\mathsf{N}_\mu \sim \mathcal{N}(0, \sigma^2)$. The system operates then at a SNR $\Gamma = 1/(2\sigma^2)$. We denote the symbol time as T_s, and the overall packet duration as $T_p = (L + D)T_s$. Without loss of generality, the propagation delay is neglected so that the receiver at time μ observes the queue output at the same instant[1].

In this setting, the receiver aims to detect the start of incoming packets at a symbol level. Specifically, at time index μ, a vector $\mathbf{Y}_\mu = (\mathsf{Y}_\mu, ..., \mathsf{Y}_{\mu+L-1})$ of L consecutive symbols is stored, and used to detect the presence of a preamble. It is generally assumed that the detection process remains constantly active.

We study such decision problem considering two approaches as presented next.

2.1. Correlator

In this case, the known preamble sequence $\mathbf{w} = \{w_0, w_1, \ldots, w_{L-1}\}$ is correlated at symbol time μ with the received stream. The receiver operates via sliding windows of duration of L symbols, corresponding to the preamble lenght. The output C_μ of the correlator is described by the r.v. as follows

$$\mathsf{C}_\mu := \langle \mathbf{Y}_\mu, \mathbf{w} \rangle = \sum_{i=0}^{L-1} \mathsf{Y}_{\mu+i} \cdot w_i \,. \tag{2}$$

The result is then compared against a predefined threshold τ and the start of a preamble in position μ is declared whenever $\mathsf{C}_\mu \geq \tau$.

2.2. Likelihood ratio test (LRT)

In the hypothesis-testing approach, the receiver considers two possible conditions based on the observed samples vector of lenght L

$$\begin{aligned} \mathcal{H}_0 &: \mathsf{Y}_{\mu+i} \neq w_i + n_{\mu+i} && \forall i = 0, ..., L-1 \\ \mathcal{H}_1 &: \mathsf{Y}_{\mu+i} = w_i + n_{\mu+i} && \forall i = 0, ..., L-1. \end{aligned}$$

The first hypothesis represents the absence of the complete preamble sequence within the stored vector, whereas the second corresponds to the presence of the complete preamble. Leaning on this, the receiver bases its decision

[1]Note that although propagation delay is an important performance metric when evaluating a communication system, it does not have any impact on the detection process.

on the likelihood ratio

$$\Phi(\mathbf{y}) = \frac{f(\mathbf{y}|\mathcal{H}_0, \lambda)}{f(\mathbf{y}|\mathcal{H}_1, \lambda)} \tag{3}$$

where $f(\mathbf{y}|\mathcal{H}_i, \lambda)$ is pdf of the samples vector $\mathbf{Y}$ stored under hypothesis $\mathcal{H}_i \in \{\mathcal{H}_0, \mathcal{H}_1\}$ and under traffic load λ. The presence of a preamble in the L samples is declared (decision $\mathcal{D}_1$) if $\Phi(\mathbf{y})$ is lower than a defined threshold ϕ, whereas the absence of a preamble (decision D_0) is declared otherwise, specifically[2]

$$\Phi(\mathbf{y}) = \frac{f(\mathbf{y}|\mathcal{H}_0, \lambda)}{f(\mathbf{y}|\mathcal{H}_1, \lambda)} \underset{\mathcal{D}_1}{\overset{\mathcal{D}_0}{\gtrless}} \phi \tag{4}$$

Both the correlator and LRT receiver approaches will be evaluated in terms of detection probability p_{d} and of false alarm probability p_{fa}. The value of p_{d} represents the probability to identify the start of a packet given that the complete preamble sequence is present in the observed vector while p_{fa} denotes the probability to erroneously flag the beginning of a packet in the absence of the complete preamble sequence. From this standpoint, we recall that the LRT maximizes the detection probability given a tolerable false alarm rate in the considered setting (H. Van Trees., 2001). As such, it provides a reference benchmark in the remainder of our discussion.

3. Preliminaries

As a preliminary step towards characterizing the behavior of the considered detection strategies, we identify incoming sample sequences that can lead to a false alarm event. Events are enumerated and defined for convenience in Table. 1.

The simplest case, denoted by event $\mathcal{E}_0$, is obtained when the detection window is filled with L noise samples. This occurs whenever the sender queue is empty at the initial observation time, and no new packet is generated for the subsequent $(L-1)T_s$ seconds. In this event, the transmitter does not send any sample within the observed interval. Recalling equation (1), the event of processing a sequence of solely noise samples has probability

$$\Pr\{\mathcal{E}_0\} = \pi_0 \, e^{-\lambda(L-1)T_s}.$$

[2]Please note that is defined the inverse of the conventional threshold.

Table 1: Incoming sample sequences that might cause a false alarm

Event	structure of the L-sample detection window
$\mathcal{E}_0$	L noise samples
$\mathcal{E}_1(k)$	$k \in \{1, \ldots, L-1\}$ noise, $L-k$ preamble samples
$\mathcal{E}_2(m)$	$L-m$ preamble, followed by $m \leq L-1$ payload samples
$\mathcal{E}_3$	L payload samples
$\mathcal{E}_4(m)$	$m \in \{1, \ldots, L-1\}$ payload, $L-m$ noise samples
$\mathcal{E}_5(m,k)$	m payload, k noise, $L-k-m$ preamble samples
$\mathcal{E}_6(m)$	$m \in \{1, \ldots, L-1\}$ payload, $L-m$ preamble samples

Throughout our discussion, we assume that the transmission of a new generated packet can only start at discrete times μT_s, with $\mu \in \mathbb{N}$.

Secondly, a false alarm may be triggered when the observed vector is composed of k noise and $L-k$ noisy preamble samples. Denoting such an event as $\mathcal{E}_1(k)$, we obtain that

$$\Pr\{\mathcal{E}_1(k)\} = \pi_0 \, e^{-\lambda(k-1)T_s} \, (1 - e^{-\lambda T_s}).$$

The factor $\pi_0 e^{-\lambda(k-1)T_s}$ accounts for the source to remain silent during the first k observed samples. Instead, the latter factor indicates the probability of the Poisson process generating at least one new packet during the observation of the $(k-1)$-th noise sample. In this way, the sender initiates a transmission at the k-th observed sample and the rest of the observation windows is filled with a part of the preamble.

Consider now event $\mathcal{E}_2(m)$ with $m \in \{1, \ldots, L-1\}$, which occurs when the receiver observes exactly $L-m$ preamble symbols followed by m payload symbols. This is obtained if the sender is transmitting and the detection window is reading the $(L-m+1)$-th preamble symbol. Observing that the probability of picking a specific symbol within the packet duration P conditioned on having a non-empty queue is uniformly distributed in $\{1, \ldots, P\}$ we have that for any m

$$\Pr\{\mathcal{E}_2(m)\} = \frac{1}{P}(1 - \pi_0).$$

Following a similar reasoning, the probability for the detection window to be completely filled with payload symbols is

$$\Pr\{\mathcal{E}_3\} = \sum_{\ell=L+1}^{P-L+1} \frac{1-\pi_0}{P} = (1-\pi_0)\,\frac{P-2L+1}{P}.$$

where ℓ covers the symbol indexes within the packet in transmission.

A false alarm may also occur when the receiver observes exactly $m \in \{1,\dots,L-1\}$ payload samples followed by noise. Each of these configurations has probability

$$\Pr\{\mathcal{E}_4(m)\} = \frac{\pi_1}{P}\, e^{-\lambda(L-1)T_s}.$$

The event is indeed obtained when the sender is processing the only packet in the queue (factor π_1) and the detection window starts exactly with the $(D-m+1)$-th payload symbol (probability $1/P$). Moreover, no arrival can enter the system for a duration of $L-1$ samples as denoted by the factor $e^{-\lambda(L-1)T_s}$.

Note that, if the last condition is not verified then some preamble symbols of a new generated packet also fall within the detection window. Such event is also a possibly cause of erroneous detection which is identified in Table 1 as $\mathcal{E}_5(m,k)$. Leaning on the same approach, the probability of observing at the receiver a sequence composed of m data, k of noise and $L-k-m$ of preamble symbols can be written as

$$\Pr\{\mathcal{E}_5(m,k)\} = \frac{\pi_1}{P}\, e^{-\lambda(k+m-1)T_s}(1-e^{-\lambda T_s}).$$

Finally, a false alarm can be caused when two consecutive transmissions fall in a detection window. Specifically, we are interested in the possibility of having the last m payload symbols of a first packet followed by the initial $L-m$ preamble symbols of a second one. Such an event occurs in two different situations. The first case is verified if at the moment of the transmission of the $(L-m+1)$-th symbol of the packet in service there is already present one or more packet in the queue. Recalling that π_0 is the probability that the queue system is empty and π_1 is the probability that a packet is on service, then the first event has probability $(1-\pi_0-\pi_1)/P$. The second case is verified if initially there is one packet in the system which is service and one or more packets are generated during the processing of m payload symbols. This occurs with probability $\pi_1(1-e^{\lambda m T_s})/P$. Summing the contributions of the two disjoint events, we have that

$$\Pr\{\mathcal{E}_6(m)\} = \frac{1}{P}(1-\pi_0-\pi_1\, e^{-\lambda m T_s}).$$

Note that the events of false alarm listed in the Table plus the event of observing the entire preamble sequence are capturing all possible configurations that the sliding windows at the receiver in might observe. One can easily confirm the validity of the equation below

$$p_{\mathsf{d}} + \sum_i \Pr\{\mathcal{E}_i\} = 1$$

where p_{d} denotes the probability of correctly detecting the preamble.

4. Correlator Performance

The correlator detection probability can be derived from equation (2). Indeed, when the observed sequence corresponds to the preamble symbols plus noise, the detector output takes the form

$$\begin{aligned} \mathsf{C}_\mu &= \sum_{i=0}^{L-1} (w_i + \mathsf{N}_{\mu+i})\, w_i \\ &= L + \sum_{i=0}^{L-1} w_i\, \mathsf{N}_{\mu+i} \end{aligned}$$

where the addends of the last summation are iid random variables, i.e., $w_i\, \mathsf{N}_{\mu+i} \sim \mathcal{N}(0, \sigma^2)$. Accordingly, we have that

$$\begin{aligned} p_{\mathsf{d}} &= \Pr\left\{ \sum_{i=0}^{L-1} w_i\, \mathsf{N}_{\mu+i} \geq \tau - L \right\} \\ &= Q\left(\frac{\tau - L}{\sigma\sqrt{L}} \right) \end{aligned} \tag{5}$$

where $Q(x) = \frac{1}{2\pi} \int_x^\infty e^{-u^2/2} du$, the Q-function.

Instead, the probability of having a false alarm due to the absence of a complete preamble in the correlator window is given by the activity of the sender. Based on the seven configurations identified in Section 3 and by the law of total probability we have that

$$p_{\mathsf{fa}} = \sum_{\mathcal{E} \in \boldsymbol{\mathcal{E}}} \Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}\} \Pr\{\mathcal{E}\} \tag{6}$$

where the summation spans the set $\boldsymbol{\mathcal{E}}$ of all the events listed in Table 1. The complete expressions of all the involved conditional probabilities are derived and reported in Appendix.

In the remainder of this section we focus instead on a relevant upper bound, which offers insights on how to tackle the involved calculations. We observe that for low packet generation rates, i.e. small values of λ, the pure noise sequences play a key role in causing false alarms, i.e. event $\mathcal{E}_0$. The probability of such event is

$$\begin{aligned}\Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}_0\} &= \Pr\left\{\sum_{i=0}^{L-1} \mathsf{N}_{\mu+i}\, w_i \geq \tau\right\} \\ &\overset{(a)}{=} Q\left(\frac{\tau}{\sigma\sqrt{L}}\right) \end{aligned} \tag{7}$$

where (a) derives from the fact that $\mathsf{N}_{\mu+i}\, w_i$ are iid normal rv with zero mean and σ^2 variance. Instead for higher values of the packet generation λ, the possibility to erroneously flag the start of a packet in the presence of payload symbols becomes more relevant, i.e. event $\mathcal{E}_3$. To treat this case, let us introduce the rv $\Xi(L)$ which indicates the output of the correlation between the preamble and a set of L payload symbols, i.e.

$$\Xi(L) = \sum_{i=0}^{L-1} w_i\, \mathsf{D}_{\mu+i} \tag{8}$$

where each $\mathsf{D}_{\mu+i}$ takes values in $\{-1, 1\}$ with equal probability. $\Xi(L)$ can conveniently be expressed as $\Xi(L) = -L + 2\mathsf{U}$, where the rv U counts the number of symbols in the payload sequence for which $\mathsf{D}_{\mu+i} = w_i$ and has alphabet $\{0, \ldots, L\}$. In view of the iid distribution of the transmitted symbols, $\mathsf{U} \sim \mathrm{Bin}(L, 1/2)$, so that the pmf of $\Xi(L)$ is given by

$$p_{\Xi(L)}(-L + 2\mathsf{u}) = \binom{L}{u} \frac{1}{2^L}\,, \quad u = 0, \ldots, L.$$

Based on this result, we have

$$\begin{aligned}\Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}_3\} &\overset{(a)}{=} \Pr\left\{\Xi(L) + \sum_{i=0}^{L-1} \mathsf{N}_{\mu+i}\, w_i \geq \tau\right\} \\ &\overset{(b)}{=} \frac{1}{2^L} \sum_{u=0}^{L} \binom{L}{u} Q\left(\frac{\tau - (2u - L)}{\sigma\sqrt{L}}\right) \end{aligned} \tag{9}$$

Table 2: Conditional pdf $f(\mathbf{y}|\mathcal{E},\lambda)$ for events in Table 1, LRT

Event	Conditional pdf: $f(\mathbf{y}|\mathcal{E},\lambda)$
$\mathcal{E}_0$	$\prod_{j=0}^{L-1} \varphi(y_j)$
$\mathcal{E}_1(k)$	$\prod_{j=0}^{k-1} \varphi(y_j) \prod_{\ell=k}^{L-1} \varphi(y_\ell - w_{\ell-k})$
$\mathcal{E}_2(m)$	$\prod_{j=0}^{m-1} \varphi(y_j - w_{L-m+j}) \prod_{\ell=m}^{L-1} \frac{1}{2}[\varphi(y_\ell - 1) + \varphi(y_\ell + 1)]$
$\mathcal{E}_3$	$\prod_{j=0}^{L-1} \frac{1}{2}[\varphi(y_j - 1) + \varphi(y_j + 1)]$
$\mathcal{E}_4(m)$	$\prod_{j=0}^{m-1} \frac{1}{2}[\varphi(y_j - 1) + \varphi(y_j + 1)] \prod_{\ell=m}^{L-1} \varphi(y_\ell)$
$\mathcal{E}_5(m,k)$	$\prod_{j=0}^{m-1} \frac{1}{2}[\varphi(y_j - 1) + \varphi(y_j + 1)] \prod_{\ell=m}^{m+k-1} \varphi(y_\ell) \prod_{u=m+k}^{L-1} \varphi(y_u - w_{u-k-m})$
$\mathcal{E}_6(m)$	$\prod_{j=0}^{m-1} \frac{1}{2}[\varphi(y_j - 1) + \varphi(y_j + 1)] \prod_{\ell=m}^{L-1} \varphi(y_\ell - w_{\ell-k})$

where (a) follows from the fact that $\mathsf{Y}_{\mu+i} = \mathsf{D}_{\mu+i} + \mathsf{N}_{\mu+i}$ when is conditioned on $\mathcal{E}_3$, whereas (b) is obtained by applying the law of total probability conditioned on $\Xi(L)$.

Combining (7) and (9), a simple lower bound to the false alarm probability of the correlator, can be obtained from (6) as

$$
\begin{aligned}
p_{\mathsf{fa}} &= \sum_{\mathcal{E}\in\boldsymbol{\mathcal{E}}} \Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}\} \Pr\{\mathcal{E}\} \\
&\geq \Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}_0\} \Pr\{\mathcal{E}_0\} + \Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}_3\} \Pr\{\mathcal{E}_3\} \\
&\geq \pi_0\, e^{-\lambda(L-1)T_s}\, Q\!\left(\frac{\tau}{\sigma\sqrt{L}}\right) + \frac{(1-\pi_0)(P-2L+1)}{2^L\, P} \sum_{u=0}^{L} \binom{L}{u} Q\left(\frac{\tau-(2u-L)}{\sigma\sqrt{L}}\right).
\end{aligned}
\tag{10}
$$

The tightness of the bound obtained in (10) will be discussed in our results, Section 6.

5. Likelihood Ratio Test

The LRT can be conveniently expressed resorting to the set of disjoint events $\boldsymbol{\mathcal{E}}$ that cover the absence of a preamble in the observation window. Specifically, from (4) we obtain that

$$\Phi(\mathbf{y}) \stackrel{(a)}{=} \frac{f(\mathbf{y}, \mathcal{H}_0 \,|\, \lambda)}{\Pr\{\mathcal{H}_0 \,|\, \lambda\}\, f(\mathbf{y} \,|\, \mathcal{H}_1, \lambda)}$$
$$\stackrel{(b)}{=} \frac{\sum_{\mathcal{E}\in\boldsymbol{\mathcal{E}}} f(\mathbf{y}|\mathcal{E}, \lambda)\Pr\{\mathcal{E}\}}{f(\mathbf{y}|\mathcal{H}_1, \lambda)} \underset{\mathcal{D}_1}{\overset{\mathcal{D}_0}{\gtrless}} \phi' \tag{11}$$

where (a) follows from the application of Bayes' rules to the conditional pdf $f(\mathbf{y}|\mathcal{H}_0, \lambda)$, while (b) includes the constant value of the factor $\Pr\{\mathcal{H}_0|\lambda\}$ into ϕ' and applies the law of total probability to the numerator. Instead, the denominator of (11) can be easily derived from the fact that under hypothesis $\mathcal{H}_1$ the components of $\mathbf{Y}$ are i.i.d. with $\mathsf{Y}_i \sim \mathcal{N}(w_i, \sigma^2)$, so we have that

$$f(\mathbf{y}, \mathcal{H}_1|\lambda) = \prod_{j=0}^{L-1} \frac{1}{\sqrt{2\pi}\sigma} e^{-(\mathsf{y}_j - w_j)^2/2\sigma^2}.$$

Let us now consider instead the conditional pdfs $f(\mathbf{y} \,|\, \mathcal{E}, \lambda)$ and we focus on the two events $\mathcal{E}_0$ and $\mathcal{E}_3$. In the first case, the observed sequence is composed only by noise samples, so that

$$f(\mathbf{y} \,|\, \mathcal{E}_0, \lambda) = (\sqrt{2\pi}\sigma)^{-L} \exp(-\sum \mathsf{y}_i^2/(2\sigma^2)).$$

Conversely, when the received window contains only payload samples, then the iid received symbols are $\mathsf{Y}_i \sim \mathcal{N}(d_i, \sigma^2)$. Given that D_i takes values in $\{-1, 1\}$ with probability 1/2, by law of total probability we have that

$$f(\mathsf{y}_i \,|\, \mathcal{E}_3, \lambda) = \frac{1}{2\sqrt{2\pi}\sigma} e^{-(\mathsf{y}_i^2+1)/2\sigma^2} (e^{\mathsf{y}_i/\sigma^2} + e^{-\mathsf{y}_i/\sigma^2}).$$

The pdfs of $\mathbf{Y}$ conditioned on the other events in $\boldsymbol{\mathcal{E}}$ can be calculated following the same approach and their derivation is omitted for brevity. The obtained results for all cases are reported in Table 2, where $\varphi(y)$ denotes the pdf of a standard rv of zero mean and variance σ^2 computed in y.

In the remainder of our discussion we consider instead an approximation of the LRT. By restricting the summation in (11) to the contributions of $\mathcal{E}_0$

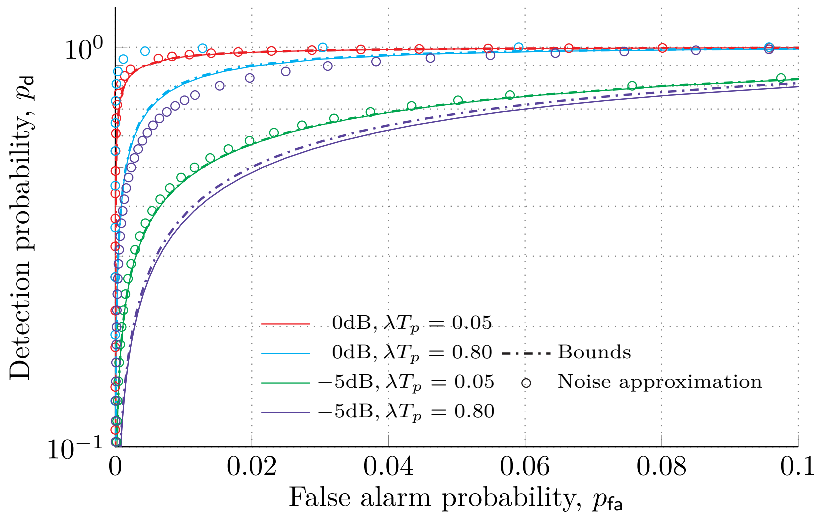

Figure 3: ROC for correlation-based detection for different SNR and traffic intensities. Results of the exact expressions in (5) and (6) are reported by solid lines while dashed lines corresponds to the bound of p_{fa} in (10). Circle markers indicate performance estimated neglecting the presence of data and considering only false alarms due to noise. In all cases, the preamble length is of $L = 16$ symbols while data is $D = 256$ symbols long.

and $\mathcal{E}_3$ and after simple manipulations we have that $\Phi(\mathbf{y})$ can be approximated as follows

$$\Phi(\mathbf{y}) \simeq \frac{f(\mathbf{y}|\mathcal{E}_0, \lambda)\Pr\{\mathcal{E}_0\} + f(\mathbf{y}|\mathcal{E}_3, \lambda)\Pr\{\mathcal{E}_3\}}{f(\mathbf{y}|\mathcal{H}_1, \lambda)}$$

$$\simeq \pi_0\, e^{-\lambda(L-1)T_s + L/(2\sigma^2)}\; e^{-\sum_{j=0}^{L-1} w_j \mathsf{y}_j/\sigma^2} + \frac{(1-\pi_0)(P-2L+1)}{P\,2^L} \prod_{j=k}^{L-1} (1 + e^{2\mathsf{y}_j w_j/\sigma^2}). \tag{12}$$

6. Numerical Results

To study the performance of the considered detection schemes we focus on the receiver operating curves (ROC), reporting the achievable $(p_{\mathsf{fa}}, p_{\mathsf{d}})$ pairs. Having in mind the transmission of short packets in LWPAN systems, we consider throughout our discussion a preamble of $L = 16$ symbols and an overall message length of $(L + D) = 256$ symbols. Specifically, we rely on the preamble sequence

$$p = [1110\,1011\,1001\,0000] \tag{13}$$

originally proposed by the Consultative Committee for Space Data Systems (CCSDS), and specifically designed to perform well for the detection of short packets thanks to its good auto-correlation properties (CCSDS, 2012). In the results presented is assumed a unitary symbol time, i.e. $T_s = 1$.

Let us focus first on Figure 3, which presents the behavior of the correlator for two relevant SNR values SNR= 0 and SNR= -5 dB, under low $\lambda T_p = 0.05$ and high $\lambda T_p = 0.8$ traffic conditions. The results confirm in all configurations that the presented bounds represented by the dashed lines are very tight and thus offer simple and convenient closed-form approximations. Moreover, as expected, better detection capabilities are obtained for higher SNR values. More interestingly, the plot reveals a strong impact of the traffic generation pattern, with higher values of λ resulting in higher false alarm rates. The effect derives from the contribution of erroneous detection decisions caused by portions of the payload such impact becomes more pronounced when packets are injected more frequently on the channel. To further investigate the relevance of this aspect, we report in Fig. 3 also the ROC obtained when the presence payload is neglected and assuming false alarms only induced by pure noise sequence. The circle-markers curves characterize the approach typically used in the literature and often regarded as a guideline for system design. From this standpoint, it is likely that such a strategy leads to a strong overestimate of the detection performance. For example, for an SNR of -5 dB and a target false alarm rate $p_{\mathsf{fa}} = 0.03$ and only considering the effect of noise one would expect a detection probability of $p_{\mathsf{d}} = 0.9$ for $\lambda T_p = 0.8$. However, the additional effect of payload on false alarm affecting practical implementations results in significantly worse performance, with an achievable detection rate of $p_{\mathsf{d}} = 0.55$. This observation not only stresses the importance of considering the traffic profile, but also confirms that the presented analysis is a useful and simple tool for proper dimensioning the system.

To complement our discussion, we report in Fig. 4 the behavior of the LRT-based detection. In this case, p_{d} and p_{fa} were estimated via Montecarlo simulations based on the expressions reported in Section 5. Also in this case, the tightness of the approximation in (12) is confirmed in all SNR and traffic configurations. We furthermore observe that, correlator and LRT

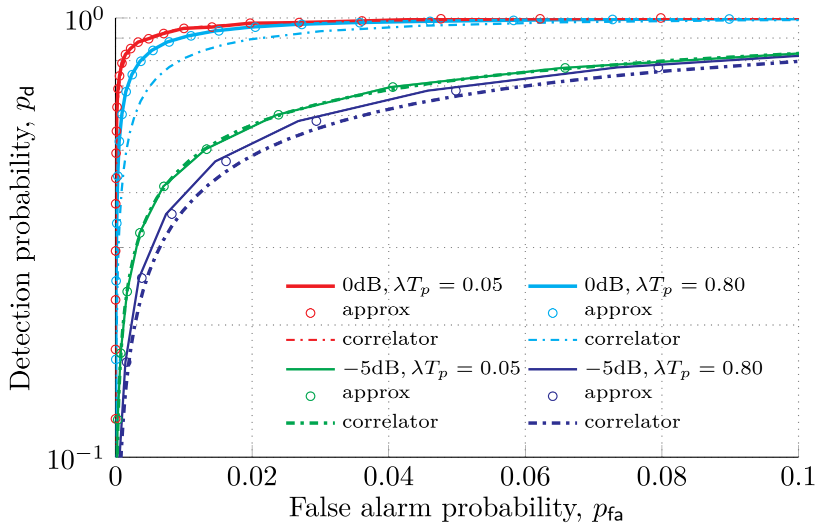

Figure 4: ROC for LRT-based detection for different SNR and traffic intensities. Results of the exact expression leaning on the probabilities given in Table 2 are reported by solid lines and circle markers lean on the approximation in (12). Dashed lines indicates the correlator performance. In all cases, the preamble length is of $L = 16$ symbols while data is $D = 256$ symbols long.

offer comparable performance for low channel load. This result is expected, as correlation is known to be optimal for detection of preambles when false alarms are dominated by noise (H. Van Trees., 2001). On the other hand, the power of the LRT approach to correctly accounting for the presence of payload emerges for larger values of λT_p. Nonetheless, the performance degradation of the correlator is rather contained and counterbalanced by a reduction in complexity, supporting the value of the approach even for higher channel loads.

As final remark, we note that, although selecting a different preamble equence may certainly impact performance, it will not alter the overall trends and key takeaways presented. Indeed, our discussion highlighted that disregarding false alarm events (e.g., $\mathcal{E}_1$ and $\mathcal{E}_3$) can lead to inaccurate estimates even with a well-chosen preamble (13) . These issues can be further magnified in the presence of preambles with poor auto-correlation properties, commonly employed in commercial LPWAN systems (LoRa, 2023; Sigfox, 2023).

7. Conclusions

In this work we studied the performance of packet detection techniques for an AWGN point-to-point link. The sender is modelled as a M/D/1 queue and different traffic generation intensities are considered, ranging from sporadic to frequent transmissions. Exact results are derived for detection and false alarm probability of a correlation-based approach and for the benchmark represented by an LRT detector. In both cases, simple yet tight closed-form approximations were as well obtained. Numerical results highlight the importance of properly capturing all causes of false alarms. In particular, including erroneous detections induced by (parts of) the data portion of incoming packets which is typically neglected in existing literature. The presented trends stress the importance of further studies on the topic, particularly taking into account the presence of multi-user interference.

Appendix

In this appendix we derive the conditional false alarm probabilities

$$\Pr\{\mathsf{C} \geq \tau \,|\, \mathcal{E}\}$$

for the correlation-based technique considering the events reported in Table. 1. Note that, for $\mathcal{E}_0$ and $\mathcal{E}_3$, results are already provided in Sec. 4.

To this aim, let us denote for convenience as

$$\Omega(k) := \sum\nolimits_{i=0}^{k-1} w_i \cdot w_{L-k+i}$$

the value at the output of the correlator when observing k noiseless preamble symbols. Leaning on this, the false alarm probability conditioned on event $\mathcal{E}_1(k)$ can be computed as

$$\begin{aligned}\Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}_1(k)\} &= \Pr\left\{\sum_{i=0}^{k-1} \mathsf{N}_i w_i + \sum_{i=k}^{L-1} (w_{i-k} + \mathsf{N}_i) w_i \geq \tau\right\} \\ &= \Pr\left\{\sum_{i=0}^{L-1} \mathsf{N}_i w_i \geq \tau - \Omega(k)\right\} = Q\left(\frac{\tau - \Omega(k)}{\sigma\sqrt{L}}\right).\end{aligned}$$

Consider instead event $\mathcal{E}_2(m)$, with the observation of $L-m$ preamble samples followed by data. In this case, after simple manipulations

$$\Pr\{\mathsf{C}_\mu \geq \tau | \mathcal{E}_2(m)\} = \Pr\left\{\sum_{i=0}^{L-m-1} w_i\, w_{m+i} + \sum_{i=L-m}^{L-1} \mathsf{D}_{i-L+m}\, w_i + \sum_{i=0}^{L-1} \mathsf{N}_i\, w_i \geq \tau\right\}.$$

Recalling the definition of Ω, as well as of the ancillary r.v. Ξ introduced in (8), the expression can be written as

$$\Pr\{\mathsf{C}_\mu \geq \tau | \mathcal{E}_2(m)\} = \sum_{\xi=-m}^{m} \Pr\left\{\sum_{i=0}^{L-1} \mathsf{N}_i w_i \geq \tau - \xi - \Omega(m)\ \Xi(m) = \xi\right\} p_{\Xi(m)}(\xi)$$

$$= \frac{1}{2^m} \sum_{u=0}^{m} \binom{m}{u} Q\left(\frac{\tau - \Omega(m) - (2u-m)}{\sigma\sqrt{L}}\right).$$

Similarly, when the receiver stores m data samples and $L-m$ noise samples we have

$$\begin{aligned}\Pr\{\mathsf{C}_\mu \geq \tau | \mathcal{E}_4(m)\} &= \Pr\left\{\sum_{i=0}^{m-1} \mathsf{D}_i w_i + \sum_{i=0}^{L-1} \mathsf{N}_i w_i \geq \tau\right\} \\ &= \frac{1}{2^m} \sum_{u=0}^{m} \binom{m}{u} Q\left(\frac{\tau - (2u-m)}{\sigma\sqrt{L}}\right).\end{aligned}$$

