This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.

Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks

Sutharshan Rajasegarar[1], Christopher Leckie[2], Marimuthu Palaniswami[1]
ARC Special Research Center for Ultra-Broadband Information Networks (CUBIN)
[1]Department of Electrical and Electronic Engineering
[2]NICTA Victoria Research Laboratory
Department of Computer Science and Software Engineering
University of Melbourne, Australia.
Email: {r.sutharshan, swami}@ee.unimelb.edu.au, caleckie@csse.unimelb.edu.au

James C. Bezdek
Computer Science Department
University of West Florida, USA.
Email: jbezdek@uwf.edu

***Abstract*—Anomaly detection is an important challenge for tasks such as fault diagnosis and intrusion detection in energy constrained wireless sensor networks. A key problem is how to minimise the communication overhead in the network while performing in-network computation when detecting anomalies. Our approach to this problem is based on a formulation that uses distributed, one-class quarter-sphere support vector machines to identify anomalous measurements in the data. We demonstrate using sensor data from the Great Duck Island Project that our distributed approach is energy efficient in terms of communication overhead while achieving comparable accuracy to a centralised scheme.**

I. Introduction

Wireless sensor networks are formed using large numbers of cheap, tiny and compact sensors which have inbuilt wireless radios for communication [1]. They have limited power, bandwidth and memory. These inherent constraints on the network make it more vulnerable to faults and malicious attacks such as denial of service attacks, black hole attacks and eavesdropping [2], [3]. Therefore, identifying misbehaviors or *anomalies* in the network is important to provide reliable and secure functioning of the network. An *anomaly* or *outlier* in a set of data is defined as an observation that appears to be inconsistent with the remainder of the data set [4].

Misbehaviors in the network can be identified by analysing either sensor data measurements or traffic related attributes in the network. Note that the underlying distribution of these measurements may not be known *a priori*. A key challenge is to identify anomalies with acceptable accuracy while minimising energy consumption in the wireless sensor network.

In sensor networks, the majority of the energy is consumed in radio communication rather than in computation [5], [6]. For example, in Sensoria sensors and Berkeley motes, the ratio between communication and computation energy consumption ranges from 10^3 to 10^4 [7]. Hence, there are advantages to increasing computational overheads in order to reduce communication requirements in the network, and thus prolong the lifetime of energy-limited wireless sensor networks. In this paper, we propose an energy efficient non-parametric distributed approach for anomaly detection in wireless sensor networks, which performs in-network processing in order to reduce the need for radio communication in the network.

Recent related work in anomaly or outlier detection in sensor networks can be found in the literature. Palpanas et al. [8] and Subramanium et al. [9] have proposed the use of kernel density estimators for online distributed outlier detection in streaming data in sensor networks. In this distributed approach, a random sample of the data set within the window of measurements are communicated between sensor nodes along with the bandwidth parameter of the kernel function that is used to model the data. Onat et al [10] have identified anomalies using a rule based technique on a predefined statistical model. Loo et al [11] have proposed a cluster based intrusion detection scheme for anomaly detection. However they have not considered co-operation between nodes.

One class support vector machines (SVMs) have been proposed as a technique for outlier detection. Techniques have been proposed based on hyperplanes [12] and hyperspheres [13]. Navia-Vazquez et al. [14] and Flouri et al. [15] have proposed distributed and incremental techniques for training SVMs in sensor networks. However, a challenge for these SVM formulations for this application is their computational complexity.

In our previous work [16] for anomaly detection, a cluster-based distributed approach was proposed, where the data measurements are clustered and summary information for each cluster is communicated between nodes for performing distributed anomaly detection and classifying the data. In this paper, we propose another communication efficient distributed technique based on a one-class quarter sphere SVM.

The rest of the paper is organised as follows. We formally introduce the problem in Section II. The quarter sphere support vector machine formulation and our distributed approach are explained in Section III. An empirical comparison of the centralised and distributed approaches is provided in Section IV.

II. Problem Statement

We consider the problem of anomaly detection in a wireless sensor network where the sensor nodes are connected by a routing tree such as Figure 1(a). The sensors are time synchronised and deployed in a homogeneous environment, where the measurements have the same unknown distribution.

1-4244-0353-7/07/$25.00 ©2007 IEEE

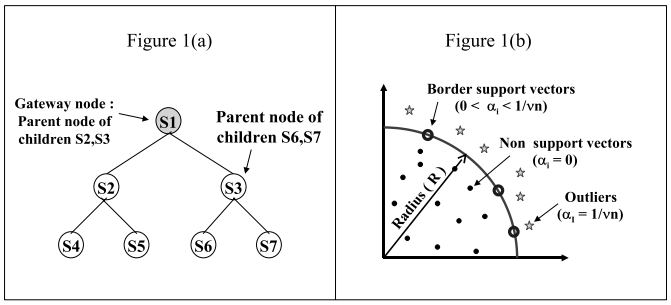


Fig. 1. (a) Example of a multi-level hierarchical organisation of sensors. (b) Geometry of the one-class quarter sphere support vector machine.

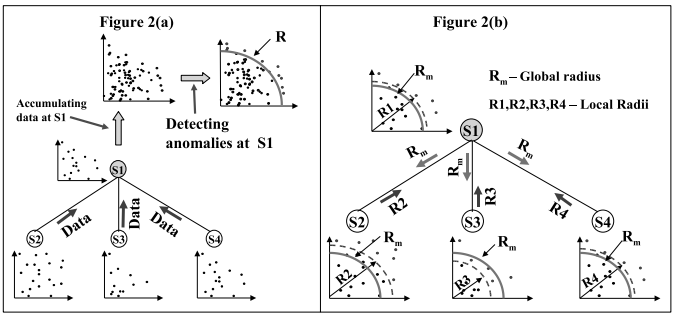


Fig. 2. Centralised and distributed detection. (a) Centralised approach: Data from children nodes are accumulated at parent node *S1*, and then anomaly detection is performed at *S1* to identify outliers. (b) Distributed approach: Children nodes *S2*, *S3*, *S4* and parent node *S1* perform local anomaly detection. Children nodes transmit their local radii (*R2*, *R3*, *R4*) to the parent node *S1*. Parent node *S1* computes the global radius R_m using the received and its own local radii. Then *S1* transmits the global radius R_m to the children nodes, which each perform global detection using R_m.

At every time interval Δ_i, each sensor s_j in a set of sensor nodes $S = \{s_j : j = 1...s\}$ measures a data vector x_i^j. Each data vector is composed of attributes x_{ik}^j, where $x_i^j = \{x_{ik}^j : k = 1...d\}$ and $x_i^j \in \Re^d$. After a window of n measurements, each sensor s_j has collected a set of measurements $X_j = \{x_i^j : i = 1...n\}$. Consider any parent node $s_p \in S$ in the network having a set of children nodes $S_c = \{s_c : c = 1...l, l \leq (s-1)\}$ and $S_c \subset S$. Our aim is to identify outliers $O \subset X$ at the parent node s_p in the combined set of measurements $X = \bigcup_{j=1..(l+1)} X_j$. This framework provides a flexible model for detecting anomalies at any level of the routing hierarchy in the network.

III. Anomaly Detection

A *naive* approach for detecting anomalies in wireless sensor networks is to use centralised detection. In this approach, all the sensor measurements are communicated to the parent node and then an anomaly detection algorithm is run at the parent node to classify the data.

Figure 2(a) shows an example of a centralised approach for a single-level hierarchical topology. Here, sensor nodes *S2*, *S3* and *S4* send their data measurements to their parent node *S1*. Node *S1* then combines its own data with the received data and performs anomaly detection on the combined data.

This approach is highly inefficient in terms of communication overhead in the network. Therefore, it is desirable to perform more in-network computation and communicate only summary information between nodes. Our proposed distributed approach performs detection on local data at each node and communicates only summary information among the sensor nodes for global classification of the data.

We now describe the anomaly detection algorithm and our proposed distributed detection approach in more detail.

A. Anomaly Detection Algorithm

Tax and Duin [13] have proposed a one-class support vector machine (SVM) formulation for outlier detection, which is based on fitting a hypersphere to the data in a higher dimensional space. It is observed in [17] that the "typical distribution of features used in intrusion detection systems (IDS) is one-sided on $\Re_0^+$". A geometric construction which takes into account this one-sidedness of the data distribution can be obtained by extending the hypersphere based one-class SVM approach. Laskov et al. have extended such an approach into a special type of SVM called a one-class quarter-sphere SVM [17]. Here we provide the mathematical formulation of the one-class quarter-sphere SVM.

In order to ease the notation burden, we drop the superscript j that denotes the data measurements of a sensor node s_j. Consider a data vector x_i in the *input space* from a set of data vectors $X_j = \{x_i : i = 1..n\}$ mapped to a higher dimensional space, called the *feature space*, by some non-linear mapping function $\phi(x_i)$. The mapped vector $\phi(x_i)$ in the feature space is called the *image vector*. The aim is to fit a hypersphere with minimal radius R, having its center fixed at the origin and encompassing a majority of the image vectors. This can be formulated as an optimisation problem as follows:

$$\begin{aligned} \min_{R\in\Re,\xi\in\Re^n} \quad & R^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i \\ \text{subject to:} \quad & \|\phi(x_i)\|^2 \leq R^2 + \xi_i, \\ & \xi_i \geq 0. \end{aligned} \tag{1}$$

where $\{\xi_i : i = 1...n\}$ are the slack variables that allow some of the image vectors to lie outside the sphere. The parameter $\nu \in (0, 1)$ is the regularisation parameter which controls the fraction of image vectors that lie outside the sphere, i.e., the fraction of image vectors that can be *outliers* or *anomalies* [17]. The Lagrange function for this optimisation is:

$$\begin{aligned} L(R, \xi_i, \alpha_i, \gamma_i) = \quad & R^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\gamma_i\xi_i \\ & - \sum_{i=1}^{n}\alpha_i(R^2 - \|\phi(x_i)\|^2 + \xi_i) \end{aligned} \tag{2}$$

where $\alpha_i \geq 0, \gamma_i \geq 0, \forall i$ are the Lagrange multipliers. Equating the partial derivatives of L with respect to R and ξ_i to zero yields:

$$\frac{\partial L}{\partial R} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n}\alpha_i = 1 \tag{3}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad \gamma_i = \frac{1}{\nu n} - \alpha_i \tag{4}$$

From (4) and using $\alpha_i \geq 0, \gamma_i \geq 0$, we can obtain $0 \leq \alpha_i \leq \frac{1}{\nu n}$. Substituting (3) and (4) in (2) results in:

$$L = \sum_{i=1}^{n} \alpha_i (\phi(x_i).\phi(x_i)) \tag{5}$$

where $(\phi(x_i).\phi(x_i)) = \|\phi(x_i)\|^2$ is the inner product of the image vector $\phi(x_i)$. Using the kernel trick, the inner product can be replaced by a kernel function $k(x_i, x_i)$ analogous to [18]. Hence, for the above primal problem (1), we can obtain the dual problem as

$$\begin{aligned} \min_{\alpha \in \Re^n} \quad & -\sum_{i=1}^{n} \alpha_i k(x_i, x_i) \\ \text{subject to:} \quad & \sum_{i=1}^{n} \alpha_i = 1, \\ & 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad i = 1...n. \end{aligned} \tag{6}$$

This dual problem (6) is a linear optimisation problem, so the $\{\alpha_i\}$ can be obtained using widely available linear optimisation techniques [19]. Compared to some existing one-class SVM formulations [13], [12], which require solving a quadratic optimisation problem, this formulation with linear optimisation is advantageous in terms of its computational complexity [17].

From the solution of (6) for $\{\alpha_i\}$, the image vectors can be classified as follows (refer to Figure 1(b)). The image vectors with $\alpha_i = 0$ will fall inside the sphere. The image vectors with $\alpha_i > 0$ are called the *support vectors*. Support vectors with $\alpha_i = \frac{1}{\nu n}$ are termed as *outliers*, which fall outside the sphere. Support vectors with $0 < \alpha_i < \frac{1}{\nu n}$ will reside *on* the surface of the sphere, and hence are called the *border support vectors*. Moreover, the radius of the sphere R can be obtained using $R^2 = k(x_i, x_i)$, for any border support vector x_i.

Further, in (6) it can be observed that the solution is affected only by the norms of the non-linear mapping of data vectors using the kernel $k(x_i, x_i)$. This creates a problem for the application of this approach with distance based kernels, as the norms of the kernels are now equal for all data vectors [17]. In order to alleviate this problem, the image vectors in the feature space are *centered* [20], [21] in that space using:

$$\tilde{\phi}(x_i) = \phi(x_i) - \frac{1}{n} \sum_{i=1}^{n} \phi(x_i).$$

In other words, the mapped vectors are subtracted from the mean in the feature space. The dot product $\tilde{K} = (\tilde{\phi}(x_i).\tilde{\phi}(x_j))$ of the centered image vectors can be obtained in terms of kernel $K = k(x_i, x_j) = (\phi(x_i).\phi(x_j))$ as follows [21], [17]:

$$\tilde{K} = K - 1_n K - K 1_n + 1_n K 1_n \tag{7}$$

where 1_n is an $n \times n$ matrix with all values equal to $1/n$. Once the image vectors are centered, the norms of the kernels are no longer equal. Hence the dual problem (6) can now be solved.

B. Distributed Anomaly Detection

We have developed a distributed approach to anomaly detection, which extends the quarter-sphere SVM scheme as follows:

- Each sensor node s_j runs the above anomaly detection algorithm (Section III-A) on its local data and identifies the local anomalies and the local radius R_j of the quarter-sphere SVM. It keeps the local radius R_j and the *norms* of the image vectors in memory. Norms of the image vectors are provided by the diagonal elements of the kernel matrix $\tilde{K}$, i.e., the squared norm of $\tilde{\phi}(x_i)$ is given by
$$\left\|\tilde{\phi}(x_i)\right\|^2 = (\tilde{\phi}(x_i).\tilde{\phi}(x_i)) = \tilde{k}(x_i, x_i)$$
- Each sensor node s_j sends its radius information R_j to its parent node s_p.
- The parent node s_p collects the radius information from its children and combines them with its own local radius. It then computes the global radius R_m, which is used for global anomaly detection.
  In order to identify the best strategy for global radius computation R_m that gives a comparable performance with that of centralised detection, we considered four strategies of global radius computations, namely, using the *mean*, *median*, *maximum* or *minimum* of the combined radii from the parent node and its children.
- Parent node s_p sends the global radius R_m to its children.
- Children nodes compare the *norms* of their data vectors with the global radius R_m and classify them as globally anomalous or normal. A data vector x_i is identified as globally anomalous if its norm $\tilde{k}(x_i, x_i) > R_m^2$.

Figure 2(b) provides an example of our scheme for distributed detection in a sensor network with a single-level hierarchical topology. In this network, first, sensor nodes *S1*, *S2*, *S3*, and *S4* perform anomaly detection on their local data and compute the radii information *R1*, *R2*, *R3* and *R4*. The local radii are shown using dotted curves in the figure. Also, they keep in memory the norms of their local data. Second, nodes *S2*, *S3* and *S4* transmit their radii *R2*, *R3* and *R4* to the parent node *S1*. Third, parent node *S1* computes the global radius R_m, based on a chosen strategy, using the combined radius information, which is formed from its own radius *R1* and its children's radius information *R2*, *R3* and *R4*. Fourth, parent *S1* sends back the global radius R_m to the children *S2*, *S3* and *S4*. The children nodes classify their local data by comparing the norms of their local data vectors with the global radius R_m. The global radius is shown using a continuous curve in the figure.

The global radius computation can be performed at any parent node (or at any level) of the hierarchy. For example, in Figure 1(a), the global radius computed at the parent node *S2* will consider the radii from its children *S4* and *S5* only. Then, *S4* and *S5* will perform global detection using the global radius sent by *S2*. In this case, the region considered for distributed detection is the region covered by the nodes *S2*,

S3 and *S4*. If the global radius is computed at the top most parent node *S1*, then it will consider radius information from children *S2*, *S3*, *S4*, *S5*, *S6* and *S7*. Here, the region considered for distributed detection will be that of all the nodes in the topology. Therefore, this distributed detection methodology and the hierarchical topology provides flexibility to the user in selecting the coverage region for distributed computation.

Our distributed detection approach involves communication of the radius information only twice between the parent and the children nodes during a time window of measurements. This is a significant reduction in communication overhead in the network compared to the centralised approach, where the whole set of data vectors is transmitted among the nodes. Further, as the sensor network size scales, anomaly detection using the centralised scheme becomes impractical as the amount of data communicated and collected at the central nodes become overwhelmingly large. However, our disributed approach is scalable as it performs anomaly detection at the local nodes and the number of data vectors involved is limited to the time window of measurements.

Moreover, our distributed detection approach is not limited to hierarchical topologies of the sensor nodes. It is applicable to any network topology where a set of sensors have communication capabilities with their neighbours in the network. In this case, any sensor node in a connected group of sensors can be selected as a *leader node* for performing the global radius computation. This also gives flexibility for the leader node to select or ignore any participating neighbours for the computation of the global radius. This provides robustness against faulty nodes in the network, which are more prevalent in energy constrained wireless sensor networks.

C. Complexity Analysis

The one-class quarter-sphere SVM algorithm (Section III-A) involves solving of a linear optimisation problem. There are several algorithms available for linear programming in the literature [19]. The simplex algorithm is extremely efficient in practice, although it has been shown to have worst case exponential complexity in the number of variables [22]. Polynomial time algorithms such as the interior point methods incur $O(n^3)$ arithmetic operations and have a complexity of $O(\sqrt{n}L)$ iterations [19], where n is the number of variables and L is the size of the optimisation problem, i.e., roughly the number of bits required to represent the problem.

In our scheme, once the optimisation is performed, each node has to keep only the norms of the data vectors and the radius value in memory. Hence the memory complexity of the algorithm for each sensor node is $O(n)$, where n is the number of data vectors in the time window of measurements.

On receipt of the radius information from the children, the parent node s_p computes the global radius using one of the strategies (mean, median, maximum, minimum). This involves a computational complexity of $O(l)$, where l is the number of children of the parent node s_p.

Once the sensor nodes receive the global radius R_m from their parent node, they compare the norms of their local data vectors with the global radius. This involves a single pass over the number of data vectors in that time window of measurements, with computational complexity $O(n)$.

IV. Evaluation

In our evaluation, we consider a three-level hierarchical organisation of wireless sensor nodes as shown in Figure 1(a). Our aim is to compare the performance of the proposed distributed approach with the centralised approach.

The data in our evaluation are a set of real sensor measurements gathered from a deployment of wireless sensors in the Great Duck Island project [23]. In 2003, a set of wireless sensors were deployed in the Great Duck Island in Maine, USA for monitoring the habitat of a sea bird called the *Leach's Storm Petrel* [24]. They recorded light, temperature, pressure and humidity measurements at 5 minutes intervals.

We consider data measurements of seven sensor nodes, namely the nodes 101,109,111,116,118,122 and 123. A 24 hour period of data recorded on 1st July 2003 was used in our evaluation. We used three attributes: humidity, temperature and pressure measurements for each data vector. The data is cleaned manually by removing erroneous and spurious measurements with the help of scatter plots. The cleaned data is labeled as *Normal* for use in our evaluation. A hierarchical topology was formed with node 101 as the top most parent (gateway node), nodes 109 and 111 as intermediate parent nodes and the others as leaf nodes.

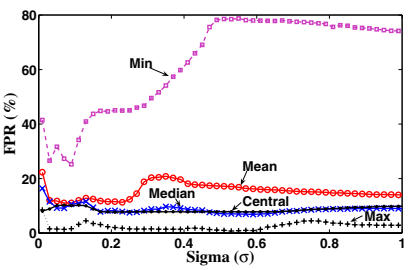
A randomly generated set of anomalous data was introduced into the tails of the distribution of each attribute for two of the sensor nodes (nodes 118 and 123). The anomalous data for each node comprised a set of 20 data vectors drawn from a uniform distribution over the tails of the collected measurements. Histogram plots for each attribute of the measurements were used to find the distribution of the attributes and the tail positions to place the anomalies. The introduced anomalous data were labeled as *Anomalies*. The data measurements were transformed to zero mean and unit variance and then normalised to the range [0,1] uniformly, using a data conditioning approach as in [16].

The distributed and centralised algorithms were implemented in Matlab, while utilising some of the functions from PRtools [25] and DDtools [26]. We have performed two simulations, one for different global radius computation strategies, and the other for various kernel functions.

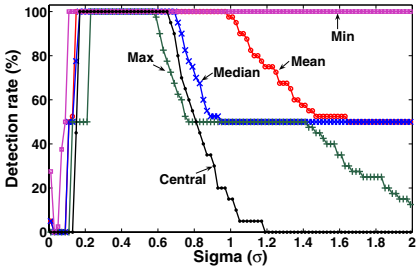
A. Evaluation for Different Global Detection Strategies

We used the RBF kernel (radial basis function) as the distance based kernel for this evaluation. The RBF kernel function for data vectors x_i and x_j is given as $k_{rbf} = \exp(-\left\|x_i - x_j\right\|^2/\sigma^2)$, where σ is the width parameter of the kernel function.

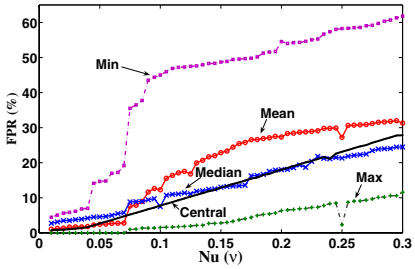
We have examined the effect of varying two parameters: the regularisation parameter ν (listed as "Nu" in Figure 3) ranging from 0.01 to 1 in intervals of 0.005, and the kernel width parameter σ ("Sigma" in Figure 3) ranging from 0.01 to 3.0 in intervals of 0.02. The radius of the sphere R is computed



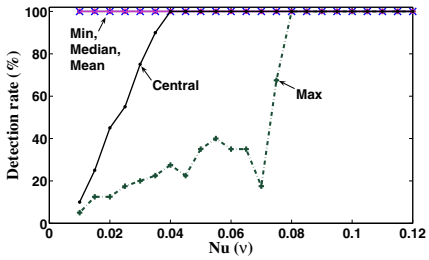
(a) False positive rate (%) Vs Sigma (σ)



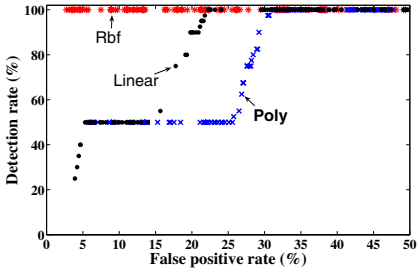
(b) Detection rate (%) Vs Sigma (σ)



(c) False positive rate (%) Vs Nu (ν)



(d) Detection rate (%) Vs Nu (ν)



(e) Receiver operating characteristics curve

Fig. 3. Centralised detection (Central) and distributed detection: (a), (b), (c), (d) Graphs for different strategies of global radius computation: maximum (Max), minimum (Min), median (Median) and mean (Mean). RBF kernel is used. (e) Receiver operating characteristics curve (ROC curve) for different kernel functions: RBF kernel (Rbf), polynomial kernel (Poly) and linear kernel (Linear).

either using any border support vector or using the mean of the support vectors should such border support vectors not exist. We considered each of the global radius computation strategies separately in our simulations, i.e., maximum (Max), minimum (Min), median (Median) and mean (Mean). In each simulation, we recorded the false positives, which occur when a normal measurement is identified as anomalous by the detector, and the true positives, which occur when an actual anomalous measurement is correctly identified by the detector. The false positive rate (FPR) is computed as the percentage ratio between the false positives and the actual normal measurements, and the detection rate (DR) is computed as the percentage ratio between the true positives and the actual anomalous measurements. