Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Wavelet-based Space Partitioning for Symbolic Time Series Analysis†

Venkatesh Rajagopalan vxr139@psu.edu
Asok Ray axr2@psu.edu
The Pennsylvania State University
University Park, PA 16802
Corresponding Author: Asok Ray

Keywords: *Symbolic Time Series Analysis, Wavelets, Fault Detection, Complex Systems*

***Abstract*— Recent literature has reported symbolic time series analysis of complex systems for real-time anomaly detection. A crucial aspect in this analysis is symbol sequence generation from the observed time series data. This paper presents a wavelet-based partitioning, instead of the currently practiced method of phase-space partitioning, for symbol generation. The partitioning algorithm makes use of the maximum entropy method. The wavelet-space and phase-space partitioning methods are compared with regard to anomaly detection using experimental data.**

I. Introduction

The concept of symbolic time series analysis has been recently proposed for anomaly detection in complex systems [8] [2]. A crucial step in symbolic time series analysis is partitioning of the phase space for symbol sequence generation [3]. Several techniques have been suggested in literature for symbol generation, primarily based on symbolic false neighbors [5] [4]. These techniques rely on partitioning the phase space of the system and may become cumbersome if the dimension of the phase space is large. Moreover, if the time series data is noise-corrupted, then the symbolic false neighbors would rapidly grow in number and require a large symbol alphabet to capture the pertinent information on the system dynamics. The wavelet transform largely alleviates these shortcomings and is particulary effective with noisy data obtained from higher dimensional dynamical systems. Usage of wavelet transform for symbolization was introduced by Ray [8] for real-time anomaly detection. This paper elaborates the underlying concept of partitioning the space of wavelet coefficients, instead of the phase-space partitioning [5], for symbol generation.

II. Wavelet Based Symbol Generation

Preprocessing of time series data is often necessary for extraction of pertinent information. Fourier analysis is adequate if the signal to be analyzed is stationary and if the time period is accurately known. However, Fourier analysis may not be appropriate if the signal has non-stationary characteristics such as drifts, abrupt asynchronous changes and frequency trends. The wavelet analysis alleviates these difficulties via adaptive usage of long windows for retrieving low frequency information and short windows for high frequency information [6] [9]. The ability to perform flexible localized analysis is one of the striking features of the wavelet transform.

In multi-resolution analysis (MRA) of wavelet transform, a continuous signal $f \in \mathbb{H}$, where $\mathbb{H}$ is a Hilbert space, is expressed as a linear combination of time translations of scaled versions of a suitably chosen scaling function $\phi(t)$ and the derived wavelet function $\psi(t)$. For a periodic signal, the difference between the original signal f and the reconstructed signal $f^{\#}$ can be made arbitrarily small by (i) choosing the shape of the scaling function ϕ to be close to one period of f, and (ii) matching the frame set with the signal through appropriate shifting across the time axis. This suggests that it would be possible to achieve high energy compaction at a suitable scale, if the shape of ϕ or ψ closely matches the shape of the signal f.

If the set $\{\phi_k\}$ is a frame for a Hilbert space $\mathbb{H}$ with the frame representation operator $\mathbb{L}$, then there exist positive real scalars *A* and *B* such that

$$A||f||^2 \leq ||\mathbb{L}f||^2 \leq B||f||^2 \quad \forall f \in \mathbb{H} \tag{1}$$

The above relationship is a norm equivalence and represents the degree of coherence of the signal f with respect to the frame set of scaling functions; it may be interpreted as enforcing an approximate energy transfer between the domains $\mathbb{H}$ and $\mathbb{L}(\mathbb{H})$. In other words, for all signals $f \in \mathbb{H}$, a scaled amount of energy is distributed in the coefficient domain where the scale factor lies between *A* and *B* [6]. However, the energy distribution is dependent on the signal's degree of coherence with the underlying frame $\{\phi_k\}$. For a signal f, a tight frame representation in Eq. (1) implies that only a few coefficients contain the bulk of the signal energy and hence have a relatively large magnitude. Similarly, since a noise signal w is incoherent with the set $\{\phi_k\}$, the associated frame representation must be spread out over a very large number of coefficients. Consequently, these coefficients have a relatively small magnitude [9].

Let $\tilde{f}$ be a noise corrupted version of the original signal f

†This work has been supported in part by Army Research Office (ARO) under Grant No. DAAD19-01-1-0646

0-7803-9568-9/05/$20.00 ©2005 IEEE

expressed as

$$\tilde{f} = f + \sigma\ w \tag{2}$$

where w is additive white gaussian noise with zero mean and unit variance and σ is the noise level. Then, the inner products with ϕ_k are related as:

$$< \tilde{f}, \phi_k >= \underbrace{< f, \phi_k >}_{\text{large}} + \sigma \underbrace{< w, \phi_k >}_{\text{small}} \tag{3}$$

The term due to noise may further be reduced if the scales over which coefficients are obtained are properly chosen.

For every wavelet, there exists a certain frequency called the center frequency F_c that has the maximum modulus in the Fourier transform of the wavelet. The pseudo-frequency f_p of the wavelet at a particular scale α is given by the following formula [1]:

$$f_p = \frac{F_c}{\alpha\ \Delta t} \tag{4}$$

where Δt is the sampling interval. Figure 1 depicts the center frequency associated with the Daubechies wavelet 'db4'.

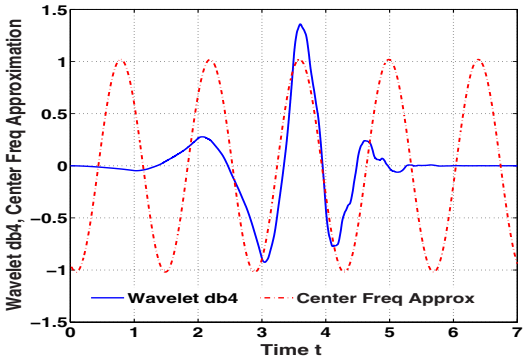

Fig. 1. Center Frequency Approximation for Wavelet db4

The Power Spectral Density (PSD) of the signal provides the information about the frequency content of the signal. This information along with Eq. (4) can be used for scale selection. The procedure of selecting the scales is summarized below:

- PSD analysis of the time series data over a selected period to find the frequencies of interest.
- Substitution of the above frequencies in place of f_p in Eq. (4) to obtain the respective scale α in terms of the known parameters F_c and Δt

The wavelet coefficients of the signal are significantly large when the pseudo-frequency f_p of the wavelet corresponds to the locally dominant frequencies in the underlying signal. Example 1 in section III illustrates how the choice of wavelet and scale affect the coefficients. Examples 2 and 3 illustrate suppression of noise and robustness issues. Example 4 illustrates enhancement of anomaly detection using symbolic dynamics and compares wavelet-space partitioning with phase-space partitioning for anomaly detection.

Once the wavelet and the scales are chosen, the wavelet coefficients are evaluated for each scale. The graphs of wavelet coefficients versus scale, at selected time shifts, are stacked starting with the smallest value of scale and ending with its largest value and then back from the largest value to the smallest value of the scale at the next instant of time shift. The arrangement of the resulting *scale series* data in the wavelet space is similar to that of the time series data in the phase space. The wavelet space is partitioned into segments of coefficients on the ordinate separated by horizontal lines [8].

A threshold-based partitioning scheme was used in [8] [2]. In this approach, the maximum and minimum of the scale series are evaluated and the ordinates between the maximum and minimum are divided into equal-sized regions. These regions are obviously mutually disjoint and thus form a partition. Each region is then labelled with one symbol from the alphabet. If the data point lies in a particular region, it is coded with the symbol associated with that region. Thus, a sequence of symbols is created from a sequence of scale series data. This type of partitioning is called uniform partitioning in the sequel.

