3-13-2020

Low Latency Bearing Fault Detection of Direct-Drive Wind Turbines Using Stator Current

Samrat Nath

Jingxian Wu

Yue Zhao

Wei Qiao

Received February 2, 2020, accepted February 23, 2020, date of publication March 2, 2020, date of current version March 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977632

Low Latency Bearing Fault Detection of Direct-Drive Wind Turbines Using Stator Current

SAMRAT NATH[1], (Student Member, IEEE), JINGXIAN WU[1], (Senior Member, IEEE), YUE ZHAO[1], (Senior Member, IEEE), AND WEI QIAO[2], (Fellow, IEEE)

[1]Department of Electrical Engineering, University of Arkansas, Fayetteville, AR 72701, USA
[2]Department of Electrical and Computer Engineering, University of Nebraska–Lincoln, Lincoln, NE 68588, USA

Corresponding author: Jingxian Wu (wuj@uark.edu)

This work was supported by the National Science Foundation under Grant ECCS-1711087.

ABSTRACT Low latency change detection aims to minimize the detection delay of an abrupt change in probability distributions of a random process, subject to certain performance constraints such as the probability of false alarm (PFA). In this paper, we study the low latency detection of bearing faults of direct-drive wind turbines (WT), by analyzing the statistical behaviors of stator currents generated by the WT in real-time. It is discovered that the presence of fault will affect the statistical distribution of WT stator current amplitude at certain frequencies. Since the signature of a fault can appear in one of the multiple possible frequencies, we need to monitor the signals on multiple frequencies simultaneously, and each possible frequency is denoted as a candidate. Based on the unique properties of WT bearing faults, we propose a new multi-candidate low latency detection algorithm that can combine the statistics of signals from multiple candidate frequencies. The new algorithm does not require a separate training phase, and it can be directly applied to the stator current data and perform online detection of various possible bearing faults. The theoretical performance of the proposed algorithm is analytically identified in the form of upper bounds of the PFA and average detection delay (ADD). The algorithm allows flexible parametric adjustment of the tradeoff between PFA and ADD.

INDEX TERMS Bearing fault, fault detection, quickest change detection, wind turbine.

I. INTRODUCTION

Wind is one of the top sources of renewable energy. Wind turbines (WTs) usually operate in harsh environments, and hence, are more susceptible to failures compared to other type of energy generators [1]. WT fault diagnosis and prediction are highly important for the wind power industry. WT faults could be caused by exhaustion or manufacturing defect, and they will cause unexpected outage and result in economic loss. One way to minimize this loss is to predict and detect a malfunction before critical damage is done. With this taken into consideration, online methods that can detect fault in real-time have a clear advantage over offline methods. In addition, online diagnosis methods allow automatic remote monitoring of the WT operation conditions, and this is crucial for WT located in remote areas.

WTs are complex systems with multiple subsystems. Each subsystem may be subject to several different types of faults. A comprehensive survey of WT subsystem faults and methods of WT fault diagnosis can be found in [1], [2]. According to [3]–[6], bearing faults constitute a significant part of failures in WT. The study of this type of failure is not new [3], [7]–[11]. The most commonly used type of bearings is the ball bearing [1]. Bearing faults can be classified into four groups: inner race, outer race, cage and rollers faults. Many defects start with the fault of bearings.

Different methods are used to identify WT compartments failures, including analyzing mechanical vibrations, acoustic emission, temperature, oil parameters, and electrical signals. Electrical signals, such as current signal, are available remotely, do not require additional sensors, and can be

The associate editor coordinating the review of this manuscript and approving it for publication was Fanbiao Li.

analyzed in real-time [12]. Since faults of a bearing are associated with mechanical defects, they introduce excitations at particular frequencies [7], [13]. For some bearings, these excitations can be captured by electrical signals [14].

Many bearing fault diagnosis methods have been developed in the literature, such as threshold-based method [13], [15], wavelet-based method [16], [17], and maximum likelihood fault detection [18]. Some existing works model the WT operation states as a Markov process, where the states of the Markov process are used to represent various stages of faults or degradations. The observed data can be used to identify the bearing state [19], [20]. Another category of fault detection methods resort to machine learning or AI-based algorithms such as artificial neural networks (ANN), support vector machine (SVM), decision trees, long short-term memory (LSTM) networks, etc. [21]–[26]. Both Markov-based and AI-based methods require extensive training before actual detection, and in practice, there might not be sufficient data for training purpose. Fault detection can also be performed with particle filters [27]–[29]. Particle filters are usually computationally expensive and also require extensive trainings beforehand.

Almost all previous works focus on the accuracy of fault detection and they seldom consider the detection delay. Detection delay is defined as the time difference between the moment that the fault occurs and the time instant that the fault is detected. A small detection delay can result in timely remedial actions, thus prevent catastrophic results due to extended damages and reduce economic loss. Therefore, detection delay is an essential design parameter for WT fault detection or diagnosis.

Quickest or low latency change detection is an online detection method aiming at minimizing the detection delay of an abrupt change in probability distributions of a random process. We propose to develop low latency fault detection method for WTs by analyzing the statistical properties of the stator currents generated by WTs. The stator current can be modeled as a random process, and the presence of bearing fault will affect the probability distributions of the stator current. The time instant that the fault occurs is a change point, and it can be modeled as a random variable. Generally, change detection methods can be classified into two categories, Bayesian and non-Bayesian (minimax) methods. If the prior probability of the change point is known, then the methods are Bayesian procedures, such as [30], [31]. On the other hand, when the prior probability of the change point is unknown, the low latency change detection methods are developed under the minimax or non-Bayesian criterion. Minimax methods are employed in many practical applications, since it is usually difficult to obtain prior information about change point distribution.

Cumulative sum (CUSUM) procedure [32] and Shiryaev-Roberts (SR) procedure [33] are two most commonly used minimax change detection methods, which aim at minimizing the delay with the worst case change point distribution, under the constraint of a lower bound of average run length (ARL) to false alarm. However, the above-mentioned minimax procedures are developed for systems with single pre-change and single post-change model, and they require the exact information regarding the distribution models before and after the change. Therefore, these procedures cannot be readily applied to WT fault detection since the fault signature might appear at one of several possible frequencies corresponding to several candidates for post-change models. There are limited works with unknown or uncertain post-change models under mimimax criterion. In [34], a quickest change detection algorithm is proposed in order to detect false data injection attacks (FDIA) in smart grids with time-varying dynamic models. An orthogonal matching pursuit CUSUM (OMP-CUSUM) algorithm is proposed to identify the buses under FDIA while minimizing the detection delay in [35]. However, these two works study the cases where post-change parameter has continuous support, whereas the problem of WT fault detection deals with multiple candidates for post-change models, i.e. the post-change parameter has finite and discrete support.

The objective of this paper is to develop low latency fault detection methods based on the unique properties of WT bearing faults. Since faults of a bearing introduce excitations (harmonics) into the spectrum of stator current, the analysis is performed in the frequency domain. Statistical analysis show that the amplitude of stator current at a given frequency can be modeled by using Gamma distribution, and the presence of fault will affect the parameters of the Gamma distribution. The theoretical frequency of the excitation caused by a certain fault is determined by the mechanical structure of the bearing. The actual frequency of the fault excitation might deviate from its theoretical value due to the uncertainty of some mechanical parameters. Therefore, in order to detect the fault, the detection algorithm needs to monitor signals on several candidate frequencies. To solve this problem, we propose a multi-candidate low latency fault detection procedure, which is an enhanced version of the CUSUM algorithm. The conventional CUSUM algorithm can only be used to detect change point over a single data stream, yet the proposed multi-candidate detection procedure can be used to detect change point that happens on one of many potential data streams, with the index of the data stream with the change point unknown to the detector. In addition, the theoretical performance bounds on the PFA and ADD of the multi-candidate procedure have been analytically derived. The performance bounds reveal the fundamental tradeoff between detection delay and false alarm. The proposed procedure allows flexible parametric adjustment of the tradeoff between ADD and PFA. Moreover, this detection method does not require a training stage.

The remainder of this paper is organized as follows. Section II describes the experimental setup and the process of data collection. The methods for data pre-processing and feature extraction in the frequency domain are presented in III. Section IV studies the statistical behavior of the feature behavior and establishes probability models.

The multi-candidate low latency fault detection procedure is proposed and analyzed in Section V. Experiment and simulation results are presented in Section VI, and Section VII concludes the paper.

II. EXPERIMENT SETUP AND DATA COLLECTION

A. DATA COLLECTION

Stator current signals obtained from a 160-W Southwest Wind Power Air Breeze direct-drive permanent magnet synchronous generator (PMSG) WT are used in this paper to test the fault detection algorithm. The procedure for obtaining the stator current signal was described in [13]. The WT used to record the data has six pole pairs ($p = 6$). The data was obtained while the turbine operated in variable wind speed (from 0 to 10 m/s) condition. The shaft of the stator was rotating within the speed range of 6−13 Hz. The stator current signal was recorded using National Instruments acquisition hardware with a sampling frequency $f_{\text{so}} = 10$ kHz. The signal was recorded within 100-second-long periods every 20 minutes. The total operation time of the WT was about 25 hours. The test bearing (7C55MP4017) was pretreated by removing the lubricant, in order to accelerate degradation. A bearing outer-race fault was generated artificially for a test bearing, as illustrated in [13, Fig. 12]. One of the two bearings supporting the shaft was removed to simulate eccentricity. The test wind turbine stopped rotating at the end of the experiment, due to the broken bearing cage.

Although the data were prerecorded and stored offline, they are processed sequentially in an online manner by the faulty detection algorithm to emulate practical application scenarios. That is, during the testing of the fault detection algorithm, the data are fed to the algorithm sequentially in real-time. The algorithm makes decisions based on all current and past data, without any knowledge of the future data that have not been fed to the algorithm. Even though the data were collected over a time period of about 25 hours, experiment results demonstrate that the algorithm can detect wind turbine faults just a few seconds after the occurrence of the fault. Details will be given in Section VI.

B. FAULT SIGNATURES

Bearing faults introduce excitations on particular frequencies of the stator current. Those excitations are referred to as fault signals. There are four fault modes for bearings: inner race fault, outer race fault, cage fault, and rollers defect. Each one is characterized by a frequency of a fault signal. Denote the frequencies corresponding to inner race fault, outer race fault, cage fault, and rollers defect as $f_{\text{i}}, f_{\text{o}}, f_{\text{c}}$, and f_{b}, respectively, and they can be calculated as [6]:

$$f_{\text{i}} = 0.5 \cdot N_b \cdot f_r \left(1 + \frac{D_b \cdot \cos\phi}{D_p}\right) \tag{1}$$

$$f_{\text{o}} = 0.5 \cdot N_b \cdot f_r \left(1 - \frac{D_b \cdot \cos\phi}{D_p}\right) \tag{2}$$

$$f_{\text{c}} = 0.5 \cdot f_r \left(1 - \frac{D_b \cdot \cos\phi}{D_p}\right) \tag{3}$$

TABLE 1. Fault frequencies.

Faulty compartment	f_{bf} (Hz)	$f_1 \pm f_{\text{bf}}$ (Hz)
Inner race	49.69	10.31, 109.69
Outer race	30.3	29.7, 90.3
Cage	3.78	56.22, 63.78
Rollers	2.28	57.72, 62.28

$$f_{\text{b}} = f_r \cdot \left(\frac{D_b}{D_p}\right) \cdot \left[1 - \left(\frac{D_b \cdot \cos\phi}{D_p}\right)^2\right] \tag{4}$$

where f_r is the bearing rotation frequency, D_b is the diameter of rollers, D_p is the rollers pitch diameter, N_b is the number of rollers, and ϕ is the rollers' contact angle with races.

The bearing under investigation supports the main shaft, which connects the generator with wind blades. For that reason, the vibrations excited in the bearing affect the stator current by modulating the amplitude of the fundamental frequency f_1. The fundamental frequency is defined as the frequency of the main harmonic generated by the wind turbine. Frequencies of fault signals in stator current can be calculated as

$$f_{\text{fault}} = f_1 \pm k \cdot f_{\text{bf}} \tag{5}$$

where $f_{\text{bf}} \in \{f_{\text{i}}, f_{\text{o}}, f_{\text{b}}, f_{\text{c}}\}$, and $k = 1, 2, \ldots$. Throughout this paper the value $k = 1$ is used.

For the WT that was used to obtain stator current, $f_r = \frac{f_1}{p}$, where $f_1 = 60$ Hz is the fundamental frequency, and $p = 6$ is the number of pole pairs. The parameters of the bearing supporting the shaft are, $D_b = 8$ mm, $D_p = 33$ mm, $N_b = 8$. The contact angle of the rollers is unknown and it is assumed that $\phi = 0$. The frequencies corresponding to the four different types of faults are listed in Table 1.

III. DATA PREPROCESSING AND FEATURE EXTRACTION

The procedures for data preprocessing and feature extraction are described in this section. All procedures described in this section are sequential online algorithms that process new data as they sequentially come in.

A. SYNCHRONOUS RESAMPLING

The key for fault detection is to extract fault signatures, which are located on particular frequencies relative to fundamental frequency f_1. The extraction of fault signatures (5) of a signal with non-stationary frequency f_1 is rather difficult. The signal-to-noise ratio (SNR) of bearing fault signals is usually small. High frequency resolution is required to reduce the impact of the noise, and more samples for FFT are required. However, the fundamental frequency f_1 may drift over time, which leads to time-varying frequencies corresponding to different faults. This makes it difficult to extract the information corresponding to a specific bearing fault.

In order to resolve the problem of fundamental frequency variation, the method of synchronous resampling can be applied [13]. The basic idea is to perform non-uniform sampling such that the phase difference between any two adjacent

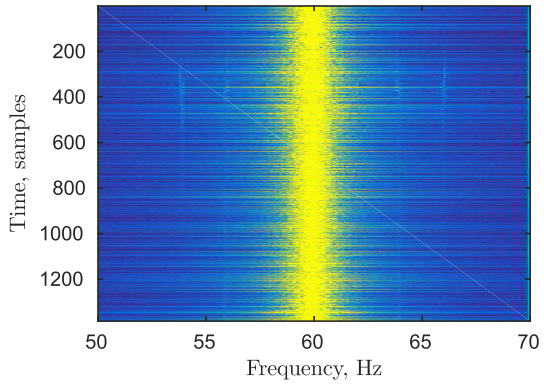

FIGURE 1. FFT of the resampled stator current. Frequency resolution is 0.1 Hz. Excitations due to faults of the bearing are visible on frequencies corresponding to bearing cage and rollers defects.

samples is constant. Details of synchronous resampling can be found in [13]. The resampling frequency used in this paper is $f_s = 1920$ Hz.

B. FREQUENCY-DOMAIN FEATURE EXTRACTION

After synchronous resampling, the time domain data are converted to the frequency domain via FFT. The resampled time-domain data are divided into length-N_f frames, and FFT is performed on each frame. The time duration for each frame is $T_f = \frac{N_f}{f_s}$ second, and the frequency resolution after FFT is $f_0 = \frac{1}{T_f} = \frac{f_s}{N_f}$. In this paper, we use $f_s = 1920$ Hz and $N_f = 19200$, and the frequency resolution after FFT is $f_0 = 0.1$ Hz.

Fig. 1 shows the spectrogram of the data after synchronous resampling. In the spectrogram, we can clearly see the main harmonic of the stator current at 60 Hz. In addition, excitations due to different types of faults are clearly visible around frequencies 54, 56, 58, 62, 64, and 66 Hz.

The fault frequencies shown in the spectrogram do not match with the theoretical values given in Table 1. The mismatch is mainly due to the assumption $\phi = 0$, which is usually non-zero. Nevertheless, the theoretical value gives a rough estimate of the exact location of the fault frequencies. Therefore, during the analysis, we can analyze a range of frequency components centering at the theoretical fault frequencies as $[f_{\text{fault}} - f_w, f_{\text{fault}} + f_w]$, where f_{fault} is one of the possible fault frequencies, and the window size $2f_w$ is chosen to be an integer multiple of the frequency resolution f_0.

For a given fault frequency f_{fault}, we can extract the amplitude of the signals over the block of frequencies in $[f_{\text{fault}} - f_w, f_{\text{fault}} + f_w]$ and form them as a feature vector. For the n-th FFT frame, denote the signal amplitude vector over the frequency range $[f_{\text{fault}} - f_w, f_{\text{fault}} + f_w]$ as $s_n = [s_{n1}, s_{n2}, \cdots, s_{nw}]$, where s_{nj} is the signal amplitude at frequency $f_{\text{fault}} - f_w + (j-1)f_0$ in the n-th frame, and $w = 2\frac{f_w}{f_0} + 1$ is the number of elements in the feature vector. Since the fault feature could appear at any one of the frequencies in $\{f_{\text{fault}} - f_w + (j-1)f_0\}_{j=1}^{w}$, we denote the set of frequencies as candidate frequencies. In this paper, $f_w = 0.5$ Hz is used, which results in a feature vector size $w = 11$.

C. ENERGY NORMALIZATION

Due to the fluctuation of the operation environment, such as wind speed, the power of the feature vector changes with respect to time. Energy normalization is performed to compensate for power fluctuations.

The energy normalization is performed by using a sliding window approach. The energy of each sample is normalized by the average energy of samples in the current and the previous N_w frames. The normalization of the j-th element in the n-th frame can be expressed as

$$x_{nj} = \frac{s_{nj}}{\sqrt{e_{nj}}} \tag{6}$$

where

$$e_{nj} = \frac{1}{N_w + 1} \sum_{k=n-N_w}^{n} |s_{kj}|^2 \tag{7}$$

is the average energy of the current and past N_w samples on the j-th candidate element.

With the normalized samples, we can form the normalized feature vector as $x_n = [x_{n1}, x_{n2}, \cdots, x_{nw}]^T$.

The normalized feature vector x_n is a random vector. We will detect the presence of fault by analyzing the statistical properties of the components in x_n.

IV. STATISTICAL ANALYSIS OF FEATURE VECTOR

Statistical analysis is performed over the frequency domain feature vector to identify their statistical properties with and without different types of faults.

A. STATISTICAL DISTRIBUTION

We obtain histogram by using the experiment data corresponding to the samples on the j-th candidate frequency $\{x_{nj}\}_n$. It is found that the distribution of x_{nj} can be accurately modeled by using Gamma distribution with probability density function (PDF) given as

$$p(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)}, \tag{8}$$

where $\Gamma(\alpha)$ is the Gamma function, and α and β are the parameters of the Gamma distribution. The parameters α and β depend on the element index j as well as the presence of fault.

The mean, m_x, and variance, σ_x^2, of a Gamma distributed random variable x are

$$m_x = \frac{\alpha}{\beta}, \quad \sigma_x^2 = \frac{\alpha}{\beta^2} \tag{9}$$

B. ONLINE PARAMETER ESTIMATION

We propose to estimate the distribution parameters, α and β, online by using a sliding window approach. At each frame, the parameters for signals on one candidate frequency can be estimated by using the corresponding data from the current and the past N_w frames.

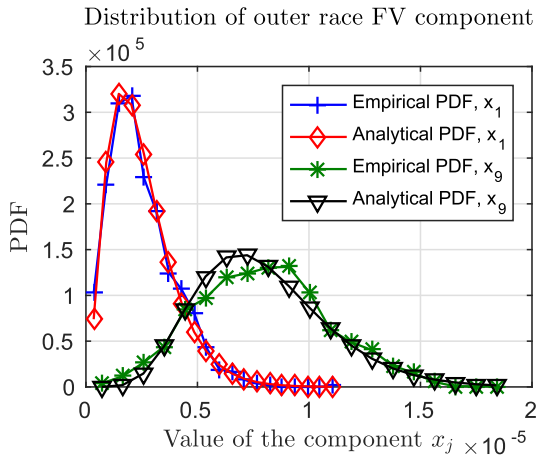

FIGURE 2. Empirical and analytical Gamma distribution of outer race feature vector components.

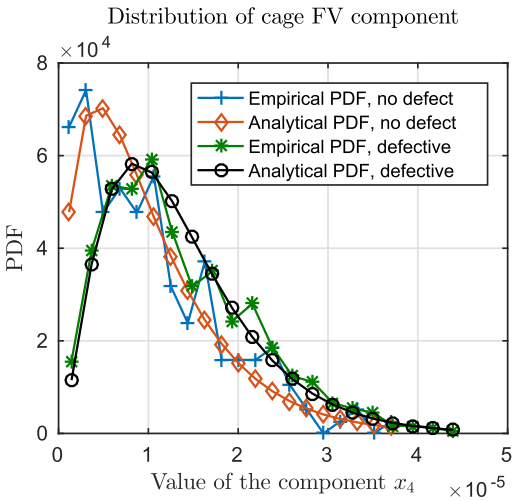

FIGURE 3. Empirical and analytical Gamma distribution of cage feature vector components in normal and defective conditions.

The mean and variance of the j-th element at the n-th frame can be estimated as

$$\hat{m}_{nj} = \frac{1}{N_w + 1} \sum_{k=n-N_w}^{n} x_{kj} \tag{10}$$

$$\hat{\sigma}^2_{nj} = \frac{1}{N_w} \sum_{k=n-N_w}^{n} (x_{kj} - \hat{m}_{nj})^2 \tag{11}$$

It should be noted that the factor $\frac{1}{N_w}$ is used in (11) to ensure the estimation is unbiased.

From (9), at the n-th frame, we can obtain an estimate of the parameters α and β for the j-th element as

$$\hat{\alpha}_{nj} = \frac{\hat{m}^2_{nj}}{\hat{\sigma}^2_{nj}}, \quad \hat{\beta}_{nj} = \frac{\hat{m}_{nj}}{\hat{\sigma}^2_{nj}} \tag{12}$$

Fig. 2 shows the empirical and analytical pdfs of x_{n1} and x_{n9} from the feature vector related to the outer race fault, by using parameters estimated with (12). In the experiment setup, the outer race fault is artificially introduced at the beginning of the experiment and its signature is shown in x_{n9}, which is at the frequency of 30.0 Hz. There is no fault signature at x_{n1}, which is at the frequency of 29.2 Hz. The analytical pdf is obtained by evaluating (8) by using parameters estimated from (12). The empirical pdf is obtained by normalizing the histogram of the experiment data. There is a good match between the analytical and empirical pdf. Also, there is an obvious difference in distribution between x_{n1} and x_{n9} due to the presence of fault signature on x_{n9}.

Fig. 3 shows the distribution of x_{n4} from the feature vector related to the cage fault, before and after the fault occurs. In the experiment, the cage fault starts to appear at $n = 130$ frame, and the fault signature is at $56.2 - f_w + (j - 1)f_0 = 56$ Hz, which corresponds to an index of $j = 4$ in the feature vector. The pdf corresponding to the non-defective case is calculated by using data with $n < 130$, and the pdf corresponding to the defective case is calculated by using data with $n \geq 130$. There is a clear difference between the distributions before and after the fault occurs.

Moreover, when the fault is not present, it is discovered that the elements in a fault feature vector are identically distributed. That is, when the fault is not present, $\hat{\alpha}_{nj} \approx \hat{\alpha}_{mj}$ and $\hat{\beta}_{nj} \approx \hat{\beta}_{mj}$ for $m \neq n$. Such a property can be used to further improve the estimation accuracy of the parameters.

It should be noted the proposed frequency domain modeling is robust against noise. Noise in real-world systems can be modeled as additive white Gaussian noise (AWGN), which has uniform power spectrum density in the frequency domain. The flat noise spectrum means the noise power is uniformly distributed over the entire frequency bands, which results in a very low noise power over a small frequency range. On the other hand, the fault signature is embedded over a very small frequency range. Thus we can significantly increase the signal-to-noise ratio (SNR) by focusing on the analysis of the signal over a very small range of frequency bands. As a result, the presence of noise will have negligible impacts regarding the performance of proposed algorithm.

V. LOW LATENCY FAULT DETECTION

The sequential low latency fault detection algorithm is presented in this section. The detection algorithm is formulated by exploring the statistical distributions of the feature vector before and after the fault occurs.

A. FORMULATION OF HYPOTHESIS

For the fault detection, define the two hypotheses for the variable x_{nj}

$$H_0 : \text{No fault} \tag{13}$$

$$H_1 : \text{Fault is present} \tag{14}$$

As shown in Figs. 2 and 3, the presence of fault will affect the values of α and β for the corresponding element in the

feature vector. Denote the parameters before and after the change point as $(\alpha^{(0)}, \beta^{(0)})$ and $(\alpha^{(1)}, \beta^{(1)})$, respectively.

For a Gamma distributed random variable x with energy normalized to unit, from (9), we have

$$m_x^2 + \sigma_x^2 = \frac{\alpha^2}{\beta^2} + \frac{\alpha}{\beta^2} = 1 \tag{15}$$

which yields

$$\beta = \sqrt{\alpha(\alpha+1)} \tag{16}$$

Therefore we only need to characterize the impacts of fault on α, and the corresponding β can be inferred from (16).

To quantify the impacts of fault, define a new parameter

$$\alpha_\Delta = \frac{\alpha^{(1)}}{\alpha^{(0)}} \tag{17}$$

For the outer race fault shown in Fig. 2, the average value of α_Δ is 1.4. For the cage fault shown in Fig. 3, the average value of α_Δ is 2.9.

In practical system design, α_Δ is unknown beforehand. We will treat α_Δ as an adjustable design parameter that can be set by using prior knowledge. We will show the impact of the choice of α_Δ on the detection performance in the section of experiment results.

With the definition of α_Δ and the estimation of $\hat{\alpha}_j^{(0)}$ using (12), the hypotheses can be formulated as

$$H_0 : x_{nj} \sim \text{Gamma}\left(\hat{\alpha}_j^{(0)}, \hat{\beta}_j^{(0)}\right) \tag{18}$$

$$H_1 : x_{nj} \sim \text{Gamma}\left(\hat{\alpha}_j^{(1)}, \hat{\beta}_j^{(1)}\right) \tag{19}$$

where

$$\hat{\alpha}_j^{(1)} = \hat{\alpha}_j^{(0)}\alpha_\Delta \tag{20}$$

$$\hat{\beta}_j^{(k)} = \sqrt{\hat{\alpha}_j^{(k)}(\hat{\alpha}_j^{(k)}+1)}, \quad \text{for } k = 0, 1 \tag{21}$$

With the hypothesis formulated in (18), define the likelihood ratio (LR) as

$$\lambda_{nj} = \frac{p(x_{nj}; \hat{\alpha}_j^{(1)}, \hat{\beta}_j^{(1)})}{p(x_{nj}; \hat{\alpha}_j^{(0)}, \hat{\beta}_j^{(0)})} \tag{22}$$

Combining (8) and (22) yields

$$\begin{aligned}\log \lambda_{nj} = {} & \hat{\alpha}_j^{(1)} \log \hat{\beta}_j^{(1)} - \hat{\alpha}_j^{(0)} \log \hat{\beta}_j^{(0)} \\ & + \left(\hat{\alpha}_j^{(1)} - \hat{\alpha}_j^{(0)}\right) \log x_{nj} - x_{nj}\left(\hat{\beta}_j^{(1)} - \hat{\beta}_j^{(0)}\right) \\ & + \log \Gamma(\hat{\alpha}_j^{(0)}) - \log \Gamma(\hat{\alpha}_j^{(1)})\end{aligned} \tag{23}$$

B. LOW LATENCY FAULT DETECTION ALGORITHM

Based on the hypothesis and LR formulated in the previous subsection, we propose a multi-candidate low latency fault detection algorithm.

Assume a certain fault occurs at frame θ and the fault signature is on the μ-th candidate frequency in the feature vector. Both θ and μ are random variables. That is, when $n < \theta$, x_{nj} follows the distribution specified by H_0; when $n \geq \theta$, $x_{n\mu}$ follows the distribution specified by H_1, and x_{nk} follows the distribution of H_0 for all n and $k \neq \mu$.

Denote the detected change point as $\hat{\theta}$. Then the PFA can be defined as

$$\text{PFA} = \text{P}(\hat{\theta} < \theta) = \sum_{k=1}^{\infty} \pi_k \text{P}(\hat{\theta} < k) \tag{24}$$

where $\pi_k = \text{P}(\theta = k)$ is the prior probability of the change point.

The well known CUSUM and SR procedures are designed for the change point detection in a single time sequence. They cannot be readily applied to the problem encountered in this paper, where there are multiple parallel data streams (multiple candidate frequencies), and the change point occurs in at most one of the data stream (candidate frequency).

To address this problem, we propose a multi-candidate low latency detection procedure as follows.

Definition 1 (Multi-candidate detection procedure): With the frequency domain feature vector x_n, the detected change point is

$$\hat{\theta} = \inf\{n : C_n \geq A\} \tag{25}$$

where we set $\inf\{\emptyset\} = \infty$, A is a pre-defined threshold determined by the PFA, the test statistic C_n is defined as $C_n = \sum_{j=1}^{w} C_{n,j}$, $C_{n,j}$ is the test statistic for the j-th element, and it can be recursively calculated as

$$C_{n+1,j} = \max(1, C_{n,j})\lambda_{n+1,j} \tag{26}$$

and $C_{1,j} = 1$.

This detection procedure is an enhanced version of the CUSUM procedure. It combines the test statistic $\{C_{n,j}\}_{j=1}^{w}$ to determine the change point. The CUSUM procedure can be considered as a special case of the multi-candidate procedure when $w = 1$. It should be noted that the multi-candidate test procedure proposed in Definition 1 is different from the procedure proposed in [36]. In [36], the summation is performed over the logarithm of the CUSUM statistic, i.e., $\sum_{j=1}^{w} \log C_{n,j}$, which corresponds to the multiplication of the linear CUSUM statistic $C_{n,j}$.

C. PERFORMANCE ANALYSIS

Next, we study the performance of the multi-candidate test procedure. Let P_k and $\mathbb{E}_k$ denote the probability measure and the corresponding expectation when the change occurs at $\theta = k$. With such a notation, P_∞ and $\mathbb{E}_\infty$ can be used to represent the probability measure and expectation before the change point, that is, $\theta = \infty$.

Lemma 1: The test statistic C_n is a sub-martingale under the probability measure P_∞, and $\mathbb{E}_\infty[C_n] \leq nw$.

Proof: The proof is given in Appendix A. ■

Theorem 1: Assume the elements in the feature vector x_n are independent. With the multi-candidate procedure defined in 1, the probability of false alarm is upper bounded by

$$PFA \leq \frac{\bar{\theta} w}{A} \tag{27}$$

where $\bar{\theta} = \sum_{k=0}^{\infty} \pi_k k$ *is the prior mean of the change point* θ*, and* w *is the number of candidates.*

Proof: The proof is given in Appendix B ∎

Next, we study the detection delay of the multi-candidate test procedure. It is difficult to obtain the exact ADD of the detection procedure. Instead, we will obtain an asymptotic upper bound. To facilitate analysis, define

$$f(x) \underset{x\to x_0}{\preceq} g(x) \Longleftrightarrow \lim_{x\to x_0} \frac{f(x)}{g(x)} \leq 1 \tag{28}$$

Theorem 2: Assume the elements in the feature vector x_n *are independent, and the frames are independent in time. Let PFA* $< \alpha$*. As* $\alpha \to 0$*, we have*

$$\mathbb{E}[\hat{\theta} - \theta | \hat{\theta} > \theta] \underset{\alpha\to 0}{\preceq} \frac{1}{1-\alpha} \frac{|\log\alpha| + \log\bar{\theta} + \log w}{D_{10}} \tag{29}$$

where $\bar{\theta}$ *is the priori mean of the change point,* w *is the length of the feature vector, and*

$$D_{10} = \int p(x; \alpha^{(1)}, \beta^{(1)}) \log \frac{p(x; \alpha^{(1)}, \beta^{(1)})}{p(x; \alpha^{(0)}, \beta^{(0)})} dx \tag{30}$$

is the Kullback-Leibler (KL) divergence between the distributions $p(x; \alpha^{(1)}, \beta^{(1)})$ *and* $p(x; \alpha^{(0)}, \beta^{(0)})$*.*

Proof: The proof is given in Appendix C. ∎

In Theorem 2, the asymptotic upper bound of the ADD is inversely proportional to the KL divergence D_{10}. The KL divergence measures the difference between two distributions. When the difference between the two distributions before and after the change point is large, the detection delay is in general smaller. The asymptotic ADD upper bound is a decreasing function of the PFA α, which demonstrates the tradeoff relationship between the PFA and ADD. In addition, the asymptotic ADD upper bound is proportional to $\log w$, which means we can reduce the detection delay by reducing the number of candidate frequencies. On the other hand, the value of w must be large enough to ensure that the frequency corresponding to the faulty feature is included in the feature vector.

D. COMPLEXITY ANALYSIS

In order to apply the proposed multi-candidate detection procedure to on-line data in each time-frame, the procedure needs to go through 3 stages: 1) data preprocessing, 2) feature extraction, and 3) test statistic computation. In stage 1, the computation is dominated by the fast Fourier transform (FFT) operation that has a complexity of $\mathcal{O}(N_f \log N_f)$, where N_f is the frame size. The second stage involves operations such as energy normalization, mean and variance estimation. Each of these operations has a complexity of $\mathcal{O}(N_w)$, where N_w is the number of past time-frames (samples) used for parameter estimation and energy normalization. Finally, the third stage is dominated by the operation described in equation (26), which has a complexity of $\mathcal{O}(w)$, where w is the feature vector size, i.e., the number of candidate frequencies.

In this paper, we set $N_f = 19200$, $N_w = 100$, and $w = 11$. As a result, the overall complexity of the proposed algorithm is dominated by stage 1, which is $\mathcal{O}(N_f \log N_f)$. That means the frame size N_f is the most dominating factor in the computational complexity of the algorithm. A frame of $N_f = 19200$ samples with a sampling rate of $f_s = 1920$ Hz corresponds to a frame duration of $T_f = 10$ seconds. Thus as long as the algorithm can finish the computation within 10 seconds, the algorithm can be implemented in an online fashion. Our simulation and experiment results show that the calculation time of the algorithm is much less than 10 seconds, thus the algorithm can be directly applied to online data.

VI. NUMERICAL ANALYSIS

In this section, we demonstrate the performance of the proposed bearing fault detection algorithm with both experiment and synthesized simulation data. Since the experiment data is obtained from only one trial, it is important to obtain the results with simulation data to analyze the bounds on performance metrics.

A. EXPERIMENT RESULTS

We test the performance of the proposed bearing fault detection algorithm with the experiment data. In the data processing, we set the FFT frame size as $N_f = 19,200$, a resampling frequency of $f_s = 1,920$ Hz, which result in a frequency resolution of $f_0 = 0.1$ Hz and a frame duration of $T_f = \frac{N_f}{f_s} = 10$ sec. The size of the feature vector is $w = 11$, which corresponds to a frequency span of $wf_0 = 1.1$ Hz. The number of past time-frames (samples) used for parameter estimation and energy normalization is set as $N_w = 100$.

In the experiment setup, the inner race fault and outer race fault are artificially introduced at the beginning of the experiment. The cage fault and roller fault starts to appear at the $n = 130$th frame.

The performance of the proposed low latency fault detection algorithm can be tuned by setting the values of two parameters, the degree of change α_Δ and the detection threshold A.

The value of α_Δ is determined by the distributions before and after the change point. If the location of the change point is known, we can estimate the value of α_Δ by using data before and after the change point. The value α_Δ may be different for different types of faults. For example, for the outer race fault shown in Fig. 2, the average value of α_Δ is 1.4, while for the cage fault shown in Fig. 3, the average value of α_Δ is 2.9. We obtained these empirical values using offline analysis based on the experimental data, where we knew the exact location of the change point, i.e., the time instant when a particular fault occurs.

However, in case of online change point (fault) detection, we will not have the exact knowledge of α_Δ beforehand, as it requires the exact location of change point. Therefore, in online change detection, α_Δ is treated as an adjustable design parameter that can be set by using prior knowledge. We will show the performance of the proposed algorithm under different values of α_Δ.

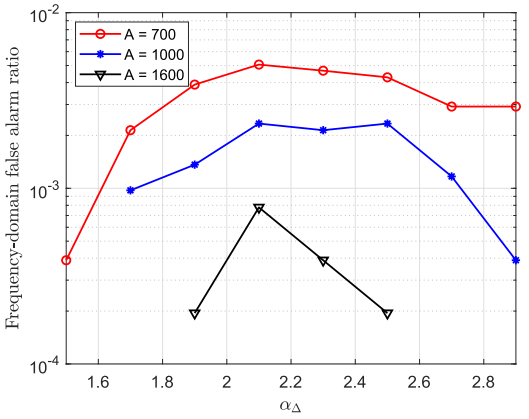

FIGURE 4. Frequency-domain false alarm ratio v.s. α_Δ.

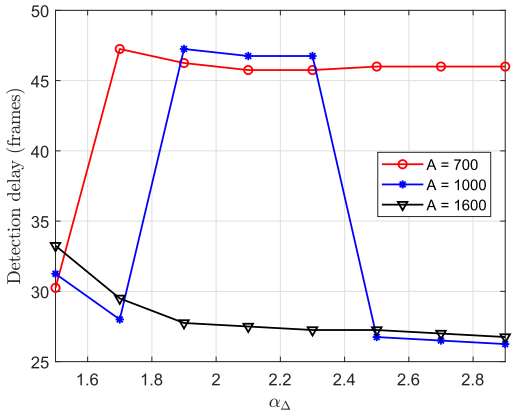

FIGURE 5. Detection delay v.s. α_Δ.

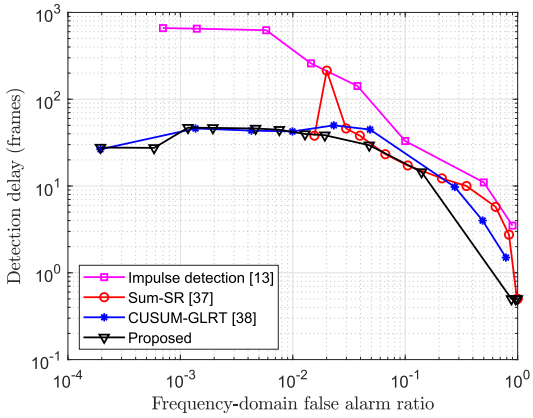

FIGURE 6. Detection delay v.s. frequency-domain false alarm ratio.

Since the experiment data are obtained under one trial, the PFA cannot be directly calculated from the experiment data. To address this issue, we propose to study the behaviors of false alarm by using the frequency-domain data streams without fault. Specifically, for each of the 4 faults, we extract features from several groups of frequency-domain data centered around frequencies which are outside the range of $[f_{\text{fault}} - f_w, f_{\text{fault}} + f_w]$, where f_{fault} denotes the theoretical fault frequency (c.f. Table 1). The detection procedure is then applied to these frequency-domain data streams and the total number of false alarms is recorded. The frequency-domain false alarm ratio is defined as the ratio between the total number of false alarms and the total number of time-frames multiplied by the number of groups. The frequency-domain false alarm ratio can serve as an indicator for PFA. Fig. 4 shows the frequency-domain false alarm ratio as a function of α_Δ, under various values of the threshold A. The relationship between frequency-domain false alarm ratio and α_Δ is not monotonic. Under a fixed threshold A, the frequency-domain false alarm ratio is concave or quasi-concave in α_Δ. The choice of α_Δ needs to consider the tradeoff between the detection delay and frequency-domain false alarm ratio.

The detection delay of the proposed multi-candidate detection algorithm is shown in Fig. 5. This metric is obtained by running the proposed detection procedure 4 times with four different sets of candidate frequencies corresponding to the 4 types of faults and then computing the average delay in detecting these 4 faults. The detection delays are shown in the units of frames, with the duration of each frame being 10 seconds. Under different configurations, the detection delay ranges between 27 to 48 frames, which correspond to an absolute delay of 270 to 480 seconds. Based on our statistical analysis of the experiment data, the average true value of α_Δ is around 2.

Fig. 6 shows the detection delay as a function of the frequency domain false alarm ratio. This figure demonstrates the tradeoff relationship between delay and false alarm. For comparison, results obtained from three other detection algorithms, namely the conventional impulse detection algorithm [13], SUM-Shiryaev-Roberts (SUM-SR) algorithm [37], and CUSUM-GLRT [38] algorithm, are also shown in the figure. SUM-SR procedure is an enhanced version of the SR procedure, where all the local SR statistics corresponding to individual post-change models are combined together. In all algorithms, we set $\alpha_\Delta = 2.0$, and different tradeoff points are achieved by adjusting the threshold value A. It is observed that the proposed multi-candidate algorithm slightly outperforms the existing algorithms.

B. SIMULATION RESULTS

Synthesized simulation data are used to verify the analytical bounds derived in this paper. In the simulation setup, we model the occurrence of the fault as a geometric distribution, that is, $\pi_k = (1-\rho)^{k-1}\rho$, where $\rho = 0.1$. The value of α_Δ is set to 2. The pre-change distribution parameters for each candidate frequency $\alpha_j^{(0)}$ follows a uniform distribution between $[1 \times 10^{-5}, 5 \times 10^{-5}]$. The simulation results are obtained by averaging over 10,000 Monte-Carlo trials, where only a single type of fault is simulated in each trial.

Fig. 7 shows the PFA as a function of the detection threshold A. Under the same threshold A, the CUSUM-GLRT algorithm has the lowest PFA, followed by the proposed algorithm, and the PFA of the SUM-SR algorithm is significantly higher than the other two algorithms. The PFA of all three algorithms are lower than the analytical upper bounds

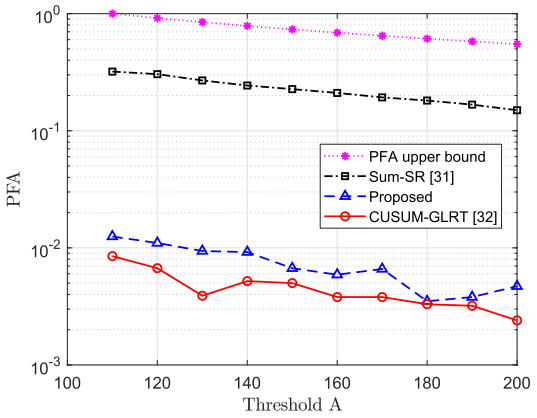

FIGURE 7. Probability of false alarm v.s. threshold.

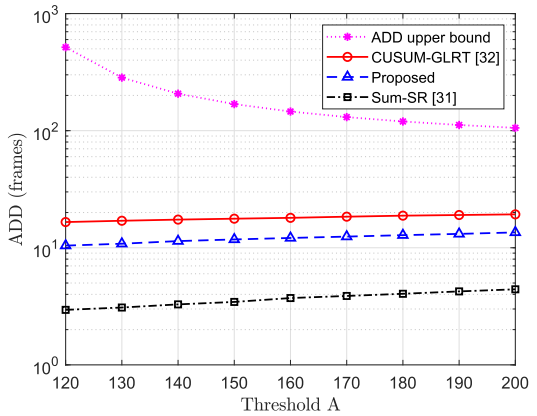

FIGURE 8. Average detection delay v.s. threshold.

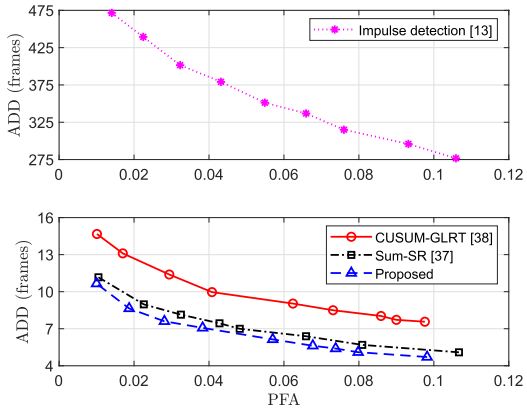

FIGURE 9. Average detection delay vs. probability of false alarm.

developed in Theorem 1. It should be noted that a lower PFA does not necessarily mean a better performance due to the effects of the ADD.

The ADDs of the three different algorithms are shown in Fig. 8 as a function of the detection threshold. Under the same threshold A, the SUM-SR algorithm has the lowest ADD at the cost of a large PFA, followed by the proposed multi-candidate detection procedure and CUSUM-GLRT, respectively. The ADD of the proposed low latency algorithm is lower than the asymptotic upper bound presented in Theorem 2.

The tradeoff relationship between ADD and PFA for the simulated data is illustrated in Fig. 9, where the ADD is shown as a function of PFA. Different points on the tradeoff curve are obtained by adjusting the detection threshold A. Among the four algorithms, the proposed algorithm achieves the best ADD-PFA tradeoff, followed by the SUM-SR algorithm, the CUSUM-GLRT algorithm, and the impulse detection algorithm, respectively. Since the impulse detection algorithm is not a quickest change detection algorithm, the ADD of the impulse detection algorithm is significantly larger than the other three algorithms, which are developed by using sequential statistics. At a PFA of 0.04, the ADDs of the proposed algorithm, the SUM-SR algorithm, the CUSUM-GLRT algorithm, and the impulse detection algorithm are 7, 7.5, 10.7, and 382 frames, respectively.

VII. CONCLUSION

The detection of bearing faults of direct-drive wind turbines have been studied in this paper under the framework of low latency change point detection. The amplitude of stator current at a given frequency was modeled by using the Gamma distribution, and the presence of fault will affect the parameters of the Gamma distribution. We defined a new parameter, α_Δ, to measure the impact of the fault on the Gamma distribution. A multi-candidate low latency change point detection algorithm, which includes the conventional CUSUM algorithm as a special case, has been proposed to detect the various faults. The theoretical performance of the proposed algorithm has been analytically identified in terms of upper bounds of the probability of false alarm and average detection delay. This algorithm can flexibly achieve different tradeoff between the PFA and ADD. This proposed low latency fault detection algorithm does not require a training phase and can be readily applied to the collected data in real-time. Experiment and simulation results demonstrated that the proposed algorithm can detect faults in real-time with minimum delays, and it outperforms existing algorithms. The proposed multi-candidate change detection can also be applied to other change detection scenarios with the change occurring in one of many possible data streams.

APPENDIXES

APPENDIX A

PROOF OF LEMMA 1

To show that C_n is a sub-martingale under P_∞, it is sufficient to prove that $\mathbb{E}_\infty[C_{n+1}|C_n] \geq C_n$. For the LR $\lambda_{n,j}$, we have

$$\mathbb{E}_\infty(\lambda_{n,j}) = \int \frac{p(x_{nj}; \hat{\alpha}_j^{(1)}, \hat{\beta}_j^{(1)})}{p(x_{nj}; \hat{\alpha}_j^{(0)}, \hat{\beta}_j^{(0)})} p(x_{nj}; \hat{\alpha}_j^{(0)}, \hat{\beta}_j^{(0)}) dx_{nj} = 1$$

Based on the definition of C_n, we have

$$\begin{aligned}\mathbb{E}_\infty[C_{n+1}|C_n] &= \sum_{j=1}^{w}\mathbb{E}_\infty[C_{n+1,j}|C_n]\\ &= \sum_{j=1}^{w}\max(1, C_{n,j})\mathbb{E}_\infty(\lambda_{n+1,j})\\ &\geq \sum_{j=1}^{w} C_{n,j} = C_n\end{aligned}$$

Thus C_n is a sub-martingale.

In addition,

$$\begin{aligned}\mathbb{E}_\infty(C_n) &= \mathbb{E}_\infty\left(\sum_{j=1}^{w}\max_{1\leq k\leq n}\Lambda_{k:n,j}\right)\\ &\leq \mathbb{E}_\infty\left(\sum_{j=1}^{w}\sum_{k=1}^{n}(\Lambda_{k:n,j})\right)\\ &= nw.\end{aligned}$$

APPENDIX B
PROOF OF THEOREM 1

C_n is a sub-martingale under P_∞. Based on Doob's inequality, we have

$$\begin{aligned}\mathrm{P}_\infty(\hat{\theta} < n) &= \mathrm{P}_\infty\left(\max_{1\leq k<n} C_n \geq A\right)\\ &\leq \frac{\mathbb{E}_\infty(C_n)}{A} \leq \frac{nw}{A}\end{aligned} \tag{31}$$

Therefore, the PFA can be calculated as

$$\begin{aligned}\mathrm{PFA} &= \sum_{k=1}^{\infty}\pi_k \mathrm{P}_k(\hat{\theta} < k)\\ &= \sum_{k=1}^{\infty}\pi_\infty \mathrm{P}_k(\hat{\theta} < k)\\ &\leq \sum_{k=1}^{\infty}\pi_k\frac{kw}{A} = \frac{\bar{\theta}w}{A}.\end{aligned}$$

APPENDIX C
PROOF OF THEOREM 2

Assume the fault happens at the k-th frame on the j-th element. Define a new stopping time

$$\tau(A) = \inf\{n \geq k : Z_j^{k:n} \geq \log A\} \tag{32}$$

where

$$Z_j^{k:n} = \sum_{l=k}^{n}\log\lambda_{l,j} \tag{33}$$

From (26), we have

$$C_{n,j} = \max_{1\leq k\leq n}\prod_{l=k}^{n}\lambda_{l,j} \tag{34}$$

Thus $C_n \geq C_{n,j} \geq \prod_{l=k}^{n}\lambda_{l,j}$ when $n \geq k$, or equivalently, $\log C_n \geq Z_j^{k:n}$. As a result, we have $\hat{\theta} \leq \tau(A)$.

Based on the strong law of large numbers, under the probability measure P_k, $\frac{1}{n}Z_j^{k:k+n-1}$ almost surely converges in probability P_k to the KL divergence $D_{10} = \mathbb{E}_k[\log\lambda_{l,j}]$,

$$\frac{1}{n}Z_j^{k:k+n-1} \xrightarrow[n\to\infty]{\mathrm{P}_k-a.s.} D_{10}. \tag{35}$$

Define

$$T_k = \sup\left\{n \geq 1 : \left|\frac{1}{n}Z_j^{k:k+n-1} - D_{10}\right| > \epsilon\right\}. \tag{36}$$

If $\tau(A)-k > T_k$, then from (36) we have

$$\left|\frac{1}{\tau(A)-k}Z_j^{k:\tau(A)-1} - D_{10}\right| \leq \epsilon, \quad \text{if } \tau(A)-k > T_k \tag{37}$$

which implies

$$\tau(A)-k \leq \frac{Z_j^{k:\tau(A)-1}}{D_{10}-\epsilon}, \quad \text{if } \tau(A)-k > T_k \tag{38}$$

Based on the definition of $\tau(A)$ in (32), we have

$$Z_j^{k:\tau(A)-1} < \log A \tag{39}$$

Combining (38) and (39) results in

$$\tau(A)-k \leq \frac{\log A}{D_{10}-\epsilon}, \quad \text{if } \tau(A)-k > T_k \tag{40}$$

When $\log(A) > 0$ and $\epsilon < D_{10}$, the following inequality is true for both $\tau(A)-k > T_k$ and $\tau(A)-k \leq T_k$

$$\tau(A)-k \leq \frac{\log A}{D_{10}-\epsilon}+T_k, \quad \text{if } \log A>0 \text{ and } \epsilon < D_{10} \tag{41}$$

Given the convergence condition in (35), we have $\mathbb{E}(T_k) < \infty$. Since ϵ can be arbitrarily small, we can let $\epsilon \to 0$. From Theorem 1, we can set $A = \frac{\bar{\theta}w}{\alpha}$ to guarantee PFA $< \alpha$. Thus when $\alpha \to 0$ and $\epsilon \to 0$,

$$\mathbb{E}_k[\tau(A) - k] \underset{\alpha\to 0}{\preceq} \frac{\log|\alpha| + \log\bar{\theta} + \log w}{D_{10}}. \tag{42}$$

Since $\hat{\theta}$ is upper bounded by $\tau(A)$, we have

$$\mathbb{E}_k[\hat{\theta} - k] \underset{\alpha\to 0}{\preceq} \frac{\log|\alpha| + \log\bar{\theta} + \log w}{D_{10}}. \tag{43}$$

When $\alpha \to 0$, the right hand side of (43) is positive. Thus the inequality in (43) still holds if we replace $\hat{\theta}_1 - k$ by $(\hat{\theta}_1 - k)^+$.

The ADD can be calculated as

$$\mathbb{E}[\hat{\theta} - \theta|\hat{\theta} > \theta] = \frac{1}{\mathrm{P}_\infty(\hat{\theta} \geq \theta)}\sum_{k=1}^{\infty}\pi_k\mathbb{E}_k(\hat{\theta} - k)^+ \tag{44}$$

With the constraint PFA $< \alpha$, we have $\mathrm{P}_\infty(\hat{\theta} \geq \theta) = 1 - \mathrm{PFA} \geq 1-\alpha$. Combining (32), (44), and the above results yields (29).

REFERENCES

[1] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault Diagnosis—Part I: Components and subsystems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6536–6545, Oct. 2015.

[2] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault Diagnosis—Part II: Signals and signal processing methods," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6546–6557, Oct. 2015.

[3] M. Blodt, P. Granjon, B. Raison, and G. Rostaing, "Models for bearing damage detection in induction motors using stator current monitoring," *IEEE Trans. Ind. Electron.*, vol. 55, no. 4, pp. 1813–1822, Apr. 2008.

[4] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *Proc. IEEE Power Electron. Mach. Wind Appl.*, Jun. 2009, pp. 1–7.

[5] Z. Daneshi-Far, G. A. Capolino, and H. Henao, "Review of failures and condition monitoring in wind turbine generators," in *Proc. XIX Int. Conf. Electr. Mach. (ICEM)*, Sep. 2010, pp. 1–6.

[6] X. Gong and W. Qiao, "Bearing fault diagnosis for direct-drive wind turbines via current-demodulated signals," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3419–3428, Apr. 2013.

[7] R. R. Schoen, T. G. Habetler, F. Kamran, and R. G. Bartfield, "Motor bearing damage detection using stator current monitoring," *IEEE Trans. Ind. Appl.*, vol. 31, no. 6, pp. 1274–1279, Nov./Dec. 1995.

[8] F. Immovilli, M. Cocconcelli, A. Bellini, and R. Rubini, "Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals," *IEEE Trans. Ind. Electron.*, vol. 56, no. 11, pp. 4710–4717, Nov. 2009.

[9] B. M. Ebrahimi, J. Faiz, and M. J. Roshtkhari, "Static-, Dynamic-, and mixed-eccentricity fault diagnoses in permanent-magnet synchronous motors," *IEEE Trans. Ind. Electron.*, vol. 56, no. 11, pp. 4727–4739, Nov. 2009.

[10] W. Yang, P. J. Tavner, and M. R. Wilkinson, "Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train," *IET Renew. Power Gener.*, vol. 3, no. 1, pp. 1–11, 2009.

[11] F. Immovilli, A. Bellini, R. Rubini, and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison," *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1350–1359, Jul. 2010.

[12] L. He, L. Hao, D. Pan, and W. Qiao, "Detection of single-axis pitch bearing defect in a wind turbine using electrical signature analysis," in *Proc. IEEE Int. Electric Mach. Drives Conf. (IEMDC)*, San Diego, CA, USA, May 2019, pp. 31–36.

[13] X. Gong and W. Qiao, "Current-based mechanical fault detection for direct-drive wind turbines via synchronous sampling and impulse detection," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1693–1702, Mar. 2015.

[14] X. Gong and W. Qiao, "Current-based eccentricity detection for direct-drive wind turbines via synchronous sampling," in *Proc. IEEE Energy Convers. Congr. Exposit.*, Sep. 2013, pp. 2972–2976.

[15] C. Peeters, P. Guillaume, and J. Helsen, "Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy," *Renew. Energy*, vol. 116, pp. 74–87, Feb. 2018.

[16] X. Gong, W. Qiao, and W. Zhou, "Incipient bearing fault detection via wind generator stator current and wavelet filter," in *Proc. 36th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2010, pp. 2615–2620.

[17] W. Teng, W. Wang, H. Ma, Y. Liu, Z. Ma, and H. Mu, "Adaptive fault detection of the bearing in wind turbine generators using parameterless empirical wavelet transform and margin factor," *J. Vibrat. Control*, vol. 25, no. 6, pp. 1263–1278, Jan. 2019.

[18] I. S. Bozchalooi and M. Liang, "Parameter-free bearing fault detection based on maximum likelihood estimation and differentiation," *Meas. Sci. Technol.*, vol. 20, no. 6, May 2009, Art. no. 065102.

[19] H. Ocak and K. A. Loparo, "A new bearing fault detection and diagnosis scheme based on hidden Markov modeling of vibration signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, May 2001, pp. 3141–3144.

[20] X. Zhang, R. Xu, C. Kwan, S. Y. Liang, Q. Xie, and L. Haynes, "An integrated approach to bearing fault diagnostics and prognostics," in *Proc. Amer. Control Conf.*, vol. 4, Jun. 2005, pp. 2750–2755.

[21] X. An, D. Jiang, and S. Li, "Application of back propagation neural network to fault diagnosis of direct-drive wind turbine," in *Proc. World Non-Grid-Connected Wind Power Energy Conf.*, Nov. 2010, pp. 1–5.

[22] Y. Wang and D. G. Infield, "Neural network modelling with autoregressive inputs for wind turbine condition monitoring," in *Proc. Int. Conf. Sustain. Power Gener. Supply (SUPERGEN)*, 2012, pp. 1–6.

[23] M.-S. An, S.-J. Park, J.-S. Shin, H.-Y. Lim, and D.-S. Kang, "Implementation of automatic failure diagnosis for wind turbine monitoring system based on neural network," in *Multimedia and Ubiquitous Engineering*. Dordrecht, The Netherlands: Springer, 2013, pp. 1181–1188.

[24] I. Abdallah, V. Dertimanis, H. Mylonas, K. Tatsis, E. Chatzi, N. Dervilis, K. Worden, and E. Maguire, "Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data," in *Safety and Reliability—Safe Societies in a Changing World*. London, U.K.: CRC Press, 2018, pp. 3053–3061.

[25] W. Teng, H. Cheng, X. Ding, Y. Liu, Z. Ma, and H. Mu, "DNN-based approach for fault detection in a direct drive wind turbine," *IET Renew. Power Gener.*, vol. 12, no. 10, pp. 1164–1171, Jul. 2018.

[26] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on long short-term memory networks," *Renew. Energy*, vol. 133, pp. 422–432, Apr. 2019.

[27] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. Orchard, and G. Vachtsevanos, "Anomaly detection: A robust approach to detection of unanticipated faults," in *Proc. Int. Conf. Prognostics Health Manage.*, Oct. 2008, pp. 1–8.

[28] M. E. Orchard and G. J. Vachtsevanos, "A particle-filtering approach for on-line fault diagnosis and failure prognosis," *Trans. Inst. Meas. Control*, vol. 31, nos. 3–4, pp. 221–246, Jun. 2009.

[29] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. E. Orchard, and G. Vachtsevanos, "A probabilistic fault detection approach: Application to bearing fault detection," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2011–2018, May 2011.

[30] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probab. Its Appl.*, vol. 8, no. 1, pp. 22–46, Jan. 1963.

[31] S. Nath and J. Wu, "Bayesian quickest change point detection with multiple candidates of post-change models," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Anaheim, CA, USA, Nov. 2018, pp. 51–55.

[32] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, p. 100, Jun. 1954.

[33] S. W. Roberts, "A comparison of some control chart procedures," *Technometrics*, vol. 8, no. 3, pp. 411–430, Aug. 1966.

[34] S. Nath, I. Akingeneye, J. Wu, and Z. Han, "Quickest detection of false data injection attacks in smart grid with dynamic models," *IEEE J. Emerg. Sel. Topics Power Electron.*, to be published, doi: 10.1109/JESTPE.2019.2936587.

[35] I. Akingeneye and J. Wu, "Low latency detection of sparse false data injections in smart grids," *IEEE Access*, vol. 6, pp. 58564–58573, 2018.

[36] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, Apr. 2010.

[37] A. G. Tartakovsky and V. V. Veeravalli, "Change-point detection in multichannel and distributed systems," in *Applied Sequential Methodologies: Real-World Examples with Data Analysis*, vol. 173. Boca Raton, FL, USA: CRC Press, 2004, pp. 339–370.

[38] T. Leung Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.

SAMRAT NATH (Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology, in 2014. He is currently pursuing the Ph.D. degree in electrical engineering from the University of Arkansas, Fayetteville, AR, USA. His research interests include statistical signal analytics, information sensing and processing, optimization, wireless communication, and cyber-security of smart grid.

JINGXIAN WU (Senior Member, IEEE) received the B.S. (EE) degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1998, the M.S. (EE) degree from Tsinghua University, Beijing, in 2001, and the Ph.D. (EE) degree from the University of Missouri at Columbia, MO, USA, in 2005. He is currently a Professor with the Department of Electrical Engineering, University of Arkansas, Fayetteville. His research interests mainly focus on signal processing for large scale networks and wireless communications, cybersecurity for smart grids, and statistical data analytics. He served as the symposium or track co-chairs for a number of international conferences, such as the 2012 and 2019 IEEE International Conference on Communications and the 2009, 2015, and 2017 IEEE Global Telecommunications Conference. He served as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, from 2007 to 2011, and an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2011 to 2016. He is currently serving as an Associate Editor of IEEE ACCESS.

YUE ZHAO (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2010, and the Ph.D. degree in electrical engineering from the University of Nebraska–Lincoln, Lincoln, USA, in 2014.

He was an Assistant Professor with the Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, USA, from 2014 to 2015. Since August 2015, he has been with the University of Arkansas, Fayetteville, USA, where he is currently an Assistant Professor at the Department of Electrical Engineering. His current research interests include electric machines and drives, power electronics, and renewable energy systems. He has coauthored more than 50 articles in refereed journals and international conference proceedings and holds four granted U.S. patents.

Dr. Zhao is a member of Eta Kappa Nu. He was a recipient of the 2018 U.S. National Science Foundation CAREER Award and the Best Paper Prize of the 2012 IEEE Transportation Electrification Conference and Expo. He is an Associated Editor of the IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS and a Guest Associate Editor of the IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS.

WEI QIAO (Fellow, IEEE) received the B.Eng. and M.Eng. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 1997 and 2002, respectively, the M.S. degree in high-performance computation for engineered systems from Singapore-MIT Alliance, Singapore, in 2003, and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2008.

Since August 2008, he has been with the University of Nebraska–Lincoln, Lincoln, NE, USA, where he is currently a Professor with the Department of Electrical and Computer Engineering. His research interests include sustainable energy systems, smart grids, condition monitoring, power electronics, electric motor drives, energy storage systems, and emerging electrical energy conversion devices. He is the author or coauthor of more than 240 articles in refereed journals and conference proceedings and holds ten U.S. patents.

Dr. Qiao was a recipient of the 2010 U.S. National Science Foundation CAREER Award and the 2010 IEEE Industry Applications Society Andrew W. Smith Outstanding Young Member Award. He is an Editor of the IEEE TRANSACTIONS ON ENERGY CONVERSION and an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS and the IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS.