Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)

Model Selection for Anomaly Detection in Wireless Ad Hoc Networks

Hongmei Deng, Roger Xu
Intelligent Automation Inc., Rockville, MD 20855
{hdeng, hgxu}@i-a-i.com

***Abstract*-Anomaly detection has been actively investigated to enhance the security of wireless ad hoc networks. However, it also presents a difficulty on model determination, such as feature selection and algorithm parameter optimization. In this paper, we address the issue of parameter selection for one-class Support Vector Machine (1-SVM) based anomaly detection. We have investigated the performance of existing approaches, and also proposed a skewness-based outlier generation approach for parameter selection in the 1-SVM based anomaly detection model.**

I. INTRODUCTION

Significant difference between a wireless ad hoc network and its wired and wireless counterparts makes it vulnerable to various types of attacks [1]-[4]. Correspondingly, intrusion detection, as an efficient monitoring and detecting method, has been actively investigated. Anomaly detection (also known as unsupervised intrusion detection) [4][5] has shown several preferred properties compared with traditional intrusion detection approaches. Without requiring a clearly labeled dataset in the training stage, it avoids the time-consuming labeling process. It usually has less complexity, and is capable of handling large amounts of audit information with the growing network size.

However, the anomaly detection approach also presents a difficulty on model determination, such as feature selection and algorithm parameter optimization. Unlike the conventional supervised intrusion approaches, feature index can be computed from the labeled dataset, and decision boundary is supported from both sides (target class and outlier class). In the case of anomaly detection, only one class of data is available. It is therefore difficult to decide, on the basis of just one class, how strictly the boundary should fit around the data in each of the feature directions.

The purpose of this research is to investigate the issue of parameter selection for anomaly detection. In particular, we mainly consider the one-class Support Vector Machine (1-SVM) based detection model, which has shown its effectiveness through several researches [5][11][12]. By optimizing the algorithm parameters, we are able to build a more efficient anomaly detection model to detect various types of attacks with high accuracy. We have investigated the performance of existing approaches, and also proposed a skewness-based outlier generation approach for parameter selection in the anomaly detection model.

The organization of the paper is as follows. In Section II, we present some background information. The proposed skewness-based parameter selection approach is described in Second III. Performance evaluation is given in Section IV, and concluding remarks are added in Section V.

II. BACKGROUND

A. Anomaly detection

Taking a data-centric point of view, the anomaly detection can be considered as an unsupervised classification or outlier detection problem, in which the decision boundary is learned only from normal network records. Using only one type of data, a decision boundary is formed around the target class to capture the behaviors of normal network operation excluding all the attacks. By considering the anomaly detection process as an unsupervised classification problem, several unsupervised learning algorithms can be used, including artificial neural networks, data clustering, k-nearest neighbor, etc. Recently, one-class Support Vector Machines (1-SVMs) [7] has been proposed for anomaly detection, and proved its effectiveness [5][12]. 1-SVM is a direct derivative of SVM algorithm and inherits all the advanced properties of SVM algorithms. As a result, it usually achieves a better performance than most of the current approaches.

In literature, there are two similar 1-SVM algorithms available. One is called ν-SVC developed by Schölkopf [7], and the other is termed as Support Vector Data Description (SVDD) [9]. In [10], Tax showed that the SVDD gives identical solutions with the ν-SVC when the data is preprocessed to have unit norm. In the case of a Gaussian kernel, the data is implicitly rescaled to norm 1. Therefore, the solutions of the SVDD and the ν-SVC are identical when the Gaussian kernel width is equal and $C=1/\nu N$ is used. In the SVDD, the parameter C is set at a pre-specified value indicating the fraction of objects which should be rejected. In the ν-SVC, the ν directly gives the fraction of objects which is rejected. Both of them are exchangeable.

In this research, we only consider parameter selection for ν-SVC based anomaly detection model. Thus, two parameters are to be optimized, namely, the Gaussian kernel width parameter γ and the fraction of rejected objects ν. The parameter ν decides the fraction of data points in the region, and the kernel width parameter γ decides the "shape" of the region. Both of them influence the generalization performance of the anomaly detection algorithm, and a good choice of the

1-4244-0705-2/07/$20.00 ©2007 IEEE

two parameters is necessary. However, as we noted the problem is not trivial due to the basic characteristics of unsupervised learning.

B. Related work on model selection for 1-SVM

There are several approaches have been proposed to deal with this problem. One simple approach is to fix ν priori to the highest allowable fraction of misclassification of the target class. For example, only a 5% classification error is allowed on the training set, i.e. $\nu=0.05$. Then it addressed the problem by only tuning the kernel parameter γ. In [15], N. Cristianini et al. suggested several possible ways for tuning γ. One way is to minimize the number of support vectors. Another way is to maximize the margin ($\rho/\|\omega\|$) of separation from the origin in *ν*-SVC, which is equivalent to minimizing the radius of the smallest sphere enclosing the data [9]. R. Unnthorsson et al. [14] also discovered that the best classification accuracy occurred at the point where the classification curve on the training data first reached 1-ν, based on a fixed setting ν. Thus, they proposed a simple heuristic criterion for selecting γ to be: start with smaller value of σ ($\sigma=\sqrt{2\gamma}$) and increase it until the error on the training set first reaches ν. In section IV we will show that this criterion may help for picking a value of γ, but fixing the value of ν is not efficient as it may choose a large number of values.

Q.-A. Tran et al. [13] proposed another method to evaluate the generalization performance of 1-SVM by combining the size of region and the generalization fraction of data points in the region, respectively. Smaller size and greater fraction all indicate better performance. The size of region is then estimated using the fraction of support vectors, as the region with more support vectors usually has smaller size. Tran et al. also proposed a ξαρ-estimate method to estimate the generalization fraction of data points in the region. When the two evaluation measure estimates are available, the generalization performance of 1-SVM can be rewritten as follows:

$$T=\lambda R_{\xi\alpha\rho}+(1-\lambda)F_{nSV}$$

where λ controls the tradeoff between these two evaluation measures. The $R_{\xi\alpha\rho}$ is the ξαρ-estimate and F_{nSV} is defined as the ratio of number of support vectors to the number of training objects. For the definition of $R_{\xi\alpha\rho}$ please refer to [13] for detail. To obtain a good generalization performance, the *T* is to be maximized.

Another direction of investigation is to estimate the fraction of opposite outlier class that is accepted (f_{O+}) by generating artificial outliers in and around the target set. By generating artificial outliers, the error on the outlier class can be estimated. Also the fraction of the outliers which is then accepted by the 1-SVM is now an estimate of the volume in the feature space covered by the 1-SVM. In [10], a method to generate uniformly distributed outliers using the box-procedure is present. The basic idea is to construct a hyperbox around the target objects, from which the artificial outliers are uniformly drawn. Tax and Duin [17] also extended the idea to generate uniformly distributed outliers within a hypersphere. The ideas behind these two approaches are the same. But instead of generating uniformly distributed outliers in a hyperbox, it uses a hypersphere and all the outlier objects are drawn from this hyperspere. They believe that it is more likely that the hypersphere solution can fit more tightly around the target class.

In this research, we propose another artificial outlier generation method, which can fit better around the target set. As it fits better, the error estimation using the artificial outliers can be more accurate and it can be used in high dimensional spaces.

III. MODEL SELECTION

In the following, we first discuss the impact of parameters of γ and ν on the performance of 1-SVM, and then describe the proposed data-skewness based outlier generation approach for parameter optimization.

A. The impact of parameters γ and ν

Using a toy data generated from four clusters of Gaussian distributed points with unit variance, we can easily observe the impact of parameters γ and ν on 1-SVM performance. The mean values of the four Gaussian data clusters are (0, 0), (3.5, 1), (5.5, 2), and (6, 5.5) respectively. Each cluster has 500 data points and totally there are 2000 sample points in the toy data. Fig. 1 shows the decision boundary obtained by choosing different parameter combination of γ and ν.

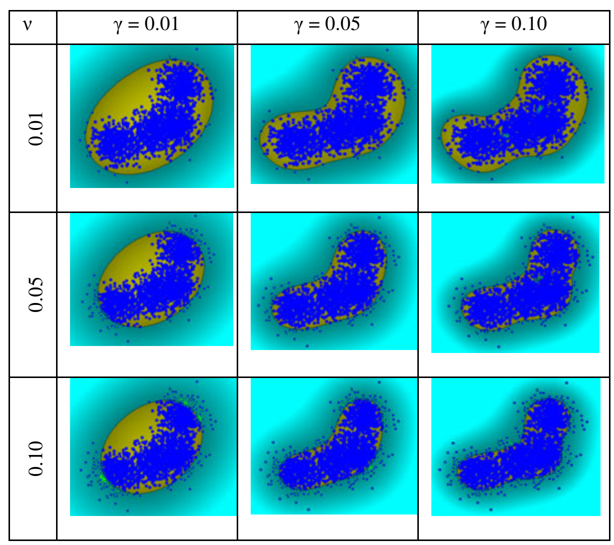


Fig. 1 The impacts of parameters γ and ν

The parameter ν controls the fraction of data points in the region. For a smaller ν, only few target objects are allowed to be outside the description. When we increase the ν from 0.01, 0.05 to 0.1, the fraction of data points rejected by the decision boundary is also increased, and the boundary is closer to the dataset. The parameter γ decides the non-linear characteristics of the decision function, in other words, it decides the "shape" of the region. For smaller γ, an approximate spherical solution is found (for $\gamma \rightarrow 0$, the rigid hypersphere solution is obtained). As we increase γ from 0.01, 0.05 to 0.10, the decision boundary becomes more sensitive to the data distribution. Schölkopf et al. [8] showed that ν is an upper bound for the fraction of target class objects outside the description. Having a larger value of ν, the approach is more tolerant to the outliers in the training data, but the generalization performance may decrease since more and more target objects in the training dataset are classified as outliers. On the other hand, having a large value of γ always make the model specific to the data description, unfortunately the over-fitting problem becomes more and more serious as we keep increasing the value of γ.

From the experiences of conventional supervised classification problem, to find good values of γ and ν, an error criterion must take the errors from both classes into account. That is, the fraction of the target class that is rejected (f_{T-}), and the fraction of opposite outlier class that is accepted (f_{O+}). The two errors are equivalent to the first and second kind errors, ε_I and ε_{II} respectively. There always exists a tradeoff between them. Shifting the decision boundary towards the opposite outlier class (or increasing the volume of data description) always decreases the first kind error ε_I, but increases the second kind error ε_{II}. The optimal choice of parameters is to balance the two errors and attempt to minimize both of them. Because only one class of target objects is available in the unsupervised classification problem, the error on the opposite outlier set (f_{O+}) has to be estimated in other ways.

B. Skewness-based Outlier Generation

We propose a new artificial outlier generation method using "skewing" technique. In the approach, the outlier objects are generated by setting slight offsets from the target objects, named "skewness". As all the outliers are skewed points of the target objects, we expect that the outlier objects are tightly distributed around the target set.

Similar to the hyperbox-based outlier generation approach, we define the error function as

$$G=\lambda f_{T-}+(1-\lambda) f_{O+}= \lambda \frac{\#SV}{N}+(1-\lambda) f_{O+}$$

We also define λ to control the tradeoff between these two evaluation measures. In 1-SVM, the target objects that do not fit into the decision boundary becomes support vectors. Thus, the fraction of the target class that is rejected (f_{T-}) can be easily estimated by the number of support vectors. The f_{O+} is estimated using artificial outliers. By minimizing the error function G, the optimal values of parameters γ and ν can be obtained. The remaining issue is how to generate the "skewness" points from the target objects.

The procedure to generate a skewness point from a target object is demonstrated as follows. For any data point $\mathbf{x} \in R^m$, assume $\mathbf{x} \equiv \{x_1, x_2, \ldots, x_m\}$. We define a skew function $y = sk(\mathbf{x})$, where $y=\{y_1, y_2, \ldots, y_m\}$ and $y_i = x_i + \alpha v_i r_i$ for all $i \in [1,m]$. Here,

- α is a positive number that describes the skew degree;
- v is the normalized standard deviation vector. If σ_i is the standard deviation on the i^{th} dimension, and
$$v_i \equiv \frac{\sigma_i}{\sqrt{\sum_{j=1}^{m} \sigma_j^2}};$$
- r is a normalized random vector. We choose to generate a random vector $\{R_i\}$ with length m with R_i is generated with a Gaussian distribution *N(0, 1)*, then $r_i \equiv \frac{R_i}{\sqrt{\sum_{j=1}^{m} R_j^2}}$ is a random unit vector.

Fig. 2 illustrates an example how skewed image of data point **x** is created. v_1 and v_2 shows the normalized standard deviation of the training data. The smaller ellipse around point **x** demonstrates the possible locations for sk(**x**). The generated skewed image of the data point **x** falls in this ellipse. Using the above skewness point generation procedure, we generate an outlier object for each target object. Thus, we have equal number of outlier objects and the target objects.

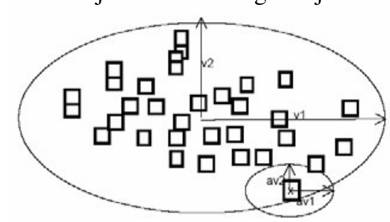


Fig. 2 Skewed Data Generation Example

As Fig. 2 indicates the skewed image of a data point **x** falls in the smaller ellipse around point **x**, there is a small chance that only a few skewness outliers falls outside the target data description. In this case, the fraction of f_{O+} is close to 1, which can not give us a good guide on selecting parameters. The problem can be mitigated by adjusting the value of skew degree α. In practice, we may generate multiple outlier objects for each target object, and the number is related to the target dimension. Generating enough outliers can help us decide the volume of the data description, and further guide the parameter selection process.

The skewed outlier generation approach is similar to both the hyperbox and hypersphere based outlier generation approaches in the sense that they all attempt to generate artificial outliers to estimate the second type of error (ε_{II} or f_{O+}). However, it is different from them in that non-uniformly distributed outlier objects are generated. In addition, the hyperbox (or hyperspher) based approach usually occupies a larger space. The proposed skewness-based outlier generation approach is expected to generate a set of "tighter" outlier objects around the target set.

IV. EVALUATION

For comparison, we have also implemented three existing parameter selection approaches discussed in Section II, including the tuning γ only method, ξαρ-estimate method, hyperbox-based uniform outlier generation method, and hypersphere-based outlier generation approach. Their performances are first investigated using the toy data set shown in Fig. 1, then, the real intrusion detection data simulated using NS-2 platform [19] are applied to evaluate their performances.

A. Toy data evaluation

(1) Fixing ν and tuning γ

The first approach we investigated is that the parameter ν is fixed priori to the highest allowable fraction of misclassification of the target class, and only the kernel parameter γ is tuned such that the f_{T+} first reaches 1-ν [14]. We set the parameter ν to 0.01, 0.05, and 0.10 respectively. That is, the highest allowable fraction of misclassification of the target class is 1%, 5% and 10% correspondingly. We vary the parameter γ from [0.20 0.15 0.10 0.05 0.01 0.005 0.001] and select the one when f_{T+} first reaches 1-ν.

Fig. 3 shows the fraction of accepted target objects vs γ values. For the case of ν=1%, the fraction of accepted target objects first reaches 99% at γ=0.05. For ν=5%, the corresponding γ value found is 0.10. The same value (γ=0.10) is also found for the ν=10%. In Fig. 1 we plotted these three selected models. The third one with ν=10% and γ=0.10 has the best performance on the training set, and the first model is poor. The reason is that the estimated error fraction (or outlier rate) on the training set is not appropriate in model 1. We have noticed that this method may help us picking a value of γ, but fixing the value of ν is not efficient as it can choose a large number of values. For different value of ν, we may obtain different γ values. If an inappropriate value of ν is used, then the model selected may not have a good performance.

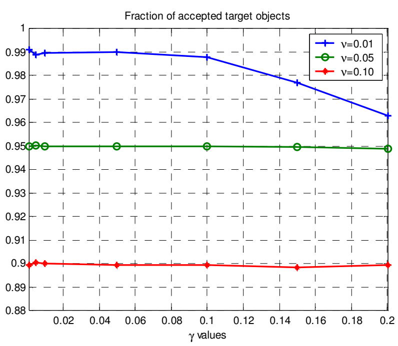


Fig. 3 The fraction of target objects that are accepted vs γ values

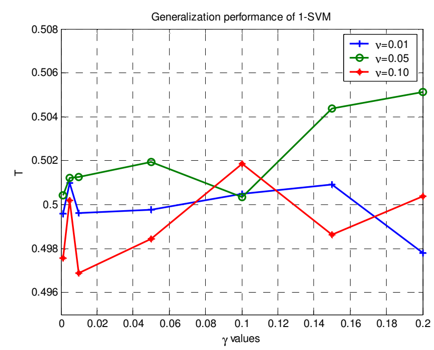
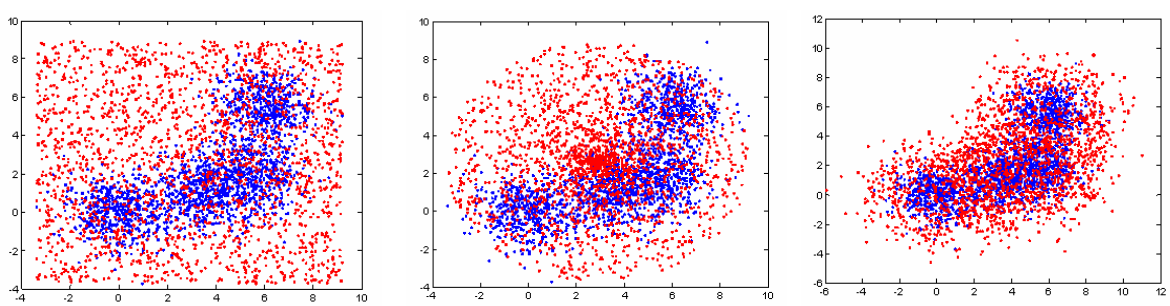


Fig. 4 Generalization of 1-SVM obtained by recall estimate method



(a) hyperbox (b) hypersphere (c) skewness

Fig. 5 Artificial outliers generated around the toy dataset
(blue dots are target objects and red dots represent outlier objects)

(2) The ξαρ-estimate method

The second approach we investigated is the ξαρ-estimate method, in which the generalization performance of 1-SVM is evaluated by the fraction of support vectors and the ξαρ-estimate of recall. We vary the parameter ν from [0.01, 0.05, 0.10] and γ from [0.001 0.005 0.01 0.05 0.10 0.15 0.20] respectively. We set equal weight to the two measures, that is λ=0.5. The generalization performance T is shown in Fig. 4.

This method results in several larger value of T, but no single solution can be obtained. One possible solution is ν=10%, γ=0.10, and the other is ν=5%, γ=0.05. Referring to the decision boundary shown in Fig. 1, both of them are reasonable solutions. But the method is not efficient on dealing with the over-fitting problem. When the model is very specific to the target distribution, in which the fraction of support vectors and the recall rate on the training set are high, the method always gives a high value. It can be easily observed in Fig. 4. When we increase the γ value further to 0.20, the T values also keep increasing. In Fig. 1 we can see that the model at these values is over-fitted to the target data distribution.

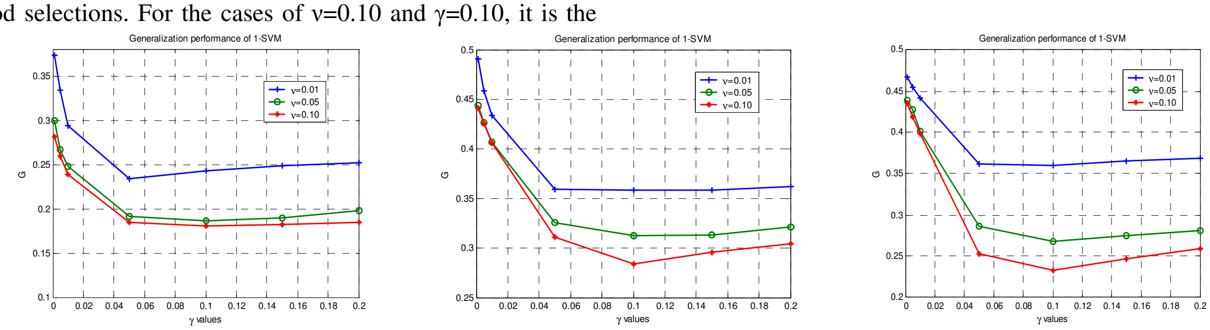
(3) Artificial outliers

In the previous Section we have discussed three artificial outlier generation methods, hyperbox-based, hypersphere-based, and our proposed skewness outlier generation. Fig. 5

shows the artificial outliers generated using these approaches around the toy dataset. The blue dots are the target data distribution, and the red dots represent the artificial outliers.

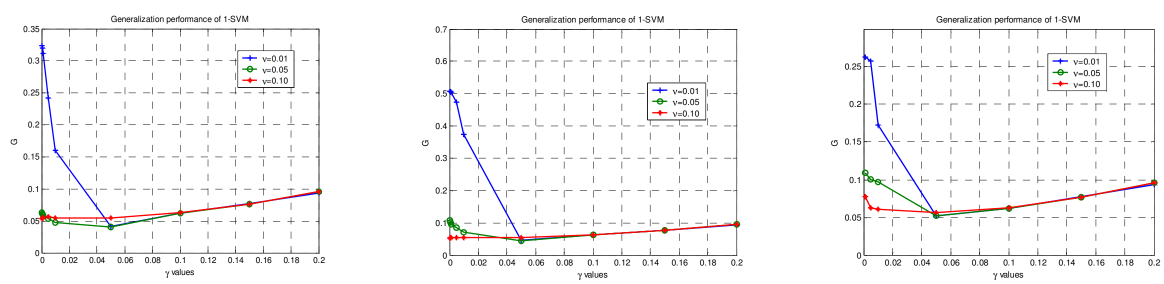
We select the parameters γ and ν by minimizing G. Initially the λ is set to 0.5. The performances of 1-SVM evaluated by different outlier generation approaches are shown in Fig. 6. In Fig. 6(a), we can see that when we set the allowable target object rejection rate to 10%, we always obtain a good performance (smaller G) than the cases of ν=0.01 and 0.05. In addition, the performance of the two cases (ν=0.05 and 0.10) are similar, and the performance of ν=0.05 is a little lower than ν=0.10. When we vary the γ value from 0.001 to 0.20, the performance measure G first decreases and then rise up. The parameter γ can be selected as the point when the G value stop decreasing and rising up.

For the performance of 1-SVM measured using the hyperbox-based artificial outliers, the best choice of parameters γ and ν is ν=0.10 and γ=0.10. Another possible choices of parameters γ and ν is ν=0.10 and γ=0.05, ν=0.05 and γ=0.10. By looking back to the decision boundaries we plotted in Fig. 1, all these choices of γ and ν combination are good selections. For the cases of ν=0.10 and γ=0.10, it is the best one from the visual observation. It is very interesting to notice that we can obtain the same observation from the other two approaches. The case of ν=0.10 achieves the best performance using all the different outlier generation methods, and the same best parameter combination is found by all these three approaches, even their performance are different slightly. We can conclude that all the three outlier generation approaches are efficient for this toy data distribution. To further evaluate the impact of λ, we vary the λ value in a range [0.1 0.3 0.5 0.7 0.9], and evaluate the performance of these three approaches again. Smaller the λ value, less effect of first type of error f_{T-}. Thus the main contribution of the error measure G is from the f_{O+}, the second type of error. On the other hand, larger value of λ makes the f_{T-} contribute more to the performance measure. The cases of λ=0.1 and λ=0.9 have inverse performance. If we use λ=0.3, the best parameter combination is ν=0.10 and γ=0.10 again. But if we use λ=0.7, the parameters selected will be ν=0.05 and γ=0.10. In real test, λ can be set to 0.5 to equalize the weight from both the two type of errors. We can also choose a litter larger value of λ (0.6 or 0.7) to consider more on the first type of error f_{T-}.



(a) hyperbox-based (b) hypersphere-based (c) skewness-based

Fig. 6 Generalization using different outlier generation approaches



(a) hyperbox-based (b) hypersphere-based (c) skewness-based

Fig. 7 Generalization on Cancer data with 9 dimensions

B. Breast Cancer data evaluation

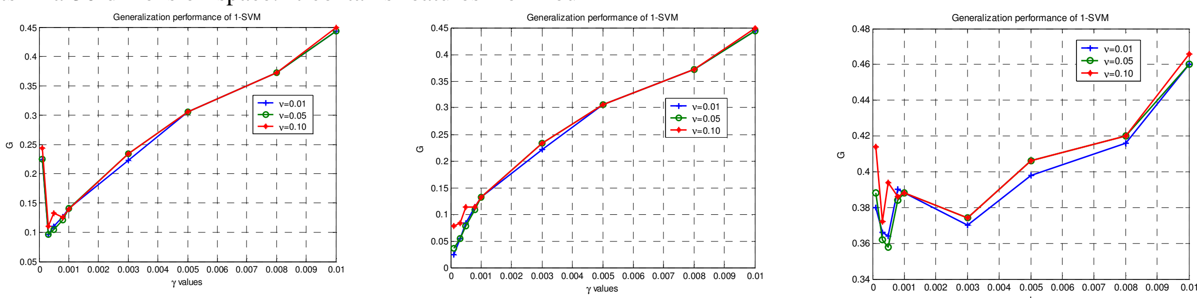
In the above experiments, we evaluated the performance of different approaches using the 2-dimensional toy data, and observed that the three approaches all performs very well. We are also interested to see their performance on higher dimensional spaces. Thus we select one common dataset from UCI Repository. The dataset we used is Wisconsin Breast Cancer (1992). The popular Wisconsin breast cancer data set contains 9 attributes, 684 instances and two classes (C=2, m=9, N=684). We select the normal patterns (non-cancer) as the training dataset. Totally there are 444 normal data records. We generate 10,000 outliers for hyperbox and hypersphere based approaches.

The generalization performances obtained are shown in Fig. 7. We have seen that all the three approaches work well for this dataset. During the experiments, we have noticed that the number of outliers generated is important for both the hyperbox and hypersphere based outlier generation approaches. Large number of outliers is to make sure there is enough outlier points falling in the target distribution and help us estimate the volume of the target data description. For the skewness based outlier generation approach, we can adjust the skewness degree α to a larger value in order to generate enough outlier points outside of the target description.

C. *Parameter Selection for Anomaly Detection*

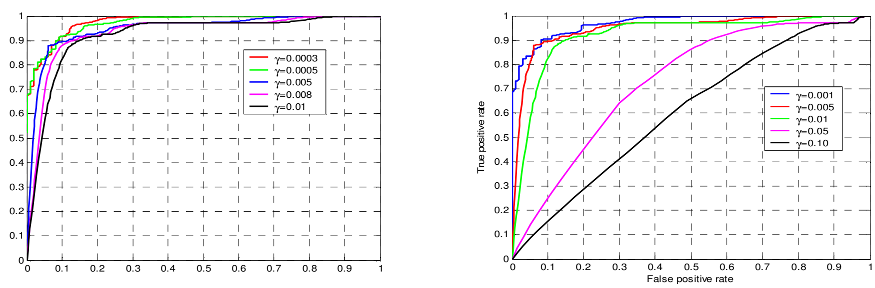
We have applied three different outlier generation approaches into the anomaly detection data obtained in our previous anomaly detection project studies [5][11], aiming to find a good parameter combination to enhance the intrusion detection performance. The dataset used here is the same as we used for evaluating anomaly detection algorithms, discussed in [5][11]. The data set contains 1000 normal data points in a 30 dimension space. It contains features from four aspects, routing packet propagation, route table changes, data packet transmission, and node mobility**.**

We organized the experiments as follows. First, we evaluate the three different approaches for parameter selection in anomaly detection model. Then we show the performance of the selected parameters. The classification performance on another independent test set is used as the evaluation criteria. Since all the training dataset are from normal network operations, the estimated outlier rate will be very low. We vary the parameter ν from [0.01, 0.05, 0.10] and γ from [0.0001 0.0003 0.0005 0.0008 0.001 0.003 0.005 0.008 0.01] respectively. For both the hyperbox-based and hypersphere-based approach, we generate 200000 outlier objects to test the performance. For skewed outlier generation, we set the skewness degree to 2.



(a) hyperbox-based (b) hypersphere-based (c) skewness based

Fig. 8 Generalization for parameter selection on anomaly detection dataset



(a) ν=0.01 (b) ν=0.05

Fig. 9 Performance of the anomaly detection

Fig. 7 shows the performance of the three different outlier generation approaches. We can see that the hyperbox-based and skewness based approaches performs well. When we set the allowable target object rejection rate to 1% or 5%, we obtain better performance (smaller G) than the cases of ν=0.10. In addition, the performance of the two cases (ν=0.01 and ν=0.05) are similar. When we vary the γ value from 0.0001 to 0.01, the performance measure G first decreases and then rise up. For the performance of 1-SVM measured using the hyperbox-based artificial outliers, the best choice of parameters γ and ν is ν=0.01 and γ=0.0003. Another possible choice of parameters γ and ν is ν=0.05 and γ=0.0003. For the skewness based outlier generation approach, we get the similar results. The best choice is ν=0.01 and γ=0.0003 or ν=0.05 and γ=0.0005.During our experiments, we have also noticed that the hypersphere-based outlier generation approach is not efficient for the anomaly detection data. By generating 200000 outlier objects, we still can not estimate the volume occupied by the target description as only few outliers fall in the target distribution. It may due to the reason that the anomaly detection data has a large variance difference for different dimensions. In this case the hypersphere-based approach needs to cover a very large space. However, we do not have enough memory to generate so many objects.

Comparing all the approaches for anomaly detection parameter selection, the skewness based approach performs better than the hyperbox-based and hypersphere-based approaches. They require fewer outlier objects and the outlier objects are more tightly around the target distribution. We also want to see how reasonableness of the selected parameters for anomaly detection. We plotted the classification accuracy in ROC curves for different parameter combinations in Fig. 9. It can be easily see that the selected parameters (ν=0.01, γ=0.0003 and ν=0.05, γ=0.0005) all perform very well on the testing dataset. It further verifies the effectiveness of the proposed parameter selection approach.

V. Conclusions

In this paper, we have investigated the parameter selection issue for anomaly detection in wireless ad hoc networks. The purpose of this research is to enhance the efficiency and effectiveness of anomaly detection by selecting good parameters. In particular, we mainly addressed the 1-SVM based anomaly detection model, as it has shown its effectiveness through several researches.

We have investigated the performance of several existing parameter selection mechanisms, and also proposed a skewness-based outlier generation approach. From experimental results, we have observed that the skewness based approach performed better than the hyperbox-based and hypersphere-based approaches. It requires fewer outlier objects and the generated outlier objects are more tightly around the target distribution. Generating a “tighter” outlier set can help us accurately estimate the volume of target distribution, thus improves the effectiveness of parameter selection.

Our current work on parameter selection is based on scanning technique. That is, we scan the possible choice of parameters, and selected the one with best performance. We will further extend the work by optimizing the parameters using Genetic Algorithms. Finally we note that this research considers the parameter selection for 1-SVM based anomaly detection model. It can also be used in any unsupervised classification or outlier detection problem for parameter optimization.

Acknowledgment

The research effort depicted is sponsored by the US Army Research Office and the content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

[1] S. Marti, T. Giuli, K. Lai, and M. Baker, “Mitigating Routing Misbehavior in Mobile Ad Hoc Networks,” Proceedings of the 6th International Conference on Mobile Computing and Networking (MOBICOM’00), pp.255-265, August 2000.

[2] H. Deng, W. Li, and Dharma P. Agrawal, "Routing Security in Ad Hoc Networks," IEEE Communications Magazine, Special Topics on Security in Telecommunication Networks, Vol. 40, No. 10, October 2002, pp. 70-75.

[3] Y. Zhang and W. Lee, “Intrusion Detection in Wireless Ad-Hoc Networks,” Proceedings of the 6th International Conference on Mobile Computing and Networking, MobiCom 2000, pp. 275-283, August 2000.

[4] Y. Huang, W. Fan, W. Lee and P. S. Yu, “Cross-Feature Analysis for Detecting Ad-hoc Routing Anomalies,” Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS'03), May 2003.

[5] H. Deng, R. Xu, J. Li, F. Zhang, R. Levy, W. Lee, “Agent-based Cooperative Anomaly Detection for Wireless Ad Hoc Networks,” The Twelfth International Conference on Parallel and Distributed Systems (ICPADS’06), 2006.

[6] D. Sterne, P. Balasubramanyam, D. Carman, B. Wilson, R. Talpade, C. Ko, “A general cooperative intrusion detection architecture for MANETs,” Proceedings of the Third IEEE International Workshop on Information Assurance, March 2005.

[7] B. Schölkopf, R. Williamson, A. Smola, J. S. Taylor and J. Platt, “Estimating the support of a high-dimensional distribution,” Technical report MSR-TR-99-87, Microsoft Research, 1999.

[8] B. Schölkopf, R. Williamson, A. Smola, J. S. Taylor, “SV estimation of a distribution’s support,” Advances Neural Information Processing Systems, 1999.

[9] D. M. J. Tax and R. P. W. Duin, “Support vector domain description,” Pattern Recognition letters, 10(11-13):1191-1199, 1999.

[10] D. M. J. Tax, One-class classification, PhD thesis, Delft University of Technology, http://www.ph.tn.tudelft.hl/~davidt/thesis.pdf, 2001.

[11] H. Deng, R. Xu, F. Zhang, C. Kwan, L. Haynes, “Agent-based Distributed Intrusion Detection Methodology for Wireless Ad Hoc Networks,” The International Conference on Security and Management, June 26-29, 2006

[12] Q. A. Tran, H. Duan, X. Li, “One-Class Support Vector Machines for Anomaly Network Traffic Detection,” Proceedings of the APAN Network Research Workshop, 2004.

[13] Q. A. Tran, Q. L. Zhang, and X. Li, “Evolving Training Model Method for One-Class SVM,” Proceedings of the 2003 IEEE International Conference on Systems, Man & Cybernetics (SMC 2003), pp 2388~2393, Washington D.C., USA, 2003.

[14] R. Unnthorsson, T. P. Runarsson, M. T. Johnson, “Model selection in one-class ν-SVMs using RBF kernels,” http://www.hi.is/~runson/svm/.

[15] N. Cristianini and j. S. Taylor, “An introduction to Support Vector Machines and Other kernel-based learning methods,” Cambridge University Press, 2000.

[16] T. Joachims, “Estimating the generalization performance of an SVM efficiently,” Proceedings of ICML’00, 17th International Conference on Machine Learning, 2000.

[17] D. M. J. Tax, R. P. W. Duin, “Uniform object generation for optimizing one-class classifiers,” Journal of Machine Learning Research, pp.155-173, 2001.

[18] M. Luban and L. P. Staunton, “An efficient method for generating a uniform distribution of points within a hypersphere,” Computers in Physics, 2(6):55-60, 1988.

[19] K. Fall and e Varadhanm, The ns Manual (formely ns Notes and Documentation), 2000.