Truncated Sequential Non-Parametric Hypothesis Testing Based on Random Distortion Testing
Prashant Khanduri, Dominique Pastor, Vinod Kumar Sharma, Pramod Varshney

To cite this version:
Prashant Khanduri, Dominique Pastor, Vinod Kumar Sharma, Pramod Varshney. Truncated Sequential Non-Parametric Hypothesis Testing Based on Random Distortion Testing. IEEE Transactions on Signal Processing, 2019, 67 (15), pp.4027-4042. 10.1109/TSP.2019.2923140. hal-02173094

HAL Id: hal-02173094
https://imt-atlantique.hal.science/hal-02173094
Submitted on 4 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—In this work, we propose a new algorithm for sequential non-parametric hypothesis testing based on Random Distortion Testing (RDT). The data based approach is non-parametric in the sense that the underlying signal distributions under each hypothesis are assumed to be unknown. Our previously proposed non-truncated sequential algorithm, \textit{Seq}RDT, was shown to achieve desired error probabilities under a few assumptions on the signal model. In this work, we show that the proposed truncated sequential algorithm, \textit{T-Seq}RDT, requires even fewer assumptions on the signal model, while guaranteeing the error probabilities to be below pre-specified levels and at the same time makes a decision faster compared to its optimal fixed-sample-size (FSS) counterpart, \textit{Block}RDT. We derive bounds on the error probabilities and the average stopping times of the algorithm. Via numerical simulations, we compare the performance of \textit{T-Seq}RDT to \textit{Seq}RDT, \textit{Block}RDT, sequential probability ratio test (SPRT) and composite sequential probability ratio tests. We also show the robustness of the proposed approach compared to standard likelihood ratio based approaches.

Index Terms—Truncated sequential testing, non-parametric testing, robust hypothesis testing, random distortion testing (RDT), sequential probability ratio test (SPRT).

I. INTRODUCTION

In today’s world, many applications are characterized by the availability of large amounts of complex-structured data. It is not always possible to fit the data to predefined models or distributions [2], [3]. Model dependent signal processing approaches are often susceptible to mismatches between the data and the assumed model. In cases where the data does not conform to the assumed model, providing sufficient performance guarantees becomes a challenging task [4], [5]. Therefore, it is important to devise methods that are model-independent, robust, provide sufficient performance guarantees for the task at hand and, at the same time, are simple to implement. In this work, we propose one such algorithm: \textit{T-Seq}RDT, for binary hypothesis testing.

Standard binary hypothesis testing problems [6], based on a fixed number of samples, test the null ($\mathcal{H}_0$) versus the alternate ($\mathcal{H}_1$) hypotheses. The decision is usually made through the Bayesian, minimax or Neyman-Pearson frameworks. Such tests are referred to as fixed-sample-size (FSS) tests. However, a majority of decision making problems are inherently sequential in nature, i.e., observations are collected sequentially and are processed one after the other [7]–[10]. In his seminal works [11], [12], Wald proposed his celebrated sequential procedure, namely, the sequential probability ratio test (SPRT) for testing two simple hypotheses. SPRT is optimal in the sense that it makes a decision faster on average, compared to all the procedures including FSS tests achieving the same probabilities of error. However, this optimality is lost in some cases when there is a mismatch between the assumed and true models for the underlying hypotheses to be tested [13]–[15], i.e., SPRTs can have larger stopping times on average compared to FSS tests that achieve the same error probabilities. To avoid these scenarios, a truncated version of SPRT was proposed in [15], where the truncation time was chosen based on the FSS test. However, the error probabilities achieved by truncated SPRTs are usually higher than those yielded by non-truncated SPRTs. In addition, larger truncation times are needed to guarantee error probabilities below predefined levels. The purpose of the present work is to propose a truncated sequential algorithm for non-parametric binary hypothesis testing.

A key motivation in our approach is that SPRT loses its optimality properties if the hypotheses to be tested are composite [5], [12], [16]–[18]. For composite binary hypothesis testing problems, variants of SPRT have been developed. Of particular interest are invariant SPRT (ISPRT), weighted SPRT (WSPRT) and generalized SPRT (GSPRT) [16]. ISPRT relies on the principle of invariance [17], [19] to reduce the composite hypothesis to a simple one, which then makes it possible to apply Wald’s SPRT [12]. However, this reduction imposes strong restrictions on the hypotheses to be tested [16], [20]. On the other hand, WSPRT assigns a suitable weight function to the unknown parameters [16], although it is not always possible to upper bound the probabilities of error and find an appropriate weight function, even in asymptotic regimes. In contrast, GSPRT approximates the likelihood ratio by replacing the unknown parameters in the likelihood by their maximum likelihood (ML) estimates [16], [21], [22]. Various versions of GSPRT have been proposed in the literature with different thresholds [23]–[25] and most of the literature is focused on the design of one-sided tests for testing single parameter families of distributions. Moreover, most algorithms are developed for exponential families of distributions and guarantees are asymptotic, which do not
upper bound the probabilities of error \[16, 23–25]. Importantly, GSPRT based approaches have heavy computational complexity even for simplest of models and, therefore, are difficult to implement online \[26]. In contrast, in this work, we consider a two-sided binary hypothesis testing problem and propose a non-parametric truncated sequential testing algorithm, namely T-SeqRDT, to solve the hypothesis testing problem. T-SeqRDT has the following properties:

- The upper bounds on the probabilities of false alarm (PFA) and missed detection (PMD) are guaranteed to stay below pre-specified levels even in non-asymptotic regimes, which is naturally of practical interest. Moreover, the algorithm is guaranteed to be faster compared to its optimal FSS version, BlockRDT, to achieve the same performance.
- The underlying signal distribution under each hypothesis is assumed to be unknown, and importantly, the results derived do not rely on independence (or i.i.d) assumptions on the observations either. This makes the algorithm robust to mismatches in the distributions of the signals, compared to likelihood ratio based approaches.
- The algorithm is simple in structure with low computational complexity and, therefore, is easy to implement online.

It must be noted that non-parametric sequential hypothesis testing approaches have been considered in the past, with limited to no success, as guaranteeing both PFA and PMD below certain pre-specified levels may not be feasible for such non-parametric sequential testing problems \[5, 17, 18]. The approaches proposed in \[5] are based on approximating the likelihood ratio by employing estimates of the unknown parameters to be tested. These approaches impose restrictive assumptions on these estimates to guarantee robustness and asymptotic optimality when there is a mismatch between the assumed and the true distribution. This is of limited use in practical problems, which are non-asymptotic in nature.

From a general point of view, all the hypothesis testing procedures discussed above are based on likelihood theory. As such, they all assume prior knowledge of the distribution of the observations under each hypothesis, perhaps up to a few unknown parameters, to derive or approximate the likelihood ratios. Likelihood ratio based procedures are limited, because, usually, precise distribution models for the data in play are not available. Moreover, likelihood ratio based tests are not robust to model mismatches \[5, 16, 20, 27]. In addition, as described earlier, many approaches in sequential hypothesis testing make stationarity as well as i.i.d. assumptions on the observed process under each hypothesis \[16, 23]. Such assumptions are questionable in practice and emphasize the need for devising testing approaches that assume little knowledge of the underlying signals to be tested. In this respect, the truncated non-parametric sequential hypothesis test, T-SeqRDT, proposed in this paper and featuring the properties discussed earlier, results from an alternative sequential binary hypothesis testing formulation aimed at avoiding the aforementioned limitations.

To begin with, assume that Y is a one-dimensional observation, with probability distribution parameterized by ξ. Consider a two-sided hypothesis testing problem \(H_0 : ξ = ξ_0\) vs \(H_1 : ξ ≠ ξ_0\). In practice, testing the signal for a precise value of ξ_0 might be too stringent due to measurement errors, environmental fluctuations other than noise and other factors \[27]. Therefore, it is reasonable to allow for some fluctuations around ξ_0 and design the null hypothesis \(H_0\) to test for the signal in the neighborhood of ξ_0. In this respect, we assume that Y is a corrupted observation of the signal to be tested, Ξ, and that Ξ is a random distorted version of ξ_0 with unknown distribution. The hypothesis testing problem then becomes:

\[
H_0 : |Ξ - ξ_0| ≤ τ vs H_1 : |Ξ - ξ_0| > τ
\]

where τ ∈ [0, ∞) represents the distortion. Note that, as discussed earlier, designing an asymptotically optimal GSPRT might not be feasible. However, one can formulate a WSPRT by assuming Ξ to be deterministic and observations to be Gaussian \[12, Chapter 4\], as discussed later in Section \[VI\].

Problem \(\Pi\) was first considered in the form of random distortion testing (RDT) in \[27\], where the signal of interest, Ξ, with an unknown distribution, was embedded in i.i.d. Gaussian noise. The authors showed that the optimal tests (under certain criteria) were simple in design and, at the same time, independent of the signal distributions, thereby did not need the computation of likelihood ratios. To enhance detection performance with multiple samples, the authors extended the RDT formulation to FSS tests, BlockRDT, in \[28\]. The authors generalized the RDT formulation by replacing the signal Ξ, in \(\Pi\), by its empirical mean over time. Although the detection performance improved with the number of samples, the designer had control only over PFA and no control over PMD. To overcome this limitation, the authors in \[29, 30\], extended the RDT formulation to sequential hypothesis testing and proposed SeqRDT. SeqRDT guarantees both PFA and PMD to be below certain levels, while making a decision faster on average. This was achieved by introducing a few key assumptions on the signal along with a buffer. Again, just like standard likelihood ratio based approaches, SeqRDT is not guaranteed to always make a decision faster compared to BlockRDT \[29, 30\]. To avoid such scenarios, we propose a truncated version of SeqRDT, T-SeqRDT, in this paper. We show that, the proposed algorithm requires fewer assumptions on the signal compared to SeqRDT, and at the same time always makes a decision faster compared to its FSS counterpart, BlockRDT. Similar to truncated SPRT \[15\], PFA and PMD of T-SeqRDT can be guaranteed to stay below pre-specified levels by designing larger truncation times. All this can be achieved with T-SeqRDT without the need for a buffer. We provide an analysis of the algorithm and show that it can be an alternative to the likelihood ratio based approaches as discussed earlier, when the signal distributions in play are unknown. The proposed approach is completely data dependent rather than being model dependent. The proposed algorithm is robust to model mismatches and is relatively simple to implement. In addition, the algorithm is able to provide sufficient performance guarantees with very few assumptions on the signal to be tested. Preliminary results were presented in \[1\] without proofs, for a fixed choice of thresholds and without control over the stopping time of the algorithm. In contrast, in the present work, we give the algorithm designer
freedom to choose the thresholds according to the desired performance guarantees. We provide complete analysis of the stopping time and probabilities of error behavior and their trade-offs with the designed thresholds. Moreover, we provide a method to choose the thresholds such that the stopping time is minimized. Summarizing, the main contributions of this work are as follows:

- We extend the RDT formulation for sequential non-parametric hypothesis testing and introduce a new truncated sequential algorithm, T-SeqRDT, to solve the binary hypothesis testing problem.
- We derive bounds on PFA and PMD of T-SeqRDT and show that the designed thresholds can guarantee pre-specified PFA and PMD.
- We analyze the average stopping time of T-SeqRDT and provide insights into the trade-off between the average stopping time and the error probabilities of T-SeqRDT.

In Section II, we formulate the problem. In Section III, we present the FSS algorithm, BlockRDT. In Section IV, we propose the truncated sequential algorithm, T-SeqRDT. In Section V, we derive the bounds on the error probabilities and stopping time of T-SeqRDT. In Section VI, we perform simulations to gain further insights into the behavior of T-SeqRDT. Finally, in Section VII, we conclude the paper.

Notation: All the random variables are defined on the same probability space \((\Omega, \mathcal{F}, \mathbb{P})\), \(M(\Omega, \mathbb{R})\) denotes the set of all real random variables defined on \((\Omega, \mathcal{F})\). Given \(U \in M(\Omega, \mathbb{R})\): 

\[
\mathbb{P}_U(B) = \mathbb{P}[U \in B] \quad \text{with} \quad \{U \in B\} = \{\omega \in \Omega : U(\omega) \in B\}
\]

when \(B\) is a Borel set of \(\mathbb{R}\). A domain \(B\) of \(U\) is any Borel set \(B\) of \(\mathbb{R}\) such that \(\mathbb{P}_U(B) = 1\). Given \(\xi \in \mathbb{R}\) and \(\sigma \in (0, \infty)\), \(Z \sim \mathcal{N}(\xi, \sigma^2)\) implies \(Z\) is Gaussian distributed with mean \(\xi\) and variance \(\sigma^2\). The Generalized Marcum Function \(Q_\frac{1}{2}\) with order 1/2 is denoted by \(Q_\frac{1}{2}(\cdot, \eta)\) and Remark V.3.

Let \(\Xi = (\Xi_n)_{n \in \mathbb{N}}\) be an element of \(M(\Omega, \mathbb{R})^\mathbb{N}\). This discrete time random process models the random mixture of a distorted signal of interest and possible interferences. Standard hypothesis testing approaches assume that the random process under null hypothesis \(H_0\) is generated from an unknown underlying joint distribution \(\mathbb{P}_0\), i.e., \(\Xi = (\Xi_n)_{n \in \mathbb{N}} \sim \mathbb{P}_0\), and under alternate hypothesis \(H_1\), by any underlying, yet unknown, arbitrary joint distribution, other than \(\mathbb{P}_0\), i.e., \(\Xi = (\Xi_n)_{n \in \mathbb{N}} \sim \mathbb{P}_0\). No assumption is made on the stationarity or the distribution of \(\Xi = (\Xi_n)_{n \in \mathbb{N}}\). In this respect, the samples \(\Xi_n\) are not necessarily i.i.d. As pointed out in \(29, 30\), this problem is difficult to tackle as very little or no knowledge of the underlying signal distributions is assumed under both hypotheses; thereby, likelihood ratio based tests (SPRT or GSPRT) are not suitable for such problems. As an alternative to likelihood ratio based approaches, we propose tests based on RDT \(27\), where we associate a non-parametric distance related criterion with each hypothesis which is independent of the distributions of the actual hypotheses. This non-parametric criterion serves as a surrogate to the actual hypotheses to be tested. Next, we present the model in more detail.

We assume that \(\Xi\) is observed in additive and independent Gaussian noise \(X = (X_n)_{n \in \mathbb{N}}\). The observation process is \(Y = (Y_n)_{n \in \mathbb{N}}\) such that \(Y_n = \Xi_n + X_n\) for all \(n \in \mathbb{N}\), and we write \(Y = \Xi + X\). In our formulation, \(\Xi\) models the distortion around a fixed known and deterministic model \(\xi_0\). We, however, expect that, for \(N\) large enough, the empirical mean \(\bar{\Xi}\) remains close to \(\xi_0\) under \(H_0\) and drifts significantly away from \(\xi_0\) under \(H_1\). Using standard terminology in statistical inference, we say that this problem is the testing of the null hypothesis — a random event, actually \(H_0: |\bar{\Xi}_N - \xi_0| > \tau\) against the alternate hypothesis (event) \(H_1: |\bar{\Xi}_N - \xi_0| > \tau\), on the basis of observation \(Y\). The hypothesis testing problem is therefore given as:

\[
\begin{align*}
\text{Observation} : \quad & Y = \Xi + X \in M(\Omega, \mathbb{R})^\mathbb{N} \\
\text{with} : \quad & X_1, X_2, \ldots \overset{\text{iid}}{\sim} N(0, 1), \\
\Xi \text{ and } X \text{ are independent.} \\
\exists N_0 \in \mathbb{N} : \quad & H_0 : \forall N \geq N_0, \quad 0 \leq |\bar{\Xi}_N - \xi_0| \leq \tau \quad (a.s) \\
& H_1 : \forall N \geq N_0, \quad \tau < |\bar{\Xi}_N - \xi_0| \leq \tau_H \quad (a.s)
\end{align*}
\]

where, \(\tau \in (0, \infty)\) is the tolerance and \(\tau < \tau_H < \infty\). Note, that the above hypothesis testing model is the same as the BlockRDT model \(28\) for a fixed sample size, \(N\) (see Section III). Here, \(N_0\) and the tolerances \(\tau\) and \(\tau_H\) are known \textit{a priori} based on some prior knowledge (or experience) about the signal. The algorithms based on formulation \(5\) have already been used for biomedical signal processing applications, specifically for the detection of Auto-positive end expiratory pressure (Auto-PEEP) \(33\). Moreover, for illustration purposes, below we give a simple example where formulation \(5\) can be easily used.

---

1This knowledge can follow from machine learning training procedures or be based on some statistical knowledge of the signal. Discussion of these procedures is beyond the scope of this work.
Example 1 (Bounded regime testing): Given $\xi \in \mathbb{R}$ and $h \in [0, \infty)$, we say $\Xi$ follows the (bounded) regime $(\xi, h)$ and write $\Xi \sim (\xi, h)$ if, for any $N \in \mathbb{N}$, $|\langle \Xi \rangle_N - \xi| \leq h$. A sufficient condition for $\Xi \sim (\xi, h)$ is that $|\langle \Xi \rangle_N - \xi| \leq h$ (a.s) for any $n \in \mathbb{N}$. Suppose $\Xi$ satisfies either $H_0^* : \Xi \sim (\xi_0, h_0)$, with $(\xi_0, h_0)$ known or $H_1^* : \Xi \sim (\xi_1, h_1)$, with $(\xi_1, h_1)$ possibly unknown.

Say that the regimes $(\xi_0, h_0)$ and $(\xi_1, h_1)$ are separate if $|\langle \xi_1 - \xi_0 \rangle| \geq h_0 + h_1$. When $(\xi_0, h_0)$ and $(\xi_1, h_1)$ are separate, testing $H_0^*$ against $H_1^*$ when we observe $Y$ is the particular problem

\[
|\langle \xi \rangle_N - \xi_0| \geq h_0 + h_1, \quad \tau_H \geq |\langle \xi_1 - \xi_0 \rangle| + h_1
\]

Moreover, as shown later for such problems, we can design optimal FSS tests as well as sequential tests guaranteeing PFA and PMD below pre-specified levels and that are, at the same time, faster than the optimal FSS tests. Importantly, these tests guarantee performance irrespective of the underlying signal distributions.

Remark 1: Note that the RDT framework of (1) is the same as that given in [5] for $N = 1$. The formulation in [5] generalizes the RDT framework of (1) for testing with multiple samples, i.e., for the FSS test, BlockRDT, and sequential hypothesis testing approaches. An alternative testing problem would be to use $|\langle \Xi \rangle_N - \xi_0|$ in (5) instead of $|\langle \Xi \rangle_N - \xi_0|$ in [5]. This would allow a larger class of distortions. However, designing such a test would require stronger assumptions of $|\langle \Xi \rangle_N - \xi_0| \leq \tau$ under $\mathcal{H}_0$ and $|\langle \Xi \rangle_N - \xi_0| > \tau$ under $\mathcal{H}_I$ for all $n \in \mathbb{N}$ in comparison to the condition of [5], where introducing $N_0$ in [5] gives the designer the flexibility to design the testing problem for models when the condition $|\langle \Xi \rangle_N - \xi_0| \leq \tau$ (resp. $|\langle \Xi \rangle_N - \xi_0| > \tau$) might not hold true for smaller values of $N \in \mathbb{N}$ under $\mathcal{H}_0$ (resp. $\mathcal{H}_I$).

Following the standard terminology [16] with a slight change of notation, we define a sequential test for the binary hypothesis testing (5) as a pair $(T, \mathcal{D}_{N_0})$, where $T$ is the stopping time and $\mathcal{D}_{N_0}$ is a decision rule taking values in $\{0, 1, \infty\}$ such that, for each $1 \leq N \leq T$:

\[
\mathcal{D}_{N_0}(N) = \begin{cases} 
0 & \mathcal{H}_0 \text{ is accepted} \\
1 & \mathcal{H}_I \text{ is accepted} \\
\infty & \text{repeat the test with } N + 1 \text{ samples.}
\end{cases}
\]

Further, $T$ is defined as

\[
T = \inf \{ N \in \mathbb{N} : N \leq N_0 + W^* - 1, \mathcal{D}_{N_0}(N) \neq \infty \}.
\]

where the condition $N \leq N_0 + W^* - 1$ guarantees that $T \leq N_0 + W^*$ and we refer to $W^*$ as the truncation window. Note that $W^* = \infty$ for non-truncated sequential procedures. It is also worth noticing that FSS tests are particular cases of tests $(T, \mathcal{D}_{N_0})$, with stopping time being a deterministic constant $T = N$ and $\mathcal{D}_{N_0}$ valued in $\{0, 1\}$. Given two specified levels $\alpha$ and $\beta$ in $(0, 1/2)$, we define the class of tests:

\[
\mathcal{C}(\alpha, \beta) = \{(T, \mathcal{D}_{N_0}) : \sup_{\mathcal{H}_0} \mathbb{P}_{FA}(\mathcal{D}_{N_0}) \leq \alpha, \sup_{\mathcal{H}_I} \mathbb{P}_{MD}(\mathcal{D}_{N_0}) \leq \beta\}
\]
the same empirical mean must yield the same decision for $T$. [Constant conditional power] means that $T$ should not yield different results for different distributions of $|(\Xi_N - \xi_0)|$, conditioned on $|(\Xi_N - \xi_0)| = \rho$.

Let the class of all BlockRDT-coherent tests with level $\gamma$ be denoted by $\mathcal{K}_\gamma$. This class can be partly pre-ordered as follows: given $T, T' \in \mathcal{K}_\gamma$, write that $T \preceq T'$ if, for any $\Xi \in \mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}}$, (i) $T$ and $T'$ satisfy [Constant conditional power] on the same domain $\mathbb{B}$ and (ii) for all $\rho \in \mathbb{B} \cap (\tau, \infty)$,

$$P[T(Y) = 1 | |(\Xi_N - \xi_0)| = \rho] \leq P[T'(Y) = 1 | |(\Xi_N - \xi_0)| = \rho].$$

According to [28], [34], the $N$-dimensional test defined for every $x \in \mathbb{R}^N$ by:

$$T_{N, \gamma}(x) = \begin{cases} \{0 \text{ if } |x - \xi_0| \leq \lambda_\gamma(\tau \sqrt{N})/\sqrt{N} \\ \{1 \text{ otherwise.} \end{cases}$$

(11)

where $\lambda_\gamma(\sqrt{N})/\sqrt{N}$ is defined using (4), is maximal in $\mathcal{K}_\gamma$: for any $T \in \mathcal{K}_\gamma$, $T \preceq T_{N, \gamma}$. Let the PFA and PMD of $T_{N, \gamma}$ for BlockRDT be denoted by $\mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma)$ and $\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma)$, respectively. We have the following proposition [28], [29].

**Proposition 3.1:** For any $\gamma \in (0, 1)$ and $\tau \geq 0$, we have:

$$\begin{align*} 
Q_2 \left(0, \lambda_\gamma(\tau \sqrt{N})/\sqrt{N} \right) &\leq \mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma) \leq 1 \\
1 - Q_2 \left(\tau \sqrt{N}, \lambda_\gamma(\tau \sqrt{N}) \right) &\leq \mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma) \leq 1 - \gamma 
\end{align*}$$

According to the above proposition, although being optimal for BlockRDT, $T_{N, \gamma}$ controls $\mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma)$ efficiently but has no control over $\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma)$.

This implies that, without further assumption and for any $\gamma \in (0, 1)$, BlockRDT cannot belong to the class $\mathcal{C}(\alpha, \beta)$ (with $\mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma)$ replaced by $\mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma)$ and $\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma)$ in (5), respectively) when $\alpha, \beta \in (0, 1/2)$. However, with Assumption 2.1 the next result implies that we can control $\mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma)$ such that BlockRDT is in $\mathcal{C}(\alpha, \beta)$.

**Proposition 3.2:** For any $\gamma \in (0, 1)$, $\mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma)$ and $\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma)$ are bounded under Assumption 2.1 as:

$$\begin{align*} 
\mathbb{P}^{\text{FA}}B_{\text{RDT}}(N, \gamma) &\leq \gamma, \\
\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma) &\leq 1 - Q_2 \left(\tau \sqrt{N}, \lambda_\gamma(\tau \sqrt{N}) \right) 
\end{align*}$$

and the upper bound on $\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma)$ decreases to 0 with $N$.

**Proof:** The bound follows from Lemma A.1 and Assumption 2.1. The upper-bound on $\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma)$ decreases with $N$ as a consequence of Lemma B.3.

Proposition 3.2 implies that for $\gamma = \alpha$ and a sufficiently large $N$ such that the bound on $\mathbb{P}^{\text{MD}}B_{\text{RDT}}(N, \gamma)$ is below $\beta$, BlockRDT is in $\mathcal{C}(\alpha, \beta)$. Since this $N$ might be very large in practice, we introduce a novel truncated sequential algorithm, T-SeqRDT, to control the number of samples.

**IV. Algorithm:** T-SeqRDT

In this section, we propose T-SeqRDT. In T-SeqRDT, if no decision has been reached until a specified time, the decision will be forced using BlockRDT [28], since Proposition 3.2 guarantees that we can attain arbitrarily small PMD for a bounded PFA. The decision variable $\mathcal{D}_{N_0}(N)$ for T-SeqRDT is defined as:

$$\begin{align*} 
\mathcal{D}_{N_0}(1) &= \mathcal{D}_{N_0}(2) = \ldots = \mathcal{D}_{N_0}(N_0 - 1) = \infty, \\
&\text{for } N_0 \leq N < N_0 + W^*, \\
\mathcal{D}_{N_0}(N) &= \begin{cases} 
0 & \text{if } |(Y)|_N - \xi_0| < \lambda_{L}(N) \\
1 & \text{if } |(Y)|_N - \xi_0| > \lambda_{H}(N) \\
\infty & \text{if } \lambda_{L}(N) < |(Y)|_N - \xi_0| < \lambda_{H}(N) \end{cases} \\
&\text{for } N = N_0 + W^*, \\
\mathcal{D}_{N_0}(N) &= \begin{cases} 
0 & \text{if } |(Y)|_N - \xi_0| \leq \lambda_{B-RDT}(N) \\
1 & \text{if } |(Y)|_N - \xi_0| > \lambda_{B-RDT}(N) \end{cases} 
\end{align*}$$

(12)

with decisions taken according to (6) and the stopping time $T$ defined in (7). At time instant $N = N_0 + W^*$, with $W^* \in \mathbb{N}$, the decision is made using BlockRDT, if a decision has not been made until then. Recall that $W^*$ is defined as the truncation window. The three thresholds $\lambda_{L}(N)$, $\lambda_{H}(N)$ and $\lambda_{B-RDT}(N)$ must be designed jointly so as to guarantee that T-SeqRDT is in $\mathcal{C}(\alpha, \beta)$. In any case, $\lambda_{H}(N)$ and $\lambda_{L}(N)$ must be such that $\lambda_{L}(N) < \lambda_{H}(N)$. Moreover, we want a decision faster compared to BlockRDT, the optimal FSS counterpart of T-SeqRDT. The thresholds are chosen with respect to these constraints.

Another sequential algorithm, SeqRDT, was proposed in [29], [30] for solving (5). SeqRDT was shown to belong to class $\mathcal{C}(\alpha, \beta)$ with the help of a buffer along with the upper and lower thresholds respectively defined as:

$$\lambda_{\alpha}(\tau \sqrt{N})/\sqrt{N} \text{ and } \lambda_{1-\beta}(\tau \sqrt{N})/\sqrt{N},$$

(13)

to control both PFA and PMD. Along with these thresholds, the buffer size was designed using $\tau^-$ and $\tau^+$ along with $\tau$ and $\tau_H$ defined in (5), where the meaning of $\tau^-$ is recalled in Remark 2 below.

**Remark 2:** SeqRDT [29], [30] imposed stricter conditions on the signal compared to T-SeqRDT. Beyond Assumption 2.1, it was assumed in the SeqRDT framework that under $\mathcal{H}_0$, for all $N \geq N_0$, $|(\Xi_N - \xi_0)| \leq (a-s)$ with $\tau^- \in (0, \tau)$. Therefore, SeqRDT requires more parameters than T-SeqRDT. In addition, performance bounds were guaranteed by SeqRDT via the use of a buffer. The buffer size was selected using $\tau^-$ and $\tau^+$ along with $\tau$ and $\tau_H$ defined in (5). In contrast, T-SeqRDT does not need to know $\tau^-$ or even $\tau_H$. It requires the knowledge of $\tau$ and $\tau^+$ only to guarantee performance.

**A. Thresholds**

SeqRDT [29], [30] is designed to belong to class $\mathcal{C}(\alpha, \beta)$ via the thresholds (13) and thanks to the use of a buffer. T-SeqRDT eliminates the need for this buffer, while being in $\mathcal{C}(\alpha, \beta)$. In view of the similarity between the T-SeqRDT statistic in (12) to that of BlockRDT in (11), we define the thresholds similar in structure to those of BlockRDT. The thresholds $\lambda_{H}(N)$, $\lambda_{L}(N)$ and $\lambda_{B-RDT}(N)$ are designed as:

$$\begin{align*} 
\lambda_{H}(N) &= \lambda_{H}(N, w_H) = \lambda_{w_H}(\tau \sqrt{N})/\sqrt{N} \\
\lambda_{L}(N) &= \lambda_{L}(N, w_L) = \lambda_{1-\beta}(\tau \sqrt{N})/\sqrt{N} \\
\lambda_{B-RDT}(N) &= \lambda_{B-RDT}(N, w_{B-RDT}) = \lambda_{w_{B-RDT}}(\tau \sqrt{N})/\sqrt{N}, 
\end{align*}$$

(14)
where $w_H, w_L$ and $w_{BH}$ give the algorithm designer control over these thresholds and are equal to or greater than 1. This constraint is necessary to ensure that T-SeqRDT is a valid sequential test, by guaranteeing that $\lambda_L(N, w_L) < \lambda_H(N, w_H)$, as shown in Proposition 4.1 below. In addition, the parameters $w_H, w_L$ and $w_{BH}$ must appropriately be chosen so as to guarantee that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$. To this end, we study the properties of the thresholds \cite{14} and establish that they satisfy suitable properties for T-SeqRDT.

Proposition 4.1: $w_L \geq 1$ and $1 \leq w_{BH} \leq w_H$ given:

(i) $\lambda_L(N, w_L) < \lambda_{BRDT}(N, w_{BH}) < \lambda_H(N, w_H)$, $\forall N \in \mathbb{N}$.

(ii) the thresholds $\lambda_H(N, w_H)$ and $\lambda_{BRDT}(N, w_{BH})$ are decreasing in $N \in \mathbb{N}$ and upper bounded by $\tau$.

(iii) For $N$ large enough, the threshold $\lambda_L(N, w_L)$ is increasing in $N$ and upper bounded by $\tau$.

(iv) all the thresholds approach $\tau$ as $N$ increases:

$$\lim_{N \to \infty} \lambda_H(N, w_H) = \lim_{N \to \infty} \lambda_{BRDT}(N, w_{BH}) = \lim_{N \to \infty} \lambda_L(N, w_L) = \tau.$$  

Proof: Since $\alpha, \beta \in (0, 1/2)$, $0 < \frac{w_H}{w_H} < \frac{w_L}{w_L} < 1/2 < 1 - \frac{w_L}{w_H}$. Thus the proof of (i) follows from Lemma B.4 Statements (ii) and (iii) follow from Lemmas B.5 and B.6 respectively. The proof of (iv) derives from Lemma B.1.

As discussed earlier, Proposition 4.1 ensures $\lambda_L(N, w_L) < \lambda_H(N, w_H)$, which is made possible by the assumption that $w_H, w_L$ and $w_{BH} \geq 1$. Moreover, all the thresholds tend to $\tau$ as $N$ increases, which intuitively implies that the chance of making a decision should be higher for larger $N$.

Now, we analyze the behavior of the thresholds with respect to parameters $w_H, w_L$ and $w_{BH}$, when $N$ is fixed.

Proposition 4.2:

(i) $\lambda_H(N, w_H)$ increases when $w_H$ increases,

(ii) $\lambda_L(N, w_L)$ decreases when $w_L$ increases,

(iii) $\lambda_{BRDT}(N, w_{BH})$ increases when $w_{BH}$ increases,

(iv) $\lim_{w_H \to \infty} \lambda_H(N, w_H) = \infty$ and $\lim_{w_{BH} \to \infty} \lambda_L(N, w_L) = 0$.

Proof: The proof of (i), (ii) and (iii) follows from Lemma B.4 Statement (iv) follows from (ii) and the fact that the Marcum function $Q_1$ is a complementary cdf.

According to Proposition 4.2, $\lambda_H(N, w_H)$ and $\lambda_L(N, w_L)$ grow further away as $w_H$ and $w_L$ increase. Therefore, thresholds designed with higher values of $w_H$ and $w_L$ should provide better $\mathbb{P}_{FA}(D_{BH})$ and $\mathbb{P}_{MD}(D_{BH})$ performance compared to thresholds tuned with lower $w_H$ and $w_L$ values, but at the expense of higher stopping times. For SeqRDT \cite{29, 30}, the error probabilities were controlled via the buffer and no control over the stopping time was provided. For T-SeqRDT, the control over the error probabilities is achieved by choosing the parameters $w_H, w_L$ and $w_{BH}$ so as to move the thresholds away from or closer to each other. This gives the designer control over the average stopping time as well. This will be discussed in more detail later.

B. Truncation window

The goal of T-SeqRDT is to make a decision faster compared to its FSS counterpart, BlockRDT, while providing sufficient performance guarantees. Thus, it makes sense to base the choice of the truncation window $W^*$ on BlockRDT as follows. For the threshold $\lambda_{BRDT}(N, w_{BH})$ given in \cite{14}, Proposition 3.2 induces that $\mathbb{P}_{FA}^{BRDT}(N, \alpha/w_{BH})$ is always upper bounded by $\alpha/w_{BH}$ and hence by $\alpha$ as $w_{BH} \geq 1$. Moreover, the upper bound on $\mathbb{P}_{MD}^{BRDT}(N, \alpha/w_{BH})$ is a decreasing function of $N$. We thus propose to choose $W^* = W^*(w_{BH}, w_{BL})$ as:

$$W^* = W^*(w_{BH}, w_{BL}) = \min \left\{ \forall N \in \mathbb{N} : 1 - Q_1 \left( \sqrt{N_0 + W, \lambda_{BH}}(\tau/\sqrt{N_0 + W}) \right) \leq \frac{\beta}{w_{BL}} \right\}$$

with $w_{BL} \geq 1$.

Remark 3: Note from Proposition 3.2 that BlockRDT with $\lambda_{BRDT}(N, w_{BH})$ given in \cite{14} for $N$ number of samples $N = N_0 + W^*$ is in $\mathcal{C}(\alpha/w_{BH}, \beta/w_{BH})$.

The parameters $w_H$ and $w_L$ defined earlier are used to control the upper and the lower thresholds, respectively, via \cite{14}. On the other hand, the parameters $w_{BH}$ and $w_{BL}$ control the truncation window, $W^*(w_{BH}, w_{BL})$, defined in (15) and the assumption $w_{BL} \geq 1$ is required to make sure that the PMD of T-SeqRDT stays below $\beta$ (see Theorem 5.2 below). All the thresholds along with the truncation window, which are thus controlled by $w_H, w_L, w_{BH}$ and $w_{BL}$, govern the performance of T-SeqRDT. Therefore, we next analyse the behavior of $W^*(w_{BH}, w_{BL})$ with $w_{BH}$ and $w_{BL}$ so that $w_{BL}, w_H, w_L$ and $w_{BH}$ can be fixed to guarantee that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$.

Proposition 4.3: We have

(i) For fixed $w_{BL}$, $W^*(w_{BH})$ does not decrease;

(ii) For fixed $w_{BH}$, $W^*(w_{BH}, w_{BL})$ does not decrease;

Proof: For any $w_{BH} \geq 1$ and any $W \in \mathbb{N}$, set:

$$UB_{MD}^{BRDT}(w_{BH}, W) = 1 - Q_1 \left( \sqrt{N_0 + W, \lambda_{BH}}(\tau/\sqrt{N_0 + W}) \right)$$

For any $w_{BL} \geq 1$,

$$W^*(w_{BH}, w_{BL}) = \min \left\{ \forall N \in \mathbb{N} : UB_{MD}^{BRDT}(w_{BH}, W) \leq \frac{\beta}{w_{BL}} \right\}$$

with:

$$A(w_{BH}, w_{BL}) = \left\{ \forall N \in \mathbb{N} : UB_{MD}^{BRDT}(w_{BH}, W) \leq \frac{\beta}{w_{BL}} \right\}$$

Proof of (i): Consider $w_{BH} \leq w_{BH}'$. According to Lemmas 1.1 and B.4, we have:

$$UB_{MD}^{BRDT}(w_{BH}', W) \leq UB_{MD}^{BRDT}(w_{BH}, W)$$

Therefore, from (16) and (17), we have $A(w_{BH}', w_{BL}) \subseteq A(w_{BH}, w_{BL})$ and thus $W^*(w_{BH}', w_{BL}) \geq W^*(w_{BH}, w_{BL})$.

Proof of (ii): Fix $w_{BH}$. If $w_{BL} \leq w_{BL}'$, then $\frac{\beta}{w_{BL}} \leq \frac{\beta}{w_{BL}}$. This implies that $A(w_{BH}, w_{BL}') \subseteq A(w_{BH}, w_{BL})$.

Hence the result.

Proposition 4.3 tells us that the smaller the required PFA and PMD for truncation by BlockRDT, the larger the truncation window for T-SeqRDT, which is natural. This will lead to the trade-off pinpointed in the next section between this truncation window and the error probabilities of T-SeqRDT. In addition, the choice of the truncation window using BlockRDT will allow for easier comparison between T-SeqRDT and BlockRDT.

Remark 4: Assumption 2.1 is instrumental in choosing an appropriate truncation window $W^*$ for T-SeqRDT (see Proposition 3.2 and (15)). But, if $W^*$ is known a priori, i.e., it is available via some preliminary training procedure or prior
experience, Assumption 2.1 is not needed, while the algorithm will still achieve the same performance. ■

Our next goal is to choose the appropriate thresholds \( \{ \alpha, \beta \} \), that such that T-SeqRDT is in \( \mathcal{F}(\alpha, \beta) \). We proceed by noticing that \( \{ \alpha, \beta \} \) and Proposition 4.3 show that this question is equivalent to choosing appropriate values of \( w_H, w_L, w_{BH} \) and \( w_{BL} \).

**V. ANALYSIS**

We calculate bounds on the PFA and PMD of T-SeqRDT. These bounds are used to derive values for \( w_H, w_L, w_{BH} \) and \( w_{BL} \), that guarantee the required performance. Then, we study the average stopping time. Finally, we discuss the relationship between the error probabilities and the average stopping time.

**A. False alarm and missed detection probabilities**

Since closed form expressions for \( P_{\text{FA}} (D_N) \) and \( P_{\text{MD}} (D_N) \) cannot be derived, we instead calculate upper and lower bounds on these error probabilities, for the thresholds \( \{ \alpha, \beta \} \). These bounds provide useful insights into the behavior of T-SeqRDT. We begin with lower bounds.

**Theorem 5.1 (Lower-bounds on \( P_{\text{FA}} (D_N) \) & \( P_{\text{MD}} (D_N) \)):**

\[
\begin{align*}
P_{\text{FA}} (D_N) & \geq Q_{\frac{1}{2}} \left( 0, \lambda_H/w_H (\tau \sqrt{N}) \right), \\
P_{\text{MD}} (D_N) & \geq 1 - Q_{\frac{1}{2}} \left( \tau H \sqrt{N}, \lambda_1 - \beta/w_L (\tau \sqrt{N}) \right),
\end{align*}
\]

**PROOF:** Since \( [D_N(T) = 1] \supseteq [D_N (N_0) = 1] \), \( \{ \alpha, \beta \} \) implies that, under \( \mathcal{H}_0 \):

\[
\begin{align*}
P_{\text{FA}} (D_N) & \geq P \left[ D_N (N_0) = 1 \right] \\
& \overset{(a)}{\geq} P \left[ Q_{\frac{1}{2}} \left( \sqrt{N} \lambda_H (N), \sqrt{N} \lambda_H (N_0, w_H) \right) \right] \\
& \overset{(b)}{\geq} Q_{\frac{1}{2}} \left( 0, \lambda_H/w_H (\tau \sqrt{N}) \right),
\end{align*}
\]

where \( (a) \) follows from Lemma A.1 \( (b) \) from \( \{ \alpha, \beta \} \) and the fact that under \( \mathcal{H}_0 \), \( 0 \leq |(\Xi)_N - \xi_0| \leq \tau \) (as). Similarly, consider the event \( [D_N(T) = 0] \) and follow the same procedure as above to get the lower bound for \( P_{\text{MD}} (D_N) \).

Although the lower bounds play no role in designing the thresholds, note that they decrease with \( N_0 \) and approach 0 as \( N_0 \to \infty \), which follows from Lemma B.3 and B.2.

**Theorem 5.2 (Upper-bounds on \( P_{\text{FA}} (D_N) \) & \( P_{\text{MD}} (D_N) \)):**

\[
\begin{align*}
P_{\text{FA}} (D_N) & \leq UB_{\text{FA}} \left( \frac{w_H^*}{w_H} + \frac{1}{w_{BH}} \right) \alpha, \\
P_{\text{MD}} (D_N) & \leq UB_{\text{MD}} \left( \frac{w_L^*}{w_L} + \frac{1}{w_{BL}} \right) \beta,
\end{align*}
\]

where \( UB_{\text{FA}} \) and \( UB_{\text{MD}} \) are given in \( \{ \alpha, \beta \} \) and \( \{ \alpha, \beta \} \), respectively, and \( W^* = W^*(w_{BH}, w_{BL}) \).

**PROOF:** We have

\[
[D_N(T) = 1] = [D_N (N_0) = 1] \bigcup_{N = N_0 + 1} \left[ \left[ D_N (N) = 1 \right] \cap [D_N (K) = \infty, \forall K \text{ s.t. } N_0 \leq K \leq N - 1] \right].
\]

Since these events are disjoint, we have

\[
P [D_N (T) = 1] = P [D_N (N_0) = 1] + \sum_{N = N_0 + 1} P \left[ \left[ D_N (N) = 1 \right] \cap [D_N (K) = \infty, \forall K \text{ s.t. } N_0 \leq K \leq N - 1] \right]
\]

\[
\geq P [D_N (N_0) = 1] + \sum_{N = N_0 + 1} P [D_N (N) = 1],
\]

and follow from the Frechet inequality. We bound each individual probability on the right hand side (rhs) of \( \{ \alpha, \beta \} \) and use the fact that under \( \mathcal{H}_0 \), \( 0 \leq |(\Xi)_N - \xi_0| \leq \tau \) and Lemma 1.1. (c) comes from \( \{ \alpha, \beta \} \).

Second, for \( N = N_0 + W^* \), we have \( \mathcal{H}_0 \):

\[
P [D_N (N) = 1] = P_{\text{FA}} [\mathcal{D}_N (N) = 1] \overset{(a)}{\leq} \alpha/w_{BH},
\]

where \( (a) \) follows from Proposition 3.2 and \( \{ \alpha, \beta \} \). Now, for all \( N_0 \leq K \leq N_0 + W^* - 1 \), we have:

\[
P [D_N (K) = \infty] = P [\left( Q_{\frac{1}{2}} \left( \sqrt{N} |(\Xi)_N - \xi_0|, \sqrt{N} \lambda_H (N, w_H) \right) \right) \leq\overset{(a)}{\leq} \alpha/w_{BH},
\]

where \( (a) \) follows from Lemma A.1 \( (b) \) from \( \{ \alpha, \beta \} \), and the monotonicity of the Marcum function, \( \{ \alpha, \beta \} \) and the fact that under \( \mathcal{H}_0 \), \( 0 \leq |(\Xi)_N - \xi_0| \leq \tau \) (as). The upper bounds on \( P_{\text{FA}} (D_N) \) follow by substituting \( \{ \alpha, \beta \} \) into \( \{ \alpha, \beta \} \) and using that \( a_1 \land a_2 \leq a_1 \). The upper bounds for \( P_{\text{MD}} (D_N) \) result from a similar procedure and the definition of \( W^* \) via \( \{ \alpha, \beta \} \).

This theorem justifies the definition of the thresholds in \( \{ \alpha, \beta \} \). It is clear that \( P_{\text{FA}} (D_N) \) and \( P_{\text{MD}} (D_N) \) of T-SeqRDT can be controlled such that T-SeqRDT is in \( \mathcal{F}(\alpha, \beta) \) by choosing appropriate parameters \( w_H, w_L, w_{BH} \) and \( w_{BL} \), which are independent of the signal model. Moreover, to do so, all these parameters have to be greater than or equal to one. Hereafter, we will work with the looser upper bounds stated in Theorem 5.2. They are simpler to analyze as they depend on fewer parameters than \( UB_{\text{FA}} \) and \( UB_{\text{MD}} \) and give useful insights into the behavior of T-SeqRDT.

We use threshold \( \lambda_{BH} (N_0 + W^*, w_{BH}) \) with \( W^* = W^*(w_{BH}, w_{BL}) \) to stop T-SeqRDT if a decision has not been taken until \( N_0 + W^* \). As pinpointed in Remark 3, the PFA
According to Lemma 1, we can write:

$$\Pr[\mathcal{D}_{N_0}(N) = \infty] = \mathbb{E} \left[ Q_{\frac{1}{2}} \left( \sqrt{N} (|\Xi| - \xi_0) - \sqrt{N} \lambda L(N, w_H) \right) \right]$$

$$- \mathbb{E} \left[ Q_{\frac{1}{2}} \left( \sqrt{N} (|\Xi| - \xi_0) - \sqrt{N} \lambda H(N, w_H) \right) \right].$$

(26)

Under $\mathcal{H}_0$, $0 \leq |(|\Xi| - \xi_0| \leq \tau$ (a-s) for all $N \geq N_0$ and thus:

$$\Pr[\mathcal{D}_{N_0}(N) = \infty] \leq Q_{\frac{1}{2}} \left( \sqrt{N} \lambda_1 - \beta/w_H \sqrt{\tau \sqrt{N}} \right)$$

$$- Q_{\frac{1}{2}} \left( 0, \lambda_\alpha/\sqrt{w_H} (\sqrt{\tau \sqrt{N}}) \right).$$

where (a) results from the monotonicity of $Q_{\frac{1}{2}}$ and (b) from (4). The bound on $\mathbb{E}[T]$ under $\mathcal{H}_0$ follows by substituting the inequality above into (25). Following a similar procedure to bound (26) under $\mathcal{H}_1$ will yield the bound under $\mathcal{H}_1$.