Let us now move to event $\mathcal{E}_5(k,m)$, characterized by m data, k noise and $L-k-m$ preamble samples in the observation window. In this case, the false alarm rate can be computed once more conditioning on the r.v. $\Xi(m)$, to obtain

$$\Pr\{\mathsf{C}_\mu \geq \tau \,|\, \mathcal{E}_5(k,m)\} = \Pr\left\{\sum_{i=0}^{m-1} \mathsf{D}_i w_i + \sum_{i=k+m}^{L-1} w_i w_{i-k-m} + \sum_{i=o}^{L-1} \mathsf{N}_i w_i \geq \tau\right\}$$

$$= \frac{1}{2^m} \sum_{u=0}^{m} \binom{m}{u} Q\left(\frac{\tau - \Omega(L-k-m) - (2u-m)}{\sigma\sqrt{L}}\right).$$

Finally, when the observed sequence is composed by m data symbols and $L-m$ preamble symbols, we have

$$\begin{aligned}\Pr\{\mathsf{C}_\mu \geq \tau | \mathcal{E}_6(m)\} &= \Pr\left\{\sum_{i=0}^{m-1} \mathsf{D}_i w_i + \sum_{i=m}^{L-1} w_i\, w_{i-m} + \sum_{i=0}^{L-1} \mathsf{N}_i w_i \geq \tau\right\} \\ &= \frac{1}{2^m}\sum_{u=0}^{m}\binom{m}{u} Q\left(\frac{\tau - \Omega(L-m) - (2u-m)}{\sigma\sqrt{L}}\right).\end{aligned}$$

References

Kang, Jae-Mo, LoRa Preamble Detection With Optimized Thresholds, IEEE Internet of Things Journal 10 2023 6525-6526.

Vangelista, Lorenzo and Cattapan, Alessandro, Start of Packet Detection and Synchronization for LoraWAN Modulated Signals, IEEE Transactions on Wireless Communications 2022 4608-4621.

Abdzadeh-Ziabari, Hamed and Shayesteh, Mahrokh G., Sufficient Statistics, Classification, and a Novel Approach for Frame Detection in OFDM Systems, IEEE Transactions on Vehicular Technology 2013 2481-2495.

Wuerll, Rene and Robert, Joerg and Kilian, Gerd and Heuberger, Gerd, Optimal One-Shot Detection of Preambles with Frequency Offset, 2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS) 2018 1–4.

Robertson, Patrick, Optimal Frame Synchronization for Continuous and Packet Data Transmission, Ph.D. Dissertation, University of Bundeswehr, Munich, Germany 1995.

Consultative Committee for Space Data Systems, TC synchronization and channel coding-summary of concept and rationale, CCSDS Report, CCSDS 230.1-G-2 2012.

LoRa Alliance, The LoRa Alliance Wide Area Networks for Internet of Things, `www.lora-alliance.org`

Sigfox, SIGFOX: The Global Communications Service Provider for the Internet of Things, `www.sigfox.com`.

Abramson, Norman, THE ALOHA SYSTEM: another alternative for computer communications, Fall Joint Computer Conference 37 1977 281–285.

Massey, J., Optimum Frame Synchronization, IEEE Transactions on Communications, 1972 20 115–119.

Chiani, Marco and Martini, Maria G., Analysis of Optimum Frame Synchronization Based on Periodically Embedded Sync Words, IEEE Transactions on Communications, 2007 55 2056–2060.

Chiani, Marco and Martini, Maria G., A lower bound for optimum frame synchronization on AWGN channels, Proc. IEEE VTC, 2005.

Nagaraj, Sumeeth and Khan, Sheehan and Schlegel, Christian and Burnashev, Marat, On Preamble Detection in Packet-Based Wireless Networks, Proc. IEEE Intr. Symp. on Spread Spectrum Techniques and Applications, 2006.

Wuerll, Rene and Robert, Joerg and Kilian, Gerd and Heuberger, Gerd, Optimal One-Shot Detection of Preambles with Frequency Offset, Proc. IEEE ANTS, 2018.

Wuerll, Rene and Robert, Joerg and Kilian, Gerd and Heuberger, Albert, A comparison of methods for detecting preambles with frequency offset at low SNR, Proc. ICCCS, 2017.

Neumüller, Clemens and Robert, Joerg and Heuberger, Albert, SIMO Frame Acquisition of Fragmented Telegrams in Interference Channels, Proc. IEEE WCNC, 2022.

J. Shortle and J. Thompson and D. Gross and C. Harris, Fundamentals of Queueing Theory, Wiley, 1974.

H. Van Trees, Detection, Estmation and Modulation Theory, Wiley, 2001.