Intuitively, it is more reasonable if the regions with more information are partitioned finer and those with sparse information are partitioned coarser. To achieve this objective, an alternative approach is proposed for partitioning, which is based on maximization of entropy and is called maximum entropy partitioning in the sequel. In this approach, the maximum entropy is achieved by the partition that induces uniform probability distribution of the symbols in the alphabet Σ. The procedure for obtaining such a partition is presented below.

Let N be the length of the scale series data and $|\Sigma|$ be the cardinality of the (finite) alphabet. The scale series data is sorted in ascending order. Starting from the first point in the sorted set, every consecutive segment of length $\lfloor \frac{N}{|\Sigma|} \rfloor$ forms a distinct element of the partition, where $\lfloor x \rfloor$ represents the greatest integer less than or equal to x.

With such a partition, a region with large information content is allotted more symbols and hence a finer partitioning is achieved in such a region. Similarly, a region with sparse information content is allotted fewer symbols and hence a coarser partitioning is achieved in such a region. So, if there is a small change in the system behavior, it is more likely to be reflected in the symbol sequence obtained under maximum entropy partitioning than under uniform partitioning.

The choice of the alphabet size $|\Sigma|$ also plays a vital role in the extraction of information. This is an area of active research. An entropy rate based approach, which has been adopted in this paper, is briefly described. Let $H(k)$ denote the entropy of the symbol sequence obtained by partitioning the data with k symbols.

$$H(k) = -\sum_{i=1}^{i=k} p_i log_2 p_i \tag{5}$$

where p_i represents the probability of occurrence of the symbol σ_i and obviously $H(1) = 0$. If the underlying data

has sufficient information content, then the entropy achieved under maximum entropy partitioning would be $log_2(k)$ corresponding to the uniform distribution. The entropy rate, with respect to the number of symbols, is given by

$$h(k) = H(k) - H(k-1) \quad \forall k \geq 2 \tag{6}$$

An algorithm for choosing the number of symbols is provided below.

1. Set k = 2. Choose a threshold ϵ_h, where $0 < \epsilon_h << 1$.
2. Sort the scale series data set (of length N) in the ascending order.
3. Every consecutive segment of length $\lfloor \frac{N}{|\Sigma|} \rfloor$ in the sorted data set (of length N) forms a distinct element of the partition.
4. Convert the scale series sequence to a symbol sequence with the partitions obtained in Step 3. If the data point lies within or on the lower bound of a partition, it is coded with the symbol associated with that partition.
5. Compute the symbol probabilities p_i, i=1,2,...k.
6. Compute the entropy $H(k) = -\sum_{i=1}^{i=k} p_i log_2 p_i$ and the entropy rate $h(k) = H(k) - H(k-1)$
7. If $h(k) < \epsilon_h$, then exit; else increment k by 1 and go to Step 3.

A primary objective of symbolic time series analysis is achieving increased computational efficiency and accuracy for real-time anomaly detection [8]. The choice of the threshold ϵ_h depends on the signal that is analyzed and may vary for individual systems. A small ϵ_h leads to a large size of the symbol alphabet, resulting in increased computation. On the other hand, a larger ϵ_h may fail to capture the small changes in dynamics because of excessive coarse graining resulting from reduced number of symbols. Hence, ϵ_h should be chosen based on these constraints and the available computational resources.

III. Simulation and Laboratory Experimentation

This section presents the proposed methodology for symbolization of time series data with examples and illustrates the advantages.

A. Example 1: Choice of Wavelet Parameters

This example illustrates how the choice of wavelet and scales affect the coefficients that, in turn, determine the symbolic dynamics for anomaly detection [8]. Let us consider the following sinusoidal signal.

$$y(t) = cos(2\pi t) \qquad \forall t \in [-5, +5] \tag{7}$$

The frequency of $y(t)$ in Eq. (7) is 1.00 Hz. The Gaussian wavelet 9 ('gaus9') closely matches the shape of $y(t)$, in the sense that the inner product of the signal and wavelet 'gaus9' over the effective support of the wavelet is relatively large (i.e., good mean-square fit) when compared with other wavelets. A suitably scaled and translated version of the wavelet is depicted along with $y(t)$ in Figure 2.

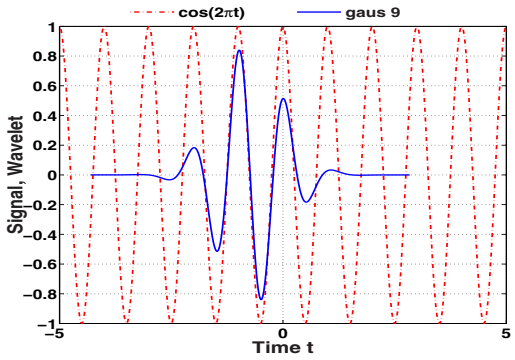

Fig. 2. Signal and Wavelet Plots

To demonstrate the impact of the chosen wavelet parameters on the coefficients, the wavelet 'db1' is also considered for comparison purposes. The signal $y(t)$ is sampled at 100 Hz ($\Delta t = 0.01s$). The wavelet coefficients of the signal $y(t)$ are obtained for various scales with both the wavelets. The norm of the coefficients corresponding to each scale and the pseudo-frequencies of the wavelet corresponding to the chosen scales are calculated. Figure 3 shows the plot of the norm of coefficients and the pseudo-frequencies of the wavelet.

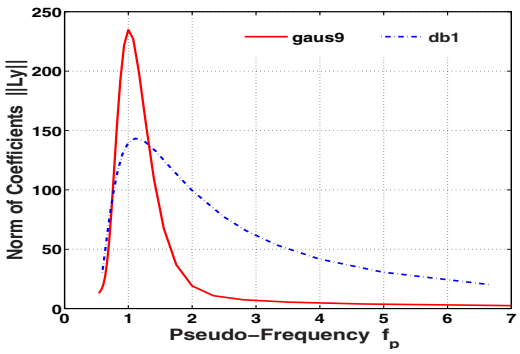

Fig. 3. Coefficient Norm and Pseudo-Frequency for Different Wavelets

It is observed from Figure 3 that, for both wavelets 'gaus9' and 'db1', the maximum of the norm is obtained at $f_p \approx 1.00$ Hz. In fact, it is exactly at 1.00 Hz for 'gaus9'. Furthermore, the value of the peak norm achieved with 'gaus9' is appreciably greater than that with wavelet 'db1'. In other words, the coefficients obtained with 'gaus9' are more significant than those obtained with 'db1'. Another observation is that the norm curve for 'gaus9' shows a greater rate of decay across pseudo-frequencies than that of 'db1'. More energy is concentrated in a narrow band frequencies around 1.00 Hz in the case of 'gaus9'. These observations imply that high energy compaction can be achieved with fewer coefficients if the wavelet and the scales are chosen as stated in section II. A favorable implication of fewer coefficients is fewer number of symbols for analysis and hence an improvement in computational efficiency.

B. Example 2: Noise Suppression

This example demonstrates the noise suppression achieved with wavelets. Let the signal $y(t)$ in Eq. (7) be corrupted with additive zero-mean white Gaussian noise $w(t)$.

$$\tilde{y}(t) = y(t) + \sigma\ w(t) \tag{8}$$

A common measure of noise in a noise-corrupted signal is the signal-to-noise (SNR) ratio that is defined as:

$$SNR_t \triangleq \frac{||y||^2}{||\sigma\ w||^2} \tag{9}$$

The subscript t denotes that y and w are functions of time. Similar to the above definition, SNR in the wavelet scale domain is defined as:

$$SNR_s \triangleq \frac{||\mathbb{L}y||^2}{||\sigma\ \mathbb{L}w||^2} \tag{10}$$

Numerical experiments have been performed with $\sigma \in \{0.05, 0.1\}$. The signal is sampled at 100 Hz (i.e., $\Delta t = 0.01s$). The scales are determined following Eq. (4), such that the pseudo-frequency of the wavelet matches the frequency of the signal. Figure 4 depicts the time domain plot and coefficient plot of the signal y and the noise. The top plate corresponds to noise level $\sigma = 0.05$ while the bottom plate corresponds to $\sigma = 0.10$. Table I lists the values of SNR_t and SNR_s, averaged over 20 simulation runs.

TABLE I
SNR VALUES

	$\sigma = 0.05$	$\sigma = 0.1$
SNR_t	191.55	50.89
SNR_s	25195	4281.5

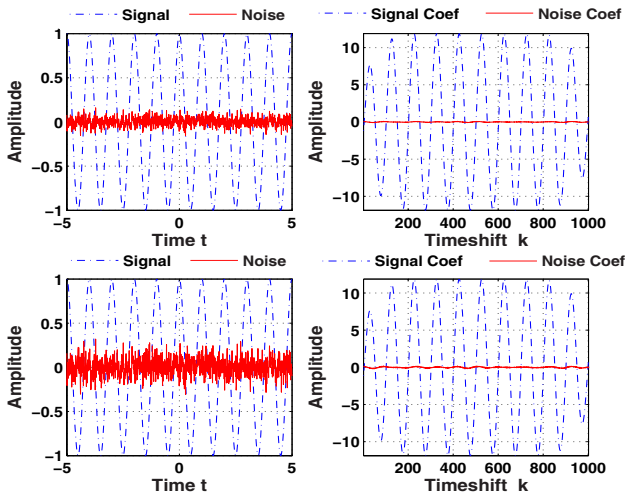

Fig. 4. Signal Denoising

Figure 4 shows that SNR_s is significantly larger than SNR_t. That is, the wavelet-transformed signal is significantly de-noised relative to the time-domain signal. This is expected because the noise is incoherent with the wavelet while the signal enjoys a great degree of coherence with the same. Thus, symbols generated from wavelet coefficients would reflect the characteristics of the signal with more fidelity than those obtained with time domain signals.

C. Example 3: Robustness of Symbol Probability Vector

The proposed symbolization scheme was developed to enhance real-time anomaly detection in complex systems [8]. Of critical importance is the symbol probability vector $\mathbf{p}$ whose elements denote the probability of occurrence of individual symbols in the symbol sequence. The vector $\mathbf{p}$ must be robust relative to measurement noise and, at the same time, be sensitive enough to detect small slowly-varying anomalies from the observed data set. Two distortion measures of a noise-corrupted signal are introduced below.

$$\delta_t \triangleq ||p_t - \tilde{p}_t||_1 \tag{11}$$

The subscript t denotes that the probability vectors correspond to symbols generated from time domain signals; and $||\bullet||_1$ is the sum of the absolute values of the elements of the vector $\bullet$. The vector p_t, with $||p_t||_1 = 1$, corresponds to the uncorrupted signal and $\tilde{p}_t$ corresponds to the corrupted signal. Similar to the above definition, distortion ratio in the wavelet scale domain is defined as

$$\delta_s \triangleq ||p_s - \tilde{p}_s||_1 \tag{12}$$

Therefore, if the distortion ratio is lower, the probabilistic representation of the corrupted signal is closer to that of the uncorrupted signal, i.e., the probability vector $\mathbf{p}$ is more robust to noise.

The partitions are obtained, in case of time domain, by employing the maximum entropy criterion on the time series data of the signal. In the wavelet domain, the partitions are obtained with the wavelet coefficients of the the time series data. In both time domain and wavelet domain, the respective probability vectors p and $\tilde{p}$ are computed with the same partitions for the uncorrupted and corrupted signals.

The cardinality $|\Sigma|$ of the symbol alphabet Σ is chosen to be 4 in this example. The partitions are obtained as mentioned before for the signal y and its transform, i.e., the coefficient vector $\mathbb{L}y$. Table II lists the values of distortion ratios δ_t and δ_s, averaged over 20 simulation runs.

TABLE II
DISTORTION RATIOS

	$\sigma = 0.05$	$\sigma = 0.1$
δ_t	0.040	0.054
δ_s	0.006	0.010

It is seen that distortion ratios are far smaller in the wavelet scale domain than those in the time domain. This observation implies that the symbol probabilities are significantly more robust to measurement noise in the wavelet domain than in the time domain. Hence, it may be inferred that symbols generated from the wavelet coefficients would be better for anomaly detection as the effects of noise to induce errors in the symbol probabilities are significantly mitigated.

D. Anomaly Detection in Nonlinear Electronic Systems

This example demonstrates efficacy of symbolic time series analysis for anomaly detection in nonlinear electronic systems. Experiments have been conducted on a laboratory apparatus [8] that emulates the forced Duffing equation. The details of the experimental apparatus are provided in [7].

$$\frac{d^2y}{dt^2} + \beta\frac{dy}{dt} + y(t) + y^3(t) = A\cos(\Omega t) \qquad (13)$$

The dissipation parameter β varies slowly with respect to the response time t of the dynamical system; $\beta = 0.1$ represents the nominal condition; and a change in the value of β is considered as an anomaly. With amplitude $A = 22.0$ and $\Omega = 5.0$ rad/sec, a sharp change in the behavior was noticed around $\beta = 0.29$, possibly due to bifurcation. The phase plots and time-response plots, depicting this drastic change behavior, are not presented here as they are provided in an earlier publication [8]. The objective of anomaly detection is to identify small changes in the parameter β as early as possible and well before it manifests a drastic change in the system dynamics.

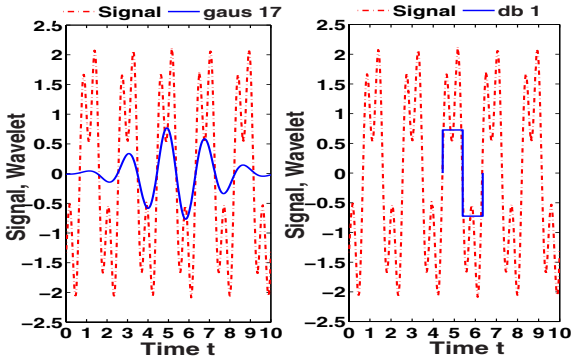

Fig. 5. Profiles of Signal and Wavelet

Choosing the wavelet basis forms the first step in the analysis. The time series data of the signal and a scaled and translated version of the wavelet 'gaus17' are shown in the left hand plate of Figure 5. For the purpose of comparison, the right hand plate of Figure 5 shows the same time series data of the signal and a scaled and translated version of the wavelet 'db1' that was used in [8] for wavelet analysis. Since the wavelet 'gaus17' matches the shape of the signal better than 'db1', in the sense of ℓ_2 distance, 'gaus17' should be more appropriate for symbolic time series analysis than 'db1'. Once the wavelet is chosen, the next step is identification of the frequencies of interest.

While frequencies in the neighborhood of 0.54 Hz are present at the nominal condition ($\beta = 0.10$), they are absent at $\beta \geq 0.29$. Therefore, the wavelet coefficients at scales, corresponding to the pseudo-frequency of 0.54 Hz, would be smaller in magnitude in the anomalous condition(s) when compared with those in the nominal condition. Hence, by choosing scales that correspond to pseudo-frequencies around 0.54 Hz, early detection can be achieved more effectively.

The next aspect of anomaly detection via symbolic time series analysis is the choice of number of symbols, i.e., cardinality $|\Sigma|$ of the symbol alphabet Σ. The scale series data, at the nominal condition, is partitioned into a symbol sequence starting with $|\Sigma| = 2$ and the threshold parameter ϵ_h is chosen to be 0.2. Figure 6 depicts the plot of entropy rate vs number of symbols $|\Sigma|$. It is seen that as $|\Sigma|$ is monotonically increased and reaches the value of 8, the entropy rate $h(\bullet)$ becomes less than ϵ_h. Accordingly, the number of symbols $|\Sigma|$ was chosen to be 8. A smaller value of ϵ_h would result in increased number $|\Sigma|$ of symbols, which will increase computation with no significant gain in accuracy of anomaly detection.

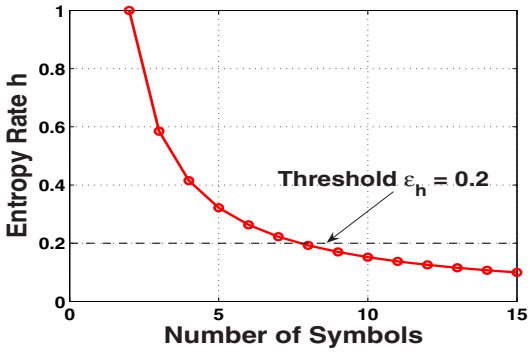

Fig. 6. Selection of Number of Symbols from Entropy Rate

The partitions are obtained using data obtained under the nominal ($\beta = 0.1$) condition. Once the partition is generated, it remains invariant. As the dynamical behavior of the system changes due to variations in β, the statistical characteristics of the symbol sequences are also altered and so are the symbol probabilities. A measure could be induced on the symbol probability vectors obtained under different anomalous conditions, to quantify these changes. Such a measure is called the anomaly measure $\mathbf{M}$. The metric $\mathbf{M_k} = d(\mathbf{p_0}, \mathbf{p_k})$ is an anomaly measure, where $\mathbf{p_0}$ and $\mathbf{p_k}$ represent the symbol probability vectors under nominal and anomalous conditions, respectively. A candidate anomaly measure is the angle between the symbol probability vectors under nominal and anomalous conditions. This measure is defined as:

$$\mathbf{M}_k = \arccos\left(\frac{< \mathbf{p_0}, \mathbf{p_k} >}{||\mathbf{p_0}||_2||\mathbf{p_k}||_2}\right) \qquad (14)$$

where $< x, y >$ is the inner product between the vectors x and y; and $||x||_2$ is the Euclidean norm of x.

Figure 7 depicts the anomaly measures obtained with wavelet 'gaus17' under maximum entropy and uniform partitioning. It also compares the anomaly measures with those obtained with 'db1' wavelet partitioning, which was used in an earlier publication [8], and phase-space partitioning via symbolic false nearest neighbor (SFNN) [5].

The number of symbols $|\Sigma|$ is chosen to be eight in all four cases.With β increasing from 0.1, there is a gradual increase in the anomaly measure much before the abrupt

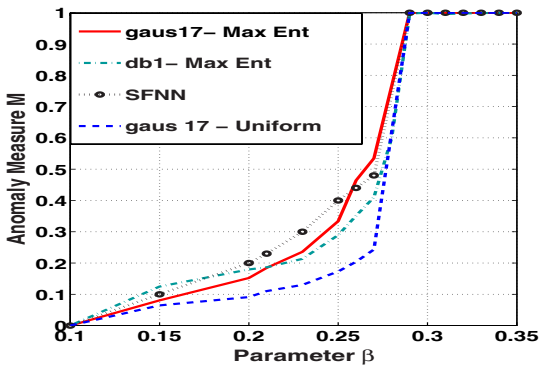

Fig. 7. Anomaly Detection on the Electronic System Apparatus

change in the vicinity of $\beta = 0.29$ takes place. This indicates growth of the anomaly even before any noticeable change in the dynamical behavior takes place. At this point, the measure starts increasing relatively more rapidly suggesting the onset of a forthcoming catastrophic failure. Under maximum entropy partitioning, the larger values of the anomaly measure at smaller values of β and gradual increase in both slope and curvature of the anomaly measure curve would facilitate anomaly detection significantly before it is possible to do so under uniform partitioning. Therefore, with regard to early detection of anomalies, maximum entropy partitioning appears to be more effective than uniform partitioning.

While the SFNN partitioning yields slightly higher values of the anomaly measure and comparable slope (i.e., $dM/d\beta$) in range of $0.2 \leq \beta \leq 0.25$, the wavelet partitioning is relatively more smooth and yields a significant change in the curvature (i.e., $d^2M/d\beta^2$) around $\beta = 0.25$, which is a clear early warning for a forthcoming disruption. Simultaneous consideration of the anomaly measure, slope, and curvature provides a robust method of failure prediction and reduces the probability of false alarms.

Although the detection performance is largely similar, the computation time for SFNN partitioning is observed to be several orders of magnitude larger than those for both cases of wavelet partitioning. For wavelet partitioning, while 'db1' uses 128 scales to generate coefficients, very few (three in this example) scales are needed for 'gaus17'. This means that, a much smaller number of coefficients are necessary for the 'gaus17' wavelet partitioning, all of which are significant, and hence much fewer symbols. Thus, choosing a small number of symbols with an appropriate wavelet leads to a substantial increase in computational efficiency without any noticeable degradation in the performance of anomaly detection.

IV. Summary and Conclusions

This paper presents a novel method of symbol sequence generation from time series data for anomaly detection in complex systems. In this approach, the wavelet transform coefficients of the time-domain signal are utilized for symbol generation instead of the time series data. Various aspects of this method, such as selection of the wavelet and scales, are systematically investigated. The advantages of using wavelet coefficient-space partitioning, instead of phase-space partitioning, are demonstrated with simulation and experimental data. It has been shown that use of wavelet coefficients suppresses the measurement noise and results in smaller symbol distortion ratios. That is, symbol sequences, generated from the wavelet coefficients of a noisy signal, capture the signal information better than those obtained directly from the time series data of the signal. It is also shown that the choice of an appropriate wavelet and scales significantly improves computational efficiency and thereby enhances implementation of the anomaly detection technique for real-time applications. An algorithm, based on entropy rate, is introduced for selection of the symbol alphabet size, i.e., the the number of symbols.

A partitioning method, based on maximum entropy, is presented and compared with the previously used method of uniform partitioning [8]. It is demonstrated that maximum entropy partitioning yields better detection performance than uniform partitioning. Wavelet-based maximum entropy partitioning has been compared with symbolic false nearest neighbor (SFNN) partitioning [5] with regard to anomaly detection. It is observed that the afore-mentioned partitioning methods yield comparable results while the computation time for SFNN partitioning is observed to be several orders of magnitude larger than wavelet-based maximum entropy partitioning.

A major conclusion based on this investigation is that maximum entropy partitioning, combined with an appropriate choice of wavelet and scales, significantly enhances computational efficiency and anomaly detection capabilities beyond what has been reported in literature [8] [2]. The field of symbolic time series analysis is relatively new and its application to anomaly detection is very recent. Therefore, the proposed method of symbol generation for anomaly detection requires further theoretical and experimental research in laboratory environment as suggested in an earlier publication [8].

References

[1] P. Abry, *Ondelettes et turbulence. multirsolutions, algorithmes de dcomposition, invariance d'chelles*, Diderot Editeur, Paris, 1997.

[2] S.C. Chin, A. Ray, and V.Rajagopalan, *Symbolic time series analysis for anomaly detection: A comparative evaluation*, Signal Processing **85** (2005), no. 9, 1859–1868.

[3] C.S. Daw, C.E.A. Finney, and E.R. Tracy, *A review of symbolic analysis of experimental data*, Review of Scientific Instruments **74** (2003), no. 2, 915–930.

[4] H. Kantz and T. Schreiber, *Nonlinear time series analysis, 2nd ed.*, Cambridge University Press, 2004.

[5] M.B. Kennel and M. Buhl, *Estimating good discrete partitions form observed data: Symbolic false nearest neighbors*, Physical Review E **91** (2003), no. 8, 084102.

[6] S. Mallat, *A wavelet tour of signal processing 2/e*, Academic Press, 1998.

[7] V. Rajagopalan, R.Samsi, A.Ray, J.Mayer, and C.Lagoa, *A symbolic dynamics approach for early detection of slowly evolving faults in nonlinear systems*, IASTED-CSS, Clearwater, FL, Nov-Dec 2004.

[8] A. Ray, *Symbolic dynamic analysis of complex systems for anomaly detection*, vol. 84, 2004.

[9] A. Teolis, *Computational signal processing with wavelets*, Birkhäuser, 1998.