Procreating V-detectors for Nonself Recognition: An Application to Anomaly Detection in Power Systems

Min Gui
Electrical & Computer Engg Dept.
Kansas State University
Manhattan, KS 66502
Tel: (785) 532-4642
Email: mingui@ksu.edu

Sanjoy Das
Electrical & Computer Engg Dept.
Kansas State University
Manhattan, KS 66502
Tel: (785) 532-4642
Email: sdas@ksu.edu

Anil Pahwa
Electrical & Computer Engg Dept.
Kansas State University
Manhattan, KS 66502
Tel: (785) 532-4654
Email: pahwa@ksu.edu

ABSTRACT

The artificial immune system approach for self-nonself discrimination and its application to anomaly detection problems in engineering is showing great promise. A seminal contribution in this area is the V-detectors algorithm that can very effectively cover the nonself region of the feature space with a set of detectors. The detector set can be used to detect anomalous inputs. In this paper, a multistage approach to create an effective set of V-detectors is considered. The first stage of the algorithm generates an initial set of V-detectors. In subsequent stage, new detectors are grown from existing ones, by means of a mechanism called procreation. Procreating detectors can more effectively fill hard-to-reach interstices in the nonself region, resulting in better coverage. The effectiveness of the algorithm is first illustrated by applying it to a well-known fractal, the Koch curve. The algorithm is then applied to the problem of detecting anomalous behavior in power distribution systems, and can be of much use for maintenance-related decision-making in electrical utility companies.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods and Search – graph and tree search strategies, heuristic methods.

General Terms

Algorithms, theory.

Keywords

Anomaly detection, immunocomputing, artificial immune systems.

1. INTRODUCTION

One of the fundamental characteristics of the mammalian immune system is the ability to recognize the presence of foreign bodies called antigens. This process is carried out by means of a class of T-cell antibodies called detectors that possess the ability to distinguish invading antigens from the body's own cells. This phenomenon is termed as self-nonself recognition. Self-nonself recognition proceeds within a specimen's immune system by generating a large variety of detector cells, systematically culling out those that erroneously categorize native cells (self) as foreign (nonself) while retaining the rest. This principle is called negative selection.

Nature-inspired algorithms based on the mammalian immune system model for self-nonself discrimination are becoming increasingly popular in engineering applications [1, 2, 3]. In particular, negative selection algorithms have been successfully applied to many anomaly detection problems [4]. Here, the anomalous region of the input vector space is considered to be equivalent to nonself. Recently, Ji and Dasgupta have also shown that negative selection algorithms perform as well as the more conventional support vector machine based classification algorithm [5]. An important engineering application of negative selection is suggested in [3] where it is applied to detect faults in squirrel cage induction motors. Computer security is another important application of negative selection [1, 6].

An approach that makes use of negative selection to create detectors is the V-detectors (variable sized detectors) approach [7]. V-detectors provide a simple, yet efficient way to yield a set of detectors that extensively covers the nonself region of the feature space. They have been used in a self-nonself discrimination application involving ball-bearing data as well as another one involving Indian Telugu characters [5]. In [8], it has been shown that V-detectors alone are adequate for most anomaly detection problems.

In this paper, we examine a multistage approach for creating V-detectors. Our algorithm generates an initial set of detectors in a manner similar to [7], but also incorporates one or more phases where offspring detectors can be created from existing ones to provide better nonself coverage. The method is especially applicable when the boundary separating self from nonself follows complicated patterns.

We demonstrate the effectiveness of our algorithm for a synthetic but difficult problem, where the boundary between self and nonself is the well-known Koch curve. The Koch curve is a fractal with a Hausdorff dimensionality of 1.26 [9]. In spite of its deceptively simple shape, the Koch curve has infinite length.

We then apply our method to a real world problem of anomaly detection in power distribution systems. Failures in distribution systems often occur due to a variety of reasons, such as animal

GECCO'07, July 7–11, 2007, London, England, United Kingdom.

activity, bad weather conditions, or extraneous factors including poor maintenance and aging equipment. An automated approach to detect anomalies in distribution systems can be of immense commercial value to electrical utility companies. Abnormally high failures in a location would highlight the need for equipment repair, replacement or other maintenance action. On the other hand, if the number of failures in one area is significantly less than expected, the limited resources of the company can be redirected elsewhere.

2. PROCREATING V-DETECTORS

2.1 The V-detectors algorithm

V-detectors provide a very simple, yet effective framework to generate a set of detectors that can cover the nonself region in an extensive manner, using negative selection. Each detector is associated with its own location in the nonself region, as well as its own detector radius. A detector detects an input if the distance between its own location and the input is less than the radius. Such an input is classified as nonself. When no detector can detect an input, the input is classified as a self point.

The V-detectors algorithm proceeds as follows. The set of detectors D is initialized to a null set. In each iteration, a new detector is generated and assigned a location xi. It is then tested to see if it overlaps the self region S. If it does not do so, the detector is added to D, else it is discarded. Each detector is assigned its own detector radius, which is the distance from the detector's location to S. The algorithm is repeated until the termination condition is satisfied. In the simple implementation of the V-detectors algorithm considered here, the termination condition is satisfied when D reaches a predetermined maximum limit.

2.2 The single phase procreating V-detectors method

The V-detectors algorithm has proved to be very effective in covering the nonself region. It has been applied to datasets both real-world as well as synthetic [5, 7, 8]. Synthetic data is particularly useful to visualize the coverage in two dimensions. Unfortunately, in the existing literature, only simple shapes (such as squares or rings) have been used as a surrogate for the self. While randomly generated detectors may be sufficient to cover such simple shapes, we have observed that that may not be the case when the boundaries between the self and nonself regions acquire more complex patterns. Under these circumstances, we propose a procreating V-detectors algorithm that has two distinct phases. The first stage is called the generative stage. As earlier [5, 7, 8], during the generative stage of the proposed method, random V-detectors are generated and tested to see if they abut upon the self region, being added to the detector set D only if they do not do so. In the second stage, called the procreative stage, new detectors are generated by a mechanism of procreation from the older ones in D. The procreative stage, which is the main contribution of this paper, is described below. At the beginning of the procreative stage, the algorithm has a non-empty set of detectors D. In each iteration, it selects a detector whose location and detection radius are xi and ri from D, and creates new offspring detectors, each located at a distance ri from xi. If an original detector in D is regarded as a sphere of radius ri in the nonself region centered around xi, the offspring detectors will be located along the sphere's circumference, i.e. at a location xi + ûri where û is some unit vector. Its radius is set to be equal to the minimum of the distance from its center to the self region S. An offspring detector that found to cover a region that is entirely covered by another existing detector in D is said to be subsumed by D. Since a new detector has additional value as long as it is not subsumed by the latter, it is allowed to enter a set of offspring detectors Do only if it is not subsumed. The algorithm is stopped when all the original detectors in D have been allowed to grow offspring.

The outline of the procedure for the generative stage is as follows.

```
D = generate(S,N)
begin
i = 1
for j = 1 to N
       xi = rand()
       if overlap(xi,S) == F
              ri = dist(xi,S)
              D = D ∪ {(xi,ri)}
              i = i + 1
       endif
endfor
end
```

The procedure for the procreative stage is provided below. The unit vectors **û** that are used to compute the locations of the offspring detectors.

```
D◦ = procreate(D,S)
begin
j = 1
D◦ = {}
for each detector (xi,ri) in D do
       for each unit vector û do
              xj = xi + ûri
              a = dist(xj,S)
              rj = min(a)
              if subsume(xj,D) == F
                     D◦ = D◦ ∪ {(xj,rj)}
                     j = j + 1
              endif
       endfor
endfor
end
```

2.3 The multiphase procreating V-detectors method

During both the generation as well as procreative stages, it is a good idea to introduce a non-negative parameter called the *threshold*, θ. A detector is considered to be non-overlapping with the self region **S** when its circumference is at least at a distance of θ away from either a given set of self samples or the entire self region. The threshold can be set to zero to maximize the coverage of the nonself region, resulting in more frequent classification of inputs as anomalies later on. On the other hand, a larger θ results in more circumspect classification. Later in this section we have considered varying degrees of detection that can be accomplished easily by varying the threshold. When computing the radius of a detector, the threshold is also subtracted from the actual Euclidean distance of its location from **S**.

It is also possible to have more than a single procreative stage, which we define as multiphase procreation. In such a case, the detectors produced by the immediate previous phase of procreation can be used to grow a new set of detectors. As seen in Section 3, such an approach is found to be especially helpful to cover arbitrarily small interstices of the self-nonself boundary that would normally be out of reach of the previous detectors. Since it is highly improbable for a randomly generated detector to land in a small interstice, relying on them entirely for coverage would require a very large number of detectors to be created and tested. On the other hand, these interstices can be filled rapidly by procreating offspring.

The outline of the procedure for the generative stage of multiphase procreating V-detectors method is as follows.

```
D = M-generate(S,N,θ)
begin
i = 1
for j = 1 to N
       x_i = rand()
       if overlap(x_i,S,θ) == F
              r_i = dist(x_i,S,θ)
              D = D ∪ {(x_i,r_i)}
              i = i + 1
       endif
endfor
end
```

The procedure for the procreative stage of multiphase procreating V-detectors method is provided below. The unit vectors **û** that are used to compute the locations of the offspring detectors.

```
D_o = M-procreate(D,S,θ)
begin
j = 1
D_o = {}
for each detector (x_i,r_i) in D do
       for each unit vector û do
              x_j = x_i + û r_i
              a = dist(x_j,S,θ)
              r_j = min(a)
              if subsume(x_j,D) == F
                     D_o = D_o ∪ {(x_j,r_j)}
                     j = j + 1
              endif
       endfor
endfor
end
```

Maintaining a threshold θ that is initially high during the generative stage, and lowering it towards zero in a stepwise manner during subsequent procreative stage is found to produce better coverage. This is because each reduction in θ leaves open an extra 'gap' surrounding the self-nonself boundary that must be covered by new detectors by the next phase of procreation. This gap can readily be filled by subsequent offspring detectors. The 'gap' causes larger detector radii to be assigned to the offspring detectors by the algorithm. The net effect of this scheme is bigger offspring detectors which indirectly results in additional coverage of the nonself region. On the other hand, keeping a constant zero threshold throughout the algorithm leaves very little space in the nonself region at the end of each phase, yielding smaller offspring detectors. In order to completely cover the entire nonself region, the threshold must be assigned a zero value at least during the last procreative phase before terminating the algorithm.

The complete training algorithm of multiphase procreating V-detectors method is outlined below.

```
algorithm M-proliferation
begin
determine θ_0 thru θ_{M-1} (θ_0≥θ_1≥…≥θ_{M-1}≥0)
θ_M = 0
θ = θ_0
D_0 = M-generate(S,N,θ)
D = D_0
for phase = 1 to M do
       θ = θ_phase
       D_phase = M-procreate(D_{phase-1},S,θ)
       D = D ∪ D_phase
end
```

2.4 Graded anomaly detection

The use of multiple phases of proliferation with progressing decreasing values of the threshold θ also works to our advantage in some applications. Instead of a binary classification of any input **z** as belonging to the self region **S** or its complementary nonself region (i.e. anomalous), whenever finer gradations of classification is desired, varying the threshold can easily produce different levels of detectors. An input feature that is categorized as an anomaly by the lower levels of detectors is more likely to be anomalous than that which is detected by a higher leveled detector, as shown in figure 1.

A similar fuzzy characterization of self-nonself regions has been suggested for a rule-based system of anomaly detection elsewhere [2]. The present approach may also be used here as well. Noting that each procreating phase produces detectors that are increasingly closer to the real boundary between the self and nonself regions, an input **z** that is detected by a detector from the generative stage $\mathbf{D}_0$, (which are associated with the largest threshold), should be considered most anomalous. Conversely, an input that is detected by a detector from the last phase $\mathbf{D}_M$ but not by any previous one, is one that is quite close to the self-nonself boundary and should therefore be considered as only minimally anomalous.

The following procedure accepts the input **z**, and returns an alarm level. A zero alarm implies that **z** lies within **S**. More anomalous inputs **z** cause higher values of the variable alarm to be returned by the procedure. Detection by any set of detectors in this procedure is accomplished simply by computing the Euclidean distance of **z** to each detector's location $\mathbf{x}_i$, and returning a '*T*' (true) output if and only if that distance is less than a detector radius r_i.

```
alarm = detect_anomaly(z,D_0,D_1,…,D_M)
begin
alarm = 0
determine L_0 thru L_M (L_0≥L_1≥…≥L_M>0)
for level = 0 to M do
       if detect(z, D_level)
              alarm = L_level
              break
       endif
endfor
end
```

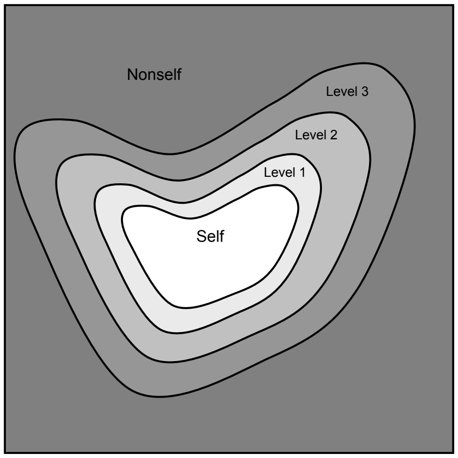

Figure 1. Degree of anomaly with different levels.

3. RESULTS WITH SYNTHETIC DATA

3.1 Data generation

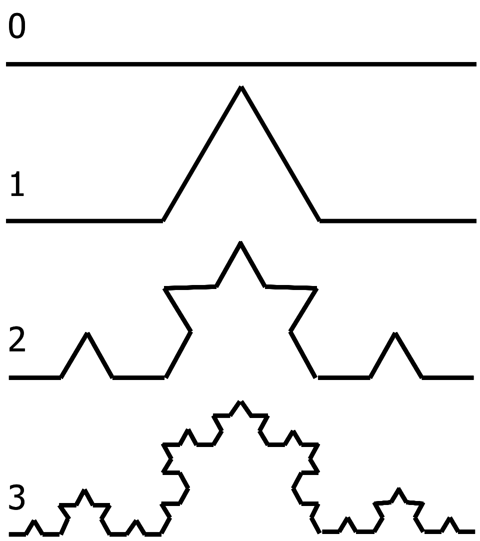

Figure 2. Generating the Koch curve.

The Koch curve is a well known fractal, a shape that possesses a non-integral Hausdorff dimensionality. Although the Koch curve is a curve that can be straightened to become a straight line, its dimensionality is in fact 1.26 and it has an infinite length. As this paper addresses nonself coverage for complex boundaries, these properties of the Koch curve make it an ideal model for the self region. The intricacies of the Koch curve are apparent even upon simple visual inspection, for it bears a strong resemblance to one of the six faces of a snowflake.

The Koch curve can be generated recursively quite easily in the following manner. Starting with a straight line segment, in each recursion every straight line segment is divided into three equal sub-segments. The middle sub-segment is then replaced with two of equal length that are rotated by angles of ±60° from the original orientation. figure 2 clearly shows how the Koch curve is generated. It must be noted that although only three recursive steps are shown in the figure, in order to generate a fractal, infinite recursions are needed. In our analysis, we have generated the self region in a manner, although rotated by 45°. The recursive process has however been limited to only a depth of six. Adding more stages of recursion increased the computations involved without causing any discernible difference in the results obtained.

3.2 Results

The search space in this set of simulations is a square shaped region, $[0, 1]^2$. The Koch curve separates the self from the nonself. The bottom left region in Figures 3-8 is defined as the self space. In figure 3 and 4, the set of detectors generated the original V-detector algorithm and the proposed single phase procreating V-detectors approach are shown. figure 5, 6, 7 and 8 show the generated detector sets using the multiphase method with changing thresholds. In all cases, a total of 5,500 detectors were generated, which were then either accepted or rejected. The number of accepted detectors are showed in figure 9 for V-detectors algorithm, single phase procreating V-detectors algorithm and multiphase procreating V-detectors algorithm. From the figure we can see the latter two procreating methods have more accepted detectors for the same number of generated detectors. In figures 4-8, the procreative stage was started after 1,000 initial detectors were tested during the generative stage. The figures clearly indicate that using the two procreating V-detectors methods it is possible to generate detectors that are more likely to be accepted than in the original V-detector algorithm. The figure also shows that multiphase procreating V-detectors method has more accepted detectors then single phase procreating V-detectors method when the threshold is close to zero. The underlying reason is that with a nonzero threshold, the multiphase procreating V-detectors method has less nonself space to cover.

After the detector set is generated, testing data points are randomly distributed over the entire space to test the performance. The multiphase procreating V-detectors method has better performance over the other two methods as it can detect anomalies most frequently, as shown in figure 10.

4. DISTRIBUTION SYSTEM FAILURES

4.1 Characteristics of weekly animal-related failures

Animal-related failures from the year 1998 to 2005 in the distribution system of Manhattan, Kansas were aggregated on a monthly basis in each year to find the pattern of outages due to animals. We observed that type of month has a considerable impact on the animal-caused failures on overhead distribution systems because of the difference in the behavioral patterns of animals in different months [10]. Hence, we have used the month as a causal factor. Further examination of animal-related failure showed that almost all the faults due to animals took place in fair weather [10]. Fair weather days have temperature within 40^{o} F and 85^{o} F. This is consistent with the fact that animals are most active in fair weather. Hence, the number of fair days per week is taken as another causal factor for weekly failure. Therefore, the feature space of weekly animal-related failures in distribution system has three dimensions, which are weekly animal-related failures, month type and the number of fair days per week.

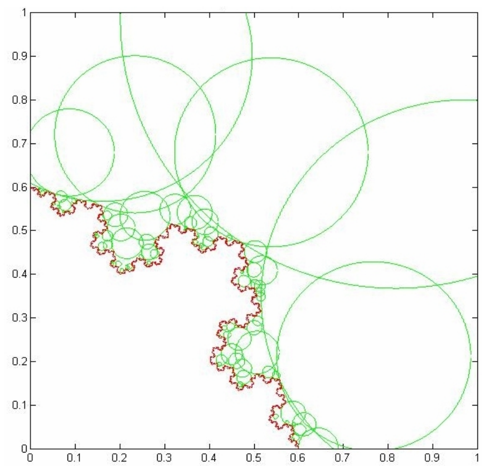

Figure 3. V-detectors method after 5500 iterations.

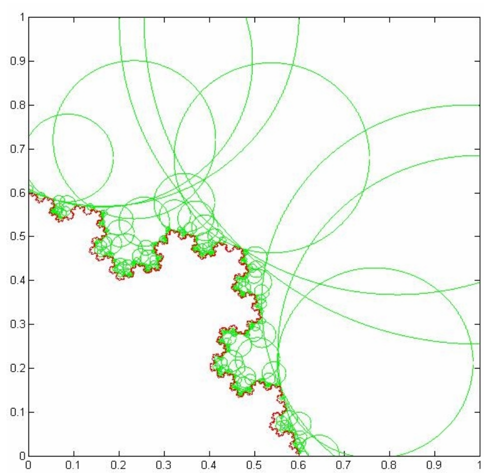

Figure 4. Single phase procreating V-detectors method after 5500 iterations.

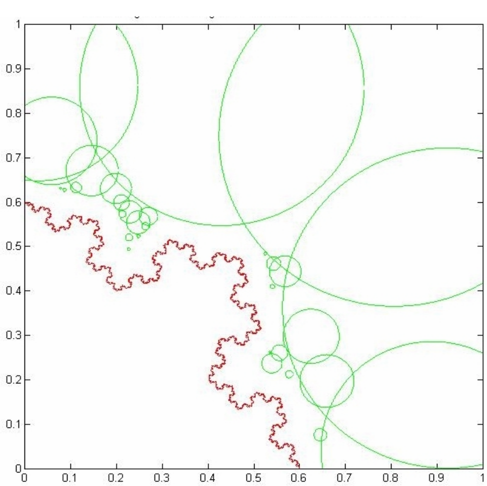

Figure 5. Multiphase procreating V-detectors method with threshold=0.5 after 1000 iterations.

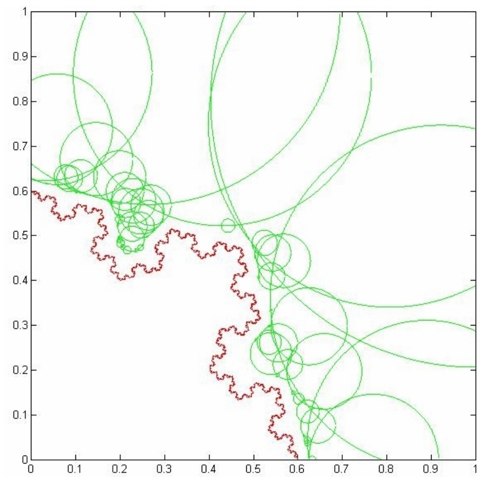

Figure 6. Multiphase procreating V-detectors method with threshold=0.25 after 2500 iterations.

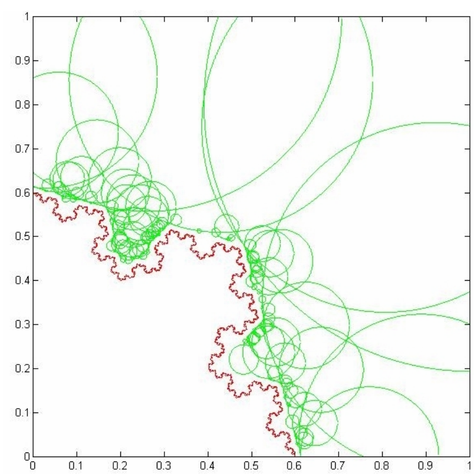

Figure 7. Multiphase procreating V-detectors method with threshold=0.125 after 4000 iterations.

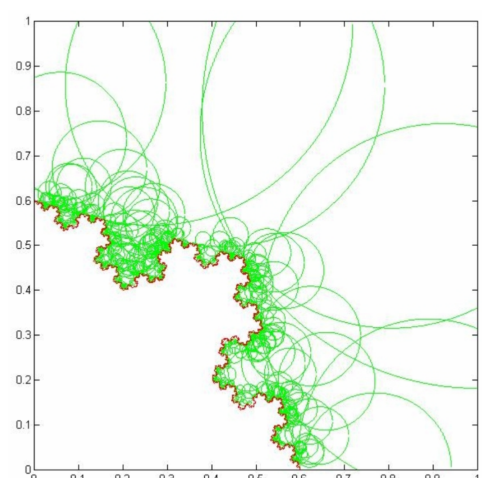

Figure 8. Multiphase procreating V-detectors method with threshold=0.0 after 5500 iterations.

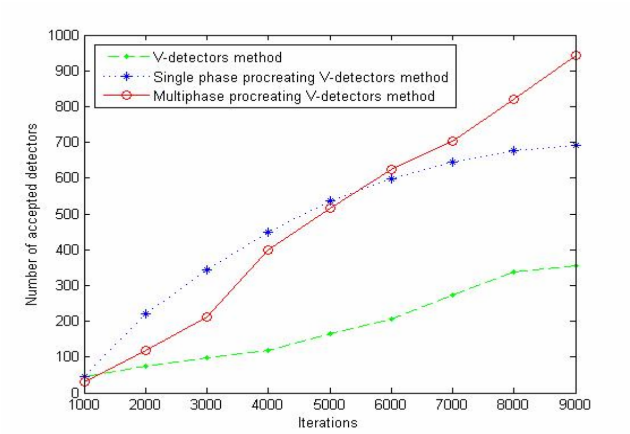

Figure 9. Number of accepted detectors by each method.

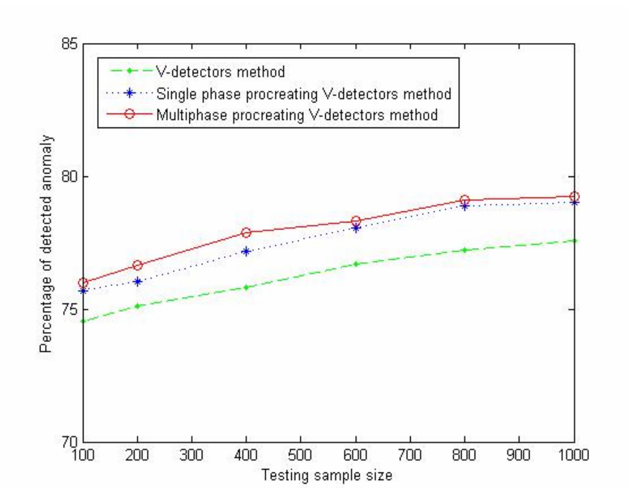

Figure 10. Percentage of anomaly detected by each method.

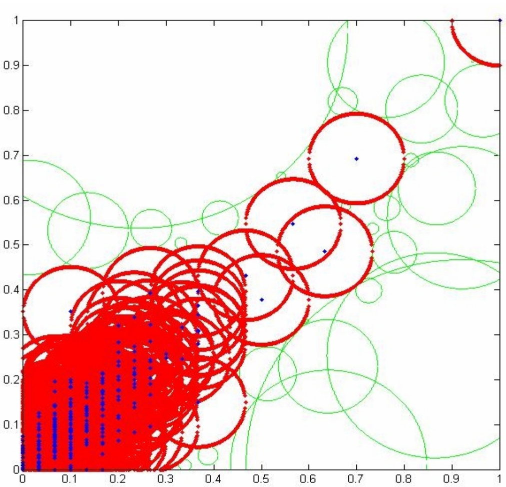

Figure 11. 2-D view of self region and detectors for animal-related failures.

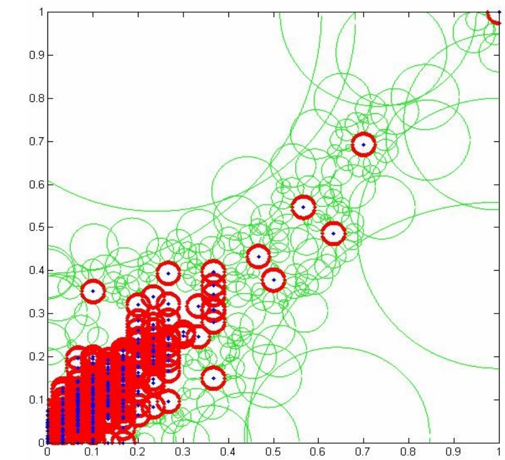

Figure 12. 2-D view of the last phase of multiphase procreating V-detectors method for animal-related failures.

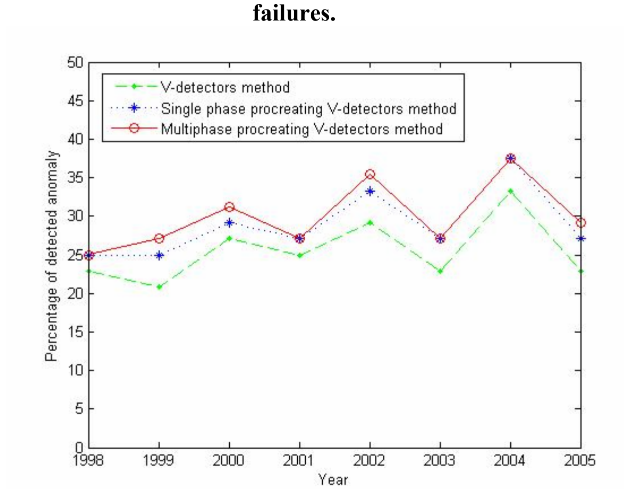

Figure 13. Percentage of anomaly detected by each method in animal-related failures.

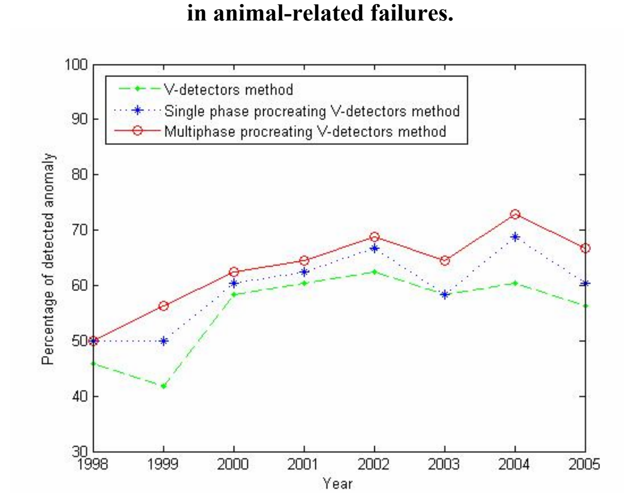

Figure 14. Percentage of anomaly detected with different thresholds by multiphase procreating V-detectors method in animal-related failures.

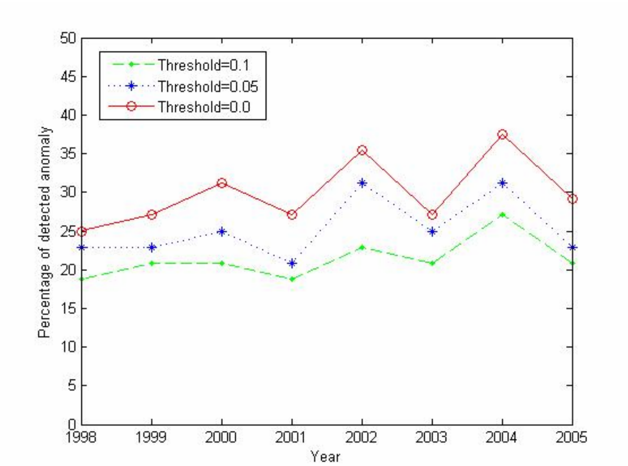

Figure 15. Percentage of anomaly detected with 30% noise added into the original animal-related failures by each method.

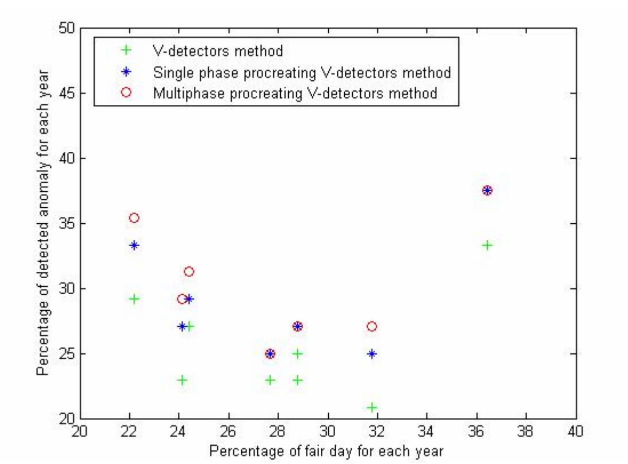

Figure 16. Relation between weather and anomaly detected in animal-related failures bv each method.

4.2 Anomaly detection for weekly animal-related failures data

The approach taken for the problem of anomaly detection in weekly animal-related failures data is similar to that taken earlier for the Koch curve. The entire searching space is normalized to a 3-D cube $[0, 1]^3$. The difference is that for three dimensions data the detectors become 3-D spheres and procreation takes place along the entire surface of the sphere. Another difference is the definition of self space. For synthetic data, as we mentioned in section 3, the region inside the Koch curve is defined as self space and the region outside the Koch curve is defined as nonself space. However, in case of distribution system data, the self space is determined indirectly through a sample set of self points. Under these circumstances, the self region is defined as the union of spheres centered around these sample points. Figures 11 and 12 show projections of the self and nonself regions in two dimensions. In figure 12, we show the view of 3-D procreating method with a smaller self radius than figure 11. In figure 11 and 12, the bold circles with a point in the center are defined as the self region whose center is a self sample. The thin lined circles are the detectors.

4.3 Results

We applied the V-detectors method, the single phase procreating V-detectors method and the multiphase procreating V-detectors method to weekly animal-related failures from the year 1998 to 2005 in the distribution system of Manhattan, Kansas. From the eight years data, we perform eight simulations, taking seven years' weekly data as self sample for training, leaving out one year's data for testing, each time. The multiphase procreating V-detectors method with a threshold of 0.0 has the best performance to detect anomaly and the V-detectors method has the worst performance, as shown in figure 13. With self radius equals 0.02 and different thresholds we have obtained the percentage of detected anomaly with multiphase procreating V-detector technique, as shown in figure 14. More anomaly is detected with smaller threshold. As mentioned in section 3, we can allow graded anomaly detection and define the degree of anomaly corresponding to different thresholds. With 30% noise added into original failures data used as testing sample, the multiphase procreating technique again has the best performance, as it can detect anomalies most frequently, whereas the V-detectors technique has the worst performance, as shown in figure 15.

Lastly, we examine if V-detectors are able to discover any relationship between weather conditions and detected anomalies. As before, we perform eight simulations, leaving out one year of data from the original data, each time. In figure 16, the percentage of anomalies detected for each year is shown as a function of the percentage of fair days that occurred in that same year. As seen from the figure, the multiphase procreating V-detectors method is more likely to classify those years with unusually high as well as unusually low percentages of fair days as anomalous. This is consistent with the fact that animal activity is directly related to the number of fair days, with higher animal activity resulting in higher failure rates and vice versa. The single phase procreating V-detectors method follows a similar pattern, although it is marginally less likely to classify points with low fair day data as anomalous. In contrast, visual inspection clearly shows that the V-detectors method is not able to distinguish between anomalous weather patterns and non-anomalous ones that easily. This study again shows that the proposed methods perform better than the original V-detectors algorithm for anomaly detection.

5. CONCLUSION

For comparison with the V-detectors method, we have proposed single phase procreating V-detectors method and multiphase procreating V-detectors method and applied them to synthetic data and weekly animal-related failure data in distribution system. The results show that the single phase procreating V-detectors method and multiphase procreating V-detectors method have better performance for both data sets than the V-detectors method. The anomaly detection of power system failures can be used for maintenance-related decision-making in electrical utility companies. For the future work, the algorithm can be applied to overall failure data in power system and in other area such as network intrusion detection.

6. ACKNOWLEDGMENTS
This work has been supported by the NSF under grant number NSF- NSF ECS 0501288 to Anil Pahwa and Sanjoy Das.

7. REFERENCES
[1] Harmer, P.K., Williams, P.D., Gunsch, G.H., and Lamont, G.B. An artificial immune system architecture for computer security applications. *IEEE Trans. Evolutionary Computation, 6, 3*(June 2002), 252-280.

[2] Branco, P.J.C., Dente, J.A., and Mendes, R.V. Using immunology principles for fault detection. *IEEE Trans. Industrial Electronics,50,2*(April 2003), 362- 373.

[3] Ian Nunn, and Tony White. The application of antigenic search techniques to time series forecasting. In *Proceedings of the 2005 conference on Genetic and evolutionary computation(GECCO'05).* Washington DC, 2005, 353-360.

[4] D. Dasgupta, S. Yu and N. Majumdar. MILA -- Multi-Level Immune Learning Algorithm and its application to Anomaly Detection. *The Soft Computing Journal, 9, 3*(March 2005), 172-184.

[5] Zhou Ji and Dipankar Dasgupta. Applicability Issues of the Real-Valued Negative Selection Algorithms. In *Proceedings of the 2006 conference on Genetic and evolutionary computation(GECCO'06)*. Seattle, Washington, July 8-12, 2006, 111-118.

[6] D. Dasgupta and F. Gonzalez. An Immunity-Based Technique to Characterize Intrusions in Computer Networks. *IEEE Trans. Evolutionary Computation, 6, 3*(June 2002), 281-291.

[7] Zhou Ji and Dipankar Dasgupta. Real-Valued Negative Selection Algorithm using Variable-Sized Detectors. In *Proceedings of the 2004 conference on Genetic and Evolutionary computation(GECCO'04).* Seattle, Washington, June 26-30, 2004, 287-298.

[8] Zhou Ji. A Boundary-Aware Negative Selection Algorithm. In *Proceedings of the 9th IASTED International conference on Artificial intelligence and soft computing (ASC' 05).* Benidorm, Spain, 2005.

[9] Zhi Wei Zhu , Zuo Ling Zhou and Bao Guo Jia. On the Lower Bound of the Hausdorff Measure of the Koch Curve. *Journal Acta Mathematica Sinica, 19, 4*(October, 2003), 715-728.

[10] S. Sahai and A. Pahwa. Failures of Overhead Distribution System Lines Due to Animals. In *Proceedings of the36th North American power symposium(NAPS'04).* University of Idaho, Moscow, Idaho, USA, Aug. 2004.