**Proof of (ii):** The result follows from the bound $\Pr[\mathcal{D}_{N_0}(N) = \infty] < 1$ for all $N \in \{N_0, \cdots, N_0 + W - 1\}$.

Theorem 5.3 states that the average stopping time of T-SeqRDT is strictly less than the BlockRDT block size $N_0 + W^*$ Therefore, on the one hand, Theorem 5.2 suggests that T-SeqRDT will lose detection performance compared to BlockRDT; but on the other hand, Theorem 5.2 shows that T-SeqRDT is faster on average than BlockRDT. Moreover, the bounds derived in the two theorems depend on the choice of parameters $w_H, w_L, w_H, w_B$, and $w_H$. As stated earlier, these parameters are used to select the three thresholds and the truncation window required for T-SeqRDT. Next, we study the behavior of the error probabilities and the stopping time with these parameters.

**C. Trade-off: Error probabilities vs stopping time**

In this subsection, we study how increasing/decreasing $\Pr_{\text{FA}}(\mathcal{D}_{N_0})$ and $\Pr_{\text{MD}}(\mathcal{D}_{N_0})$ affect the average stopping time of T-SeqRDT. Since $\Pr_{\text{FA}}(\mathcal{D}_{N_0}), \Pr_{\text{MD}}(\mathcal{D}_{N_0})$ and $\mathbb{E}[T]$ are not available in a closed form, we hereafter study the behavior of the upper bounds for $\Pr_{\text{FA}}(\mathcal{D}_{N_0})$ and $\Pr_{\text{MD}}(\mathcal{D}_{N_0})$ with respect to the upper bounds given for $\mathbb{E}[T]$.

**Proposition 5.4 (Behavior with $w_H$ and $w_L$):** Given $w_B$, $w_L$, and $w_H$, we have:

(i) As $w_H$ and $w_L$ tend to $\infty$, T-SeqRDT approaches BlockRDT in the sense that $\lim_{w_H, w_L \to \infty} \mathbb{E}[T] = N_0 + W^*$;

(ii) As $w_H$ and $w_L$ increase, the upper bounds on $\Pr_{\text{FA}}(\mathcal{D}_{N_0})$ and $\Pr_{\text{MD}}(\mathcal{D}_{N_0})$ decrease while the upper bounds on $\mathbb{E}[T]$ increase under each hypothesis.

**Proof:**

**Proof of (i):** Using the definition of the expectation, we have
The behavior of $E_{\text{PFA}}$ and $E_{\text{PMD}}$ decreases with increasing $\lambda$, which is equivalent to decreasing $w$. This implies that the upper bound $E_{\text{T-SQRTD}}$ decreases with increasing $w$. Therefore, the results follow from Proposition 5.4.

Next, we analyze the effect of $\lambda$ and $w$ on the stopping time $T_{\text{SQRTD}}$. Proposition 5.5 states that as $w$ and $\lambda$ increase, the stopping time $T_{\text{SQRTD}}$ decreases, which contradicts the intuition that $T_{\text{SQRTD}}$ increases with $w$ and $\lambda$. However, the stopping time $T_{\text{SQRTD}}$ is in fact an inverse function of $w$, whereas the lower bound $E_{\text{T-SQRTD}}$ is independent of $w$.

We have $\lambda > w$ for fixed $w$. This suggests that $\lambda$ cannot stay fixed and must satisfy $\lambda > w$. The result follows from Proposition 5.7, since $\lambda > w$ implies $\lambda > w$. Then, we have $\lambda > w$ for fixed $w$. Therefore, the results follow from Proposition 5.7.
w_{BL}) \), hence \( W^*(w_{BH}, w_{BL}) \) so as to minimize \( w_H \) (and \( w_L \)) given by (28) as:

\[

w_{BH} = \arg \min_{w_{BH}} \frac{w_{BH}W^*(w_{BH}, w_{BL})}{w_{BH} - 1} \quad \text{s.t. } w_{BH} > 1. \quad (31)
\]

Thereby, the upper bounds derived in Theorem 1 are maintained equal to \( \alpha \) and \( \beta \) and we expect to minimize the stopping time of \( \text{Seq} \). In the next section, we experimentally show the effect of \( w_{BH} \) and \( w_{BL} \) on \( w_L, w_H \) and \( E[T] \), and pinpoint that the parameters can be chosen in a wide range without significantly impacting \( E[T] \). Since we have a method to choose appropriate values for \( w_H, w_L, w_{BH} \) and \( w_{BL} \) from which derives \( W^* = W^*(w_{BH}, w_{BL}) \), we can calculate the thresholds according to (14) and, then, perform \( \text{Seq} \) algorithm. Algorithm 1 lists the steps of \( \text{Seq} \). Next, we perform some simulations to get insights into the behavior of the algorithm.

Algorithm 1: \( \text{Seq} \)

Initialize Given \( \lambda_0, \tau, \tau^+, \alpha \), and \( \beta \).

1) Choose \( w_{BL} \) and \( w_{BH} \); thereby \( W^* = W^*(w_{BH}, w_{BL}) \),

\[

w_L = w_{BH}W^*/(w_{BH} - 1) \quad \text{and} \quad w_H = w_{BH}W^*/(w_{BH} - 1)
\]

using (28) or (14).

2) Compute \( \lambda_L(N, w_H); \lambda_L(N, w_L) \) and \( \lambda_{\text{RDT}}(N, w_{BH}) \) from (14).

While \( \lambda_L(N, w_L) < |\langle Y \rangle_N - \xi_0| \leq \lambda_H(N, w_L) \) and \( N < N_0 < N < N_0 + W^* \)

\[

N = N + 1
\]

End If \( |\langle Y \rangle_N - \xi_0| \leq \lambda_L(N, w_L) \) & \( N < N_0 + W^* \)

Accept \( \mathcal{H}_0 \)

else if \( |\langle Y \rangle_N - \xi_0| > \lambda_H(N, w_H) \) & \( N < N_0 + W^* \)

Reject \( \mathcal{H}_0 \)

else if \( |\langle Y \rangle_N - \xi_0| > \lambda_{\text{RDT}}(N, w_{BH}) \) & \( N = N_0 + W^* \)

Accept \( \mathcal{H}_0 \)

else if \( |\langle Y \rangle_N - \xi_0| > \lambda_{\text{RDT}}(N, w_{BH}) \) & \( N = N_0 + W^* \)

Reject \( \mathcal{H}_0 \)

End If.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we perform some simulations to highlight the advantages of \( \text{Seq} \) compared to \( \text{Block} \) and \( \text{SPRT} \) as proposed in (28) and (11), (12), respectively. Moreover, we compare \( \text{Seq} \) to a composite hypothesis test, \( \text{WSPRT} \), as defined in (12) and \( \text{Seq} \) to highlight the differences between the two. We first present the detection problem considered in these experiments. Then we carry out the comparison of the algorithms.

A. Detection with signal distortions

We address the problem of testing the mean of a signal. Let us consider the case when \( Y_n = \Xi_n + X_n \), for \( n \in \mathbb{N} \) with \( \Xi_n = \xi_0 \) under \( \mathcal{H}_0 \) and \( \Xi_n \neq \xi_0 \) under \( \mathcal{H}_1 \). Here, \( \xi_0 \) is a deterministic constant and the noise is Gaussian, i.e., \( X_n \sim \mathcal{N}(0,1) \). This model can be formulated in the framework defined in (12) with \( \tau = 0 \) and \( N_0 = 1 \). This is the classical Gaussian mean shift detection problem.

However, in many practical systems, there might unfortunately be a mismatch between the model and the actual signal. In many practical applications, the underlying signal \( \Xi_n \) will not be a constant \( \xi_0 \) under \( \mathcal{H}_0 \), but a perturbed version of this value. These unavoidable perturbations are difficult to model in a parametric setup. Therefore, likelihood ratio based tests fail to guarantee reliable performance (5), (16), (27). However, the \( \text{Block} \) setup (28) and the \( \text{Seq} \) setup (29), (30) are not limited by these drawbacks. Therefore, instead of dealing with a perfect model as described above, we consider the case when \( \Xi_n = \xi_0 + \Delta_n \), where \( \Delta_n \) for \( n \in \{0,1 \} \) and all \( n \in \mathbb{N} \). Here, the \( \Delta_n \) model possible perturbations with unknown distribution. We thus want to experimentally assess different algorithms for testing \( \Xi = (\Xi_n)_{n \in \mathbb{N}} \) when we observe \( Y = (Y_n)_{n \in \mathbb{N}} \). We focus on algorithms in class \( \mathcal{C}_\tau(\alpha, \beta) \). If the distributions in play are perfectly known, \( \text{SPRT} \) is optimal in the sense that it makes a faster decision on average, compared to all other algorithms in class \( \mathcal{C}_\tau(\alpha, \beta) \). Otherwise, if the distributions are not completely known and only partial knowledge of the distortions is assumed, the above hypothesis testing problem can easily be formulated in the framework of (5). Then the problem can efficiently be solved by \( \text{Block} \), \( \text{Seq} \), or \( \text{Seq} \). In this respect, we hereafter benchmark \( \text{Seq} \) against \( \text{WSPRT} \), \( \text{SPRT} \), \( \text{Block} \), and \( \text{Seq} \) under experimental settings described below.

B. Experimental setup

We first list the parameters required to design each algorithm. \( \text{Block} \) only requires \( \tau \), but guarantees \( \text{FA} \). With no control over \( \text{MD} \). With additional knowledge of \( \tau^+ \), \( \text{Block} \) can control both \( \text{FA} \) and \( \text{MD} \) as illustrated in Proposition 3.2. Likewise, \( \text{Seq} \) also requires \( \tau \) and \( \tau^+ \), whereas \( \text{Seq} \) requires \( \tau^-, \tau^+, \tau_H \). On the other hand, \( \text{SPRT} \) requires complete knowledge of the signal distributions under each hypothesis. Similarly, \( \text{WSPRT} \) also requires complete knowledge of the signal distributions at least up to an unknown (possibly vector) parameter. Note that \( \text{Block} \) is an FSS algorithm whereas the rest of the algorithms are sequential and belong to class \( \mathcal{C}_\tau(\alpha, \beta) \). For the experimental setup, let us assume \( \tau^- \) to be some positive real value. We consider \( \xi_1 \) and \( \xi_0 \) such that \( |\xi_1 - \xi_0| > 4\tau^- \). We set \( \tau^+ = |\xi_1 - \xi_0| - \tau^- \) and \( \tau_H \in [\xi_1 - \xi_0] + \tau^- \). Suppose that the empirical mean of the distortion \( \Delta = (\Delta_n)_{n \in \mathbb{N}} \) exhibits the following bounded behavior: there exists some \( N_0 \in \mathbb{N} \) such that, for all \( N \geq N_0, 0 \leq |\langle \Delta \rangle_N| \leq \tau^- \) and \( \tau^+ \leq |\langle \Delta \rangle_N + \xi_1 - \xi_0| \leq \tau_H \). The first inequality captures the signal behavior under \( \mathcal{H}_0 \), whereas the second inequality captures the signal behavior under \( \mathcal{H}_1 \). The problem of testing the mean of \( \Xi \) can be rewritten as:

\[

\begin{align*}
\mathcal{H}_0 & : \forall N \geq N_0, 0 \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau^- \quad \text{(a-s)}, \\
\mathcal{H}_1 & : \forall N \geq N_0, \tau^- < \tau^+ \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau_H \quad \text{(a-s)}. \\
\end{align*}
\]

We can choose \( \tau \in (\tau^-, \tau^+) \). For simulation purposes, we set \( \tau = 2\tau^- \). Note that (32) is a special case of the testing problem (5) and can thus be tested using the \( \text{Block} \), \( \text{Seq} \), and \( \text{Seq} \) frameworks. None of these algorithms need the complete knowledge of the distortion (or signal) distributions under either hypothesis, unlike \( \text{SPRT} \) and \( \text{WSPRT} \), which require the complete knowledge of these distributions under both
hypotheses. We consider three different types of distortions, two when (32) is only required to be satisfied with high probability and the third when it is satisfied with probability 1 (in (a-s) sense).

Case 1: Gaussian distortion: We assume \( \Delta_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \) for \( n \in \mathbb{N} \). For simulation purposes, we choose \( \tau^* = \sigma/4 \). For this distortion type, the inequalities in (32) will only be satisfied with high-probability. Below, we list the probabilities corresponding to the Gaussian distortion. We have \( \Pr[|\Delta_N| \leq \tau] \geq 0.9545, \Pr[|\Delta_N + \xi_1 - \xi_0| > \tau] \geq 0.9772, \Pr[|\Delta_N| \leq \tau^*] \geq 0.6827 \) and \( \Pr[|\Delta_N + \xi_1 - \xi_0| > \tau^*] \geq 0.8413 \) for all \( N \geq N_0 \) with \( N_0 = 16 \) and \( |\xi_1 - \xi_0| \geq 2\tau \). Note that these probabilities increase with \( N \).

Case 2: Heavy-Tailed distortion: We model \( \Delta_n \) as an \( \alpha \)-stable random variable denoted as \( \Delta_n \overset{\text{i.i.d.}}{\sim} S(\alpha, \beta, \gamma, \delta) \) for \( n \in \mathbb{N} \). The parameters \( \alpha \in (0, 2], \beta \in [-1, 1], \gamma > 0 \) and \( \delta \in (-\infty, \infty) \) are the tail-index, location, dispersion and skewness parameters, respectively. In general, an \( \alpha \)-stable distribution does not admit a closed-form probability density function, except in a few special cases like the Cauchy (\( \alpha = 1, \beta = 0 \)) and Gaussian (\( \alpha = 2 \)) distributions. Moreover, for the Cauchy distribution and for \( \alpha \in (0, 1] \), none of the moments of the \( \alpha \)-stable distribution exist. For \( \alpha \in (1, 2) \) the distribution is sometimes referred to as the Pareto-Lévy distribution and for this class of distributions, all higher moments beyond the mean do not exist. For simulation purposes, we consider the following two types of heavy-tailed distortions:

Case 2(i) [Pareto-Lévy distortion (\( \alpha \in (1, 2) \))]: We assume the distortion to be Pareto-Lévy distributed with \( \Delta_n \overset{\text{i.i.d.}}{\sim} S(1.5, 0, \tau^*, 0) \) for \( n \in \mathbb{N} \). We thus have: \( \Pr[|\Delta_N| \leq \tau] \geq 0.9885, \Pr[|\Delta_N + \xi_1 - \xi_0| > \tau] \geq 0.9953, \Pr[|\Delta_N| \leq \tau^*] \geq 0.9646, \Pr[|\Delta_N + \xi_1 - \xi_0| > \tau^*] \geq 0.9832 \) for all \( N \geq N_0 \) with \( N_0 = 30 \) and \( |\xi_1 - \xi_0| \geq 2\tau \). Again, note that these probabilities increase with \( N \).

Case 2(ii) [Cauchy distortion (\( \alpha = 1 \))]: Note that, unlike in the cases involving Gaussian and Pareto-Lévy distortions, the empirical mean of i.i.d Cauchy distributed random variables is again Cauchy distributed [35] and none of the moments exist for the empirical mean as well. Therefore, the empirical mean of a Cauchy distorted signal does not converge in the neighborhood of \( \xi_0 \) and \( \xi_1 \) under \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \), respectively, in contrast to the Gaussian and Pareto-Lévy distortions. Below, we show that, although the Cauchy distortion does not exhibit the desired convergence properties, the proposed algorithms guarantee performance if (32) holds with sufficiently high probabilities. To experimentally show this, we assume the distortion to be Cauchy distributed as \( \Delta_n \overset{\text{i.i.d.}}{\sim} S(1, 0, \tau^*/10, 0) \) for \( n \in \mathbb{N} \) with the associated probabilities given as: \( \Pr[|\Delta_N| \leq \tau] = 0.9682 \) and \( \Pr[|\Delta_N| \leq \tau^*] = 0.9365 \) for all \( N \in \mathbb{N} \). Also, \( \Pr[|\Delta_N + \xi_1 - \xi_0| > \tau] \geq 0.9894 \) and \( \Pr[|\Delta_N + \xi_1 - \xi_0| > \tau^*] \geq 0.9728 \) for all \( N \in \mathbb{N} \) and \( |\xi_1 - \xi_0| \geq 2\tau \). Note that, unlike the Cases 1 and 2(i) these probabilities do not increase with \( N \) as the distribution of the empirical mean of a Cauchy distribution remains the same. As a consequence, the probabilities stay the same for all \( N \in \mathbb{N} \).

To ensure that (32) is satisfied with high probability, we need the dispersion parameter, \( \gamma \), to be small enough. Later in the section we show how the above probabilities, PFA and PMD vary with \( \gamma \) for T-SeqRDT.

Case 3 [Deterministic unknown distortion]: The distortion is assumed to be unknown deterministic with \( |\Delta_n| \leq \tau^* \) for all \( n \in \mathbb{N} \). For simulation purposes, we choose \( \Delta_n = \tau^* \). With this choice, the inequalities in (32) are satisfied with probability 1. However, not all types of distortions satisfy (32) with probability 1 as shown in Cases 1 and 2. Next, we discuss different algorithms.

---

**TABLE I: T-SeqRDT vs SeqRDT and BlockRDT for Gaussian distortion.**

| SNR = | \( |\xi_1 - \xi_0| \) | 0.4 | 0.5 | 0.6 | 0.8 |
|-------|-----------------|-----|-----|-----|-----|
| T-SeqRDT | \( E[T_{SeqRDT}] \) | 567.3 | 349.2 | 192.82 | 73.92 |
| SeqRDT | \( E[T_{SeqRDT}] \) | 3820 | 955 | 425 | 153 |
| BlockRDT | \( E[T_{BlockRDT}] \) | 252.55 | 198.51 | 185.17 | 181.8 |
| \( \alpha = \beta = 0.001 \) | \( E[T_{BlockRDT}] \) | 720.08 | 481.75 | 298.65 | 114.88 |

**TABLE II: T-SeqRDT vs SeqRDT, SPRT and WSPRT.** Here, \( PFA < 10^{-5} \) and \( PMD < 10^{-5} \) indicate that probabilities of errors are at most of the order of \( 10^{-5} \).

| SNR = | \( |\xi_1 - \xi_0| \) | 0.4 | 0.5 | 0.6 | 0.8 |
|-------|-----------------|-----|-----|-----|-----|
| T-SeqRDT | \( E[T] \) | 720.08 | 481.75 | 298.65 | 114.88 |
| WSPRT | \( E[T_{WSPRT}] \) | 304.37 | 42.39 | 28.76 | 16.05 |
| \( \alpha = \beta = 0.001 \) | \( E[T_{WSPRT}] \) | 304.37 | 42.39 | 28.76 | 16.05 |

\[ \text{TABLE II: T-SeqRDT vs SeqRDT, SPRT and WSPRT. Here, PFA < 10^{-5} and PMD < 10^{-5} indicate that probabilities of errors are at most of the order of 10^{-5}.} \]
C. Algorithms

Here, we discuss the algorithms we use to solve the above mean testing problem. Certainly, using [32], we can cast the problem in the BlockRDT, SeqRDT and T-SeqRDT frameworks. Next, we discuss likelihood ratio based parametric and semiparametric approaches for comparison purposes.

Sequential probability ratio test (SPRT): For SPRT, we assume that the probability density function, $f_i$, of the observations is known under $\mathcal{H}_i$ for $i = 0, 1$. For $\alpha, \beta \in (0, 1/2)$, and with initialization $\Lambda_0 = 1$, SPRT with stopping time and decision pair $(T_{\text{SPRT}}, D)$ is defined as:

$$T_{\text{SPRT}} = \inf \{ N \geq 0 : \lambda_N \notin (\lambda_L^{\text{SPRT}}, \lambda_U^{\text{SPRT}}) \}$$

$$D(N) = \begin{cases} 
1 & \text{if } \lambda_N \geq \lambda_U^{\text{SPRT}} \\
0 & \text{if } \lambda_N \leq \lambda_L^{\text{SPRT}} \\
\infty & \text{if } \lambda_L^{\text{SPRT}} < \lambda_N < \lambda_U^{\text{SPRT}} 
\end{cases}$$

where $\Lambda_N = \sum_{n=1}^N \frac{\exp(Y_n)}{f_0(Y_n)}$ is the likelihood ratio based on the observations, $\lambda_L^{\text{SPRT}} = \frac{\beta}{\lambda_0}$ and $\lambda_U^{\text{SPRT}} = \frac{1-\beta}{\lambda_0}$ are the lower and upper thresholds, respectively. We denote the stopping time, PFA and PMD of SPRT as $T_{\text{SPRT}}, \text{PFA}^{\text{SPRT}}$ and $\text{PMD}^{\text{SPRT}}$, respectively. For the model described above, SPRT for detecting the mean with unknown distortions we have $\Lambda_N = \exp \left( \sum_{n=1}^N (\xi_n - \xi_0) \right)$.

Composite hypothesis test, GSPRT: A simple GSPRT can be designed for the case of Gaussian distortions when the means under $\mathcal{H}_0$ and $\mathcal{H}_1$ are known but the variances are unknown. Specifically, the algorithm is aware that the distortion is zero mean Gaussian distributed, but is unaware of its variance $\sigma^2$. The generalized log likelihood ratio for such a test is given as: $\log \Lambda_N = \frac{\sum_{n=1}^N (Y_n - \frac{1}{2} (\xi_n + \xi_0))}{\frac{1}{n} \sum_{n=1}^N (Y_n - \frac{1}{2} (\xi_n + \xi_0))}$ with $s_n^2 = \frac{1}{n} \sum_{n=1}^N (Y_n - \frac{1}{2} (\xi_n + \xi_0))$ for $N \geq 2$. GSPRT uses the same thresholds as SPRT [5]. We denote the stopping time, PFA and PMD of GSPRT as $E \left[ T_{\text{GSPRT}} \right], \text{PFA}^{\text{GSPRT}}$ and $\text{PMD}^{\text{GSPRT}}$, respectively.

Composite hypothesis test, WSPRT: WSPRT considers the problem of testing $\mathcal{H}_0 : |\xi - \xi_0| < \tau$ vs $\mathcal{H}_1 : |\xi - \xi_0| > \tau$ [12] Chapter 4). In T-SeqRDT, the signal $\Xi$ is assumed to be a corrupted version of $\xi$ and the distribution of $\Xi$ is assumed to be unknown. In contrast, for WSPRT, $\xi$ is deterministic and the test can only handle the case when the observations, $Y_i$, are Gaussian distributed [12], [16]. We denote PFA, PMD and stopping time of WSPRT as $\text{PFA}^{\text{WSPRT}}$, $\text{PMD}^{\text{WSPRT}}$ and $T_{\text{WSPRT}}$, respectively. WSPRT uses the same thresholds as SPRT. However, the likelihood ratio for WSPRT is given as: $\hat{\lambda}_N = \prod_{n=1}^N f_0(Y_n e^{-\frac{1}{2}(\xi_n - \xi)^2}) f_0(Y_n e^{-\frac{1}{2}(\xi_0 - \xi)^2})$. The distribution under $\mathcal{H}_1$ is thus replaced by a weighted average of two distributions.

D. Comparison: T-SeqRDT, SeqRDT, BlockRDT, SPRT, WSPRT and GSPRT

We define $|\xi_1 - \xi_0|$ as the SNR and for simulation purposes, we assume $\tau = 0.1$. For T-SeqRDT, the thresholds $\xi_0$ and the truncation window $W^{\ast}(w_{BH}, w_{BL})$ [15] are selected via parameters $w_H, w_L, w_{BH}$ and $w_{BL}$, using Algorithm 1. For the simulations, we assume $w_{BL} = w_{BH}$, which implies...
that $w_H = w_L$ and make use of the simpler of the three methods proposed in (31) to choose the parameters. In Figures 1 and 2, we plot $w_H$ against $w_{BH}$ and $E[T]$, respectively, for different $\alpha$, $\beta$, and $\tau^-$.

Moreover, we see that we can choose $w_{BH}$ and $w_{BL}$ (hence $w_H$ and $w_L$) over a wide range. For simulation purposes, we select $w_{BH} = w_{BL} = 2$, which is then used to choose an appropriate $W^*$ via (15). We then set $w_H = w_L = 2W^*$. Albeit not unique, these choices guarantee that T-SeqRDT is in $\mathcal{C}(\alpha, \beta)$ by Theorem 5.2. In contrast to T-SeqRDT and BlockRDT, SeqRDT requires additional knowledge of $\tau^-$ and $\tau_H$ to design the test: $\tau$ along with levels $\alpha$ and $\beta$ is used to design the thresholds; $\tau^-$, $\tau^+$ and $\tau_H$ are used to choose an appropriate buffer size, $M$. The thresholds, $\lambda_\alpha(\sqrt{\alpha})/\sqrt{\alpha}$ and $\lambda_1(\sqrt{\beta})/\sqrt{\beta}$ are defined in (13). We denote by $\mathbb{P}_{FA}(D_M)$ and $\mathbb{P}_{MD}(D_M)$ the PFA and PMD, respectively, and by $T_{SeqRDT}$ the stopping time of SeqRDT.

**Case 1:** For Gaussian distortion, we compare in Table III the average stopping times of T-SeqRDT and SeqRDT to the block-size of BlockRDT, for different SNR values, $\alpha = 0.01$ and $\beta = 0.001$. For SeqRDT, the buffer size $M = 90$ is selected. From Table III, SeqRDT is the fastest on average, especially at low SNR values, but needs the most amount of information (all of $\tau^-$, $\tau^+$ and $\tau_H$ to design $M$) about the signal. BlockRDT is the slowest and requires the same information ($\tau$ and $\tau^+$ only) as T-SeqRDT. However, T-SeqRDT is considerably faster on average. Moreover, at moderate to high SNRs, T-SeqRDT is the fastest among the three algorithms and considerably outperforms SeqRDT as well. It must be noted that the stopping time of SeqRDT is limited by the need and the choice of the buffer size, which makes SeqRDT relatively slower compared to T-SeqRDT, especially at higher SNRs. In Table III, we compare the average stopping times, PFAs and PMDs of T-SeqRDT, SeqRDT, SPRT, WSPRT and GSPRT. From Table III, we notice that, because of the distortion, WSPRT and SPRT do not belong to $\mathcal{C}(\alpha, \beta)$ as $\mathbb{P}_{FA}$, $\mathbb{P}_{MD}$ and $\mathbb{P}_{FA}$ are above the pre-specified levels $\alpha$ and $\beta$. Moreover, GSPRT, even with prior knowledge of the distortion, does not belong to $\mathcal{C}(\alpha, \beta)$, as the $\mathbb{P}_{FA}$ and $\mathbb{P}_{MD}$ are orders of magnitude higher compared to $\alpha$ and $\beta$, respectively. In contrast, both SeqRDT and T-SeqRDT, with only limited knowledge about the signal under each hypothesis, are in $\mathcal{C}(\alpha, \beta)$. Importantly, by design, T-SeqRDT eliminates the need for buffer $M$, whereas SeqRDT does need such a buffer to guarantee the pre-specified levels $\alpha$ and $\beta$. Moreover, it seems that the bounds on PFA and PMD are loose for Gaussian distortion. Therefore, we next consider different types of distortions to see if the bounds are tight for some other scenarios.

**Case 2:** For heavy-tailed distortions, we again compare T-SeqRDT, SeqRDT, SPRT and WSPRT. For Cases 2(i) and 2(ii), we simulate PFA, PMD and average stopping times for $\alpha = 0.01$ and $\alpha = 0.05$ to obtain Tables III and IV respectively. The average stopping time of T-SeqRDT and SeqRDT stay similar to those obtained in the Gaussian distortion case. However, the bounds on PFA and PMD are tight, as a consequence of the heavy-tailed distribution of the distortion. Moreover, similar to Case 1, SPRT and WSPRT do not belong to $\mathcal{C}(\alpha, \beta)$ for both Cases 2(i) and 2(ii).

**Case 3:** Finally, we consider the unknown deterministic distortion case. In this case, (22) is satisfied with probability 1, unlike in Cases 1 and 2. We choose $\Delta_n = \tau^-$, and simulate PFA, PMD and average stopping times for $\alpha = 0.05$ and $\alpha = 0.01$. From Table V, T-SeqRDT and SeqRDT belong to $\mathcal{C}(\alpha, \beta)$, whereas SPRT and WSPRT fail too. Also, note that, similar to Case 2, the bounds on PFA are tight.

A note of caution: The above simulation results show that T-SeqRDT is robust to mismatches and can guarantee performance even in the cases when (22) is not always satisfied with probability 1, as shown in Cases 1 and 2 above. Now, the question that arises is: “How high do these probabilities need to be so that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$?” The simulation results of Case 1 suggest that Gaussian distortions allow for large mismatches, i.e., T-SeqRDT works even when the probabilities are not very high. On the other hand, Case 2, involving heavy-tailed distortions, requires these probabilities to be high, i.e., a relatively smaller mismatch. In the following, for two different cases, we show how high these probabilities need to be for T-SeqRDT to belong to $\mathcal{C}(\alpha, \beta)$.

**Case A:** We mentioned earlier in Case 2(ii) that, with Cauchy distortions, we needed the dispersion parameter $\gamma$ to be small. In Fig. 5, we show how PFA and PMD of T-SeqRDT and the probabilities associated in (22), when they are not satisfied in (a-s) sense, vary with increasing $\gamma$, for $\alpha = \beta = 0.05$ and SNR $= 0.8$. Notice that there exists a
95% for T-Seq as shown in all the above cases. The above discussion shows that T-Seq is completely data dependent rather than model dependent. We assume that the signal distributions in play are unknown as it is extremely valuable to improve the clarity of the paper. In conclusion, we believe that the proposed framework is capable of carrying out hypothesis testing. Another critical feature of the proposed approach is that it gives the algorithm designer freedom to choose these parameters, thus making it possible to test signals with arbitrarily low SNRs. Also, the algorithm is robust to mismatches in signal distributions as it does not rely on the underlying signal distributions. In the future, the optimality properties of the proposed tests remain to be studied. Also, extension of T-SeqRDT to multi-dimensional signals may be addressed. Moreover, the simulations indicate that the bounds derived on PFA and PMD might be loose in some cases, which highlights the need for novel assumptions and analysis in the future. Generalization of the model to distributed systems might be another future endeavor. In conclusion, we believe that the proposed hypothesis testing approach provides an appealing alternative to likelihood ratio based sequential frameworks.

VII. CONCLUSION AND PERSPECTIVES

In this work, we proposed an alternative approach to hypothesis testing. The proposed formulation can be useful when the signal distributions in play are unknown as it is completely data dependent rather than model dependent. We introduced a new non-parametric algorithm, T-SeqRDT, for sequential hypothesis testing. The work builds on earlier work where the authors proposed SeqRDT [29], [30]. SeqRDT, although marginally faster compared to T-SeqRDT, required more assumptions on the signal along with a buffer to ensure the desired performance. In contrast, T-SeqRDT guarantees the required performance and eliminates the need for a buffer by designing thresholds and truncation window appropriately. We studied the properties of these thresholds along with the trade-off between the error probabilities and the stopping time. Finally, simulations showed that T-SeqRDT, even with little knowledge of the signal, is able to provide sufficient performance guarantees while making a decision faster on average compared to BlockRDT. Furthermore, with the knowledge of a few parameters only, instead of the complete distributions, the proposed framework is capable of carrying out hypothesis testing. Another critical feature of the proposed approach is that it gives the algorithm designer freedom to choose these parameters, thus making it possible to test signals with arbitrarily low SNRs. Also, the algorithm is robust to mismatches in signal distributions as it does not rely on the underlying signal distributions. In the future, the optimality properties of the proposed tests remain to be studied. Also, extension of T-SeqRDT to multi-dimensional signals may be addressed. Moreover, the simulations indicate that the bounds derived on PFA and PMD might be loose in some cases, which highlights the need for novel assumptions and analysis in the future. Generalization of the model to distributed systems might be another future endeavor. In conclusion, we believe that the proposed hypothesis testing approach provides an appealing alternative to likelihood ratio based sequential frameworks.

ACKNOWLEDGMENT

We gratefully acknowledge the insightful comments of the reviewers and the Associate Editor whose suggestions were extremely valuable to improve the clarity of the paper.

APPENDIX A

Lemma A.1: For any $\Delta \in \mathbb{N}$ and any $\eta \geq 0$, we have:

$$
\mathbb{P} \left[ |(\bar{\Xi})_n + (\bar{X})_n - \xi_0| > \eta \right] = \mathbb{E} \left[ Q_2 \left( \sqrt{\Delta} \xi_0, \eta \sqrt{\Delta} \right) \right]
$$

Proof: By property of a conditional and taking the independence of $\langle \bar{\Xi} \rangle_n$ and $\langle (X)_n \rangle$ into account, we have:

$$
\mathbb{P} \left[ |(\bar{\Xi})_n + (\bar{X})_n - \xi_0| \leq \eta \right] = \int_0^\infty P \left[ ||\rho + (\bar{X})_n| - \xi_0| \leq \eta \right] P \left[ |\bar{\Xi}|_n - \xi_0|^{-1}(d\rho) \right]
$$
It follows from $X \sim N(0, 1)$ that, for all $\rho \in [0, \infty)$:

$$\mathbb{P} [\rho + \langle X \rangle_N \leq \eta] = \Phi \left( \sqrt{\lambda(\eta - \rho)} \right) - \Phi \left( -\sqrt{\lambda(\eta + \rho)} \right)$$

The foregoing and (33) imply the result through the equality:

$$\mathbb{P} \left[ \langle \Xi_N \rangle_N \leq \xi_0 \leq \eta \right] = 1 - \mathbb{E} \left[ Q_{1/2} \left( \sqrt{\lambda(\eta - \rho)} \right) \right].$$

**APPENDIX B**

**Lemma B.1:** For any $\gamma \in (0, 1)$:

(i) $\lim_{\rho \to \infty} (\lambda_\gamma(\rho) - \rho) = \Phi^{-1}(1 - \gamma)$;

(ii) $\lim_{\rho \to \infty} \lambda_\gamma(\rho) / \rho = 1$;

**PROOF:** We prove (i) only since it straightforwardly implies (ii). To prove that $\lim_{\rho \to \infty} \lambda_\gamma(\rho) / \rho = 1$, consider $\rho_n = \max(n, \eta)$ such that either $g_\gamma(\rho_n) > \theta + \epsilon \text{ or } g_\gamma(\rho_n) < \theta - \epsilon$. Thus, $\lim_{\rho \to \infty} \lambda_\gamma(\rho) / \rho = 1$.

Similarly, since $\lim_{\rho \to \infty} \Phi(2\rho_n + \theta) - \epsilon = 1$, there exists $\rho_n \in \mathbb{N}$ such that, for all $n \geq N_0$:

$$\Phi(2\rho_n + \theta - \epsilon) < 1 - \eta.$$ (34)

Let $n$ be any integer above $\max(N_0, N_1)$. If $g_\gamma(\rho_n) < \theta - \epsilon$, we then have $\Phi(g_\gamma(\rho_n)) < \Phi(\theta - \epsilon)$ and $\Phi(2\rho_n + g_\gamma(\rho_n)) < \Phi(2\rho_n + \theta - \epsilon)$. Eqs. (33) and (34) then imply that:

$$1 + \Phi(\theta) < \Phi(\theta - \epsilon) + \Phi(2\rho_n + \theta - \epsilon) < \Phi(\theta - \epsilon) + 1 + \eta,$$

which is impossible because of our choice for $\eta$. Therefore, we cannot have $g_\gamma(\rho_n) < \theta - \epsilon$. We cannot have $g_\gamma(\rho_n) > \theta + \epsilon$ either because, via (33) and (35), this inequality implies:

$$1 + \Phi(\theta) > \Phi(\theta + \epsilon) + \Phi(2\rho_n + \theta + \epsilon) > \Phi(\theta + \epsilon) + 1 + \eta.$$ (36)

which is contradictory to our choice for $\eta$.

**Lemma B.2 (Behavior of $Q_{1/2}$ in vanishing noise):** Consider $\tau \in [0, \infty)$ and $\rho \in (0, \infty)$ such that $\rho \neq \tau$.

$$\forall \gamma \in (0, 1), \lim_{\rho \to \infty} Q_{1/2}(\rho, \gamma(\tau/\sigma)) = \Phi(\tau, \infty) \rho.$$ (37)

**PROOF:** Let $(\sigma_n)_{n \in \mathbb{N}}$ be a sequence of positive real values such that $\lim_{n \to \infty} \sigma_n = 0$ and set $\rho_n = \tau / \sigma_n$ for each $n \in \mathbb{N}$. According to (2), $Q_{1/2}(\rho/\tau \rho_n, \lambda_\gamma(\rho_n)) = \mathbb{P} \left[ (\rho/\tau) + (X/\rho_n) > \lambda_\gamma(\rho_n) / \rho_n \right]$ for any $X \sim (0, 1)$. It follows from Lemma B.1 (ii) that $(\rho/\tau) + (X/\rho_n) > \lambda_\gamma(\rho_n) / \rho_n - 1$ (a.s). Therefore, the cdf of $(\rho/\tau) + (X/\rho_n) - \lambda_\gamma(\rho_n) / \rho_n$ converges weakly to $1_{(\rho/\tau, \infty)}$ (a.s). Since $\rho \neq \tau$, this weak convergence implies that $\lim_{n \to \infty} \mathbb{P} \left[ (\rho/\tau) + (X/\rho_n) > \lambda_\gamma(\rho_n) / \rho_n \right] = 1_{(\tau, \infty)}(\rho)$. Thence, the result since $(\sigma_n)_{n \in \mathbb{N}}$ is arbitrary.

**Lemma B.3 (Non-Asymptotic behavior of $Q_{1/2}$):** Consider $\tau \in [0, \infty), \rho \in (0, \infty)$ and $\gamma \in (0, 1)$, the map:

$$\sigma \mapsto Q_{1/2}(\rho \sigma, \gamma(\tau/\sigma))$$

**PROOF:** Given $\rho$ and $\tau$, we want to study the behavior of

$$\mathbb{Q}(\sigma) = Q_{1/2}(\rho \sigma, \gamma(\tau/\sigma)) = 1 - \Phi(\rho \sigma - \tau, \gamma(\tau/\sigma))$$ (37)

with $r_+ = \lambda_\gamma(\tau) + \rho \sigma$ and $r_- = \lambda_\gamma(\tau) - \rho \sigma$. For $\rho \neq \tau$, from (37) that $\mathbb{Q}$ is constant equal to $\gamma$. Thus, we have $1 - \Phi(\lambda_\gamma(\tau) - \rho \sigma) + \Phi(-\lambda_\gamma(\tau) + \rho \sigma) = \gamma$. After differentiating the two members of the equality above and after some routine algebra, we obtain:

$$\lambda_{\gamma}(\tau) = \left( 1 - e^{-2\pi \lambda_\gamma(\tau)} \right) / \left( 1 + e^{-2\pi \lambda_\gamma(\tau)} \right)$$ (38)

where $\lambda_{\gamma}$ is the first derivative of $\lambda_\gamma$. We now differentiate $\mathbb{Q}$ defined by (37). Some easy computation yields:

$$\mathbb{Q}(\sigma) = 1 - \frac{e^{-r_+^2(\sigma)/2} - e^{-r_-^2(\sigma)/2}}{2r_+ - \Delta_{\gamma}(\rho, \tau)}$$ (39)

with $\Delta_{\gamma}(\rho, \tau) = (1 + e^{-2\pi \lambda_\gamma(\tau)})/(1 + e^{-2\pi \lambda_\gamma(\tau)})$ and $\mathbb{Q}(\sigma) = \text{constant}$. For all $\sigma \geq 0$, the sign of $\mathbb{Q}(\sigma)$ is therefore that of $(\rho/\tau) - (\Delta_{\gamma}(\rho, \tau))$. We verify easily that:

$$\begin{align*}
\rho < \tau &\Rightarrow \Delta_{\gamma}(\rho, \tau) > 1 \Rightarrow \Delta_{\gamma}(\rho, \tau) > 1 \\
\rho = \tau &\Rightarrow \Delta_{\gamma}(\rho, \tau) = \Delta_{\gamma}(\rho, \tau) = 1
\end{align*}$$

Therefore, if $\rho < \tau$, $\rho/\tau < 1 < \Delta_{\gamma}(\rho, \tau)$, which implies that $\mathbb{Q}(\sigma) > 0$ and, thus, that $\mathbb{Q}$ is decreasing. On the other hand, if $\rho > \tau$, we have $\rho/\tau > 1 > \Delta_{\gamma}(\rho, \tau)$, so that $\mathbb{Q}$ is increasing in this case.

**Lemma B.4:** Given $\rho \in (0, \infty)$, the map $\gamma \in (0, 1) \mapsto \lambda_\gamma(\rho)$ is decreasing.

**PROOF:** It is a straightforward consequence of (37) and the decreasing nature of $Q_{1/2}$ with its second argument given in Lemma B.1.

**Lemma B.5:**

(P1) For any $\tau \in [0, \infty)$ and any $\gamma \in (0, \infty)$, the map $\sigma \mapsto Q_{1/2}(\rho \sigma, \tau/\sigma, \eta/\sigma)$ is increasing.

(P2) The map $\rho \mapsto Q_{1/2}(\rho \sigma, \tau/\sigma, \eta/\sigma)$ is decreasing, lower-bounded by $1/2$ and $\lim_{\rho \to \infty} Q_{1/2}(\rho \sigma, \tau/\sigma, \eta/\sigma) = 1/2$.

(P3) For any $\gamma \in (0, 1/2)$, the map $\rho \mapsto Q_{1/2}(\rho \sigma, \tau/\sigma, \eta/\sigma)$ is decreasing, lower bounded by $1/2$.

**PROOF:** Using (3), define $\mathbb{Q}(\sigma)$ as:

$$\mathbb{Q}(\sigma) = Q_{1/2}(\rho \sigma, \tau, \eta/\sigma) = 1 - \Phi(\eta/\sigma - \tau/\sigma) + \Phi(-\eta/\sigma - \tau/\sigma).$$

We now differentiate $\mathbb{Q}$ and some easy computation yields:

$$\mathbb{Q}(\sigma) = (1 - \sqrt{2\pi \sigma^2})e^{-\frac{(\eta - \tau)^2}{2\sigma^2}} \left[ (\eta - \tau) + (\eta + \tau)e^{-2\pi \gamma} \right].$$

Thence the result, since $\gamma \in (\tau, \infty)$ implies that $\mathbb{Q}(\sigma) > 0$. 
Proof of statement (P2): The map $\rho \in (0, \infty) \rightarrow Q_{\frac{1}{2}}(\rho, \rho)$ is decreasing as a consequence of (P1). Given $\rho \in (0, \infty)$, $Q_{\frac{1}{2}}(\rho, \rho) = 1/2 + \Phi(-2\rho)$ from (3). Whence the result. Proof of statement (P3): Let $\rho$ and $\rho'$ be two positive real numbers such that $0 < \rho < \rho'$. According to (40), we have:

\[ Q_{\frac{1}{2}}(\rho, \lambda_1(\rho)) = Q_{\frac{1}{2}}(\rho', \lambda_1(\rho')) = \gamma. \]

Since $\gamma < 1/2$ so that $1/2 < 1 - \gamma$ follows it from (P2) and (40) that $Q_{\frac{1}{2}}(\rho, \rho) > 1/2 > Q_{\frac{1}{2}}(\rho, \lambda_1(\rho))$. The decreasing behavior of $Q_{\frac{1}{2}}$ with its second argument implies that $\lambda_1(\rho) > \rho$, so that $\lambda_1(\rho)/\rho$ is lower bounded by 1. We then derive from (P1) that $x \in (0, \infty) \rightarrow Q_{\frac{1}{2}}(\rho/x, \lambda_1(\rho)/x)$ is an increasing map. Since $\rho/\rho' < 1$, we thus have $Q_{\frac{1}{2}}(\rho, \lambda_1(\rho)/\rho) > Q_{\frac{1}{2}}(\rho', \lambda_1(\rho)/\rho)$. This inequality and (40) induce that $Q_{\frac{1}{2}}(\rho', \lambda_1(\rho')) > Q_{\frac{1}{2}}(\rho', \lambda_1(\rho)/\rho)$. The decreasing nature of $Q_{\frac{1}{2}}(\rho', \cdot)$ then implies that $\lambda_1(\rho') < \rho_1(\rho)/\rho$. Thereby, $\rho \in (0, \infty) \rightarrow \lambda_1(\rho)/\rho$ is decreasing and lower bounded by 1, this map has a limit $\ell \geq 1$ when $\rho$ tends to $\infty$. The result then follows as a consequence of Lemma B.1(ii).

Lemma B.6: For $\gamma \in (1/2, 1)$ and $\rho$ large enough, the map $\rho \in (0, \infty) \rightarrow \lambda_1(\rho)/\rho$ is increasing, upper bounded by 1.

Proof: According to statement (i) of Lemma B.1 $\lambda_1(\rho) - \rho = \Phi^{-1}(1 - \gamma) + \kappa(\rho) \kappa(\rho)$ where $\kappa$ is such that $\lim_{\rho \to \infty} \kappa(\rho) = 0$.

Since $\gamma > 1/2$, $\Phi^{-1}(1 - \gamma) < 0$. Given $\eta$ such that $0 < \eta < \Phi^{-1}(1 - \gamma)$, there exists $\rho_0$ such that, for all $\rho \geq \rho_0$, $\kappa(\rho) \leq \eta$. Therefore, for all $\rho \geq \rho_0$, $\lambda_1(\rho) - \rho \leq \Phi^{-1}(1 - \gamma) + \eta < 0$.

We have hence proved that $\lambda_1(\rho) - \rho < \rho$ for large enough $\rho$.

With $h_\rho(\rho) = \lambda_1(\rho)/\rho$, $\Phi(\rho(h_\rho(\rho) - 1)) - \Phi(\rho(h_\rho(\rho) + 1)) = 1 - \gamma$. By differentiation of this equality with respect to $\rho$ and since $h_\rho$ is differentiable via the implicit function theorem, we find $h_\rho'(\rho)$ has the same sign as $T(\rho) = (1 - h_\rho(\rho))(\frac{2\lambda_1(\rho)}{\lambda_1(\rho) + \Phi'(\rho)}).$ For $\rho$ large enough, $h_\rho(\rho) < 1$ by the first part of the proof and Lemma B.1 implies that $\lim_{\rho \to \infty} T(\rho) = \infty$. Therefore, $T(\rho) > 0$ for $\rho$ large enough and the proof is complete.

References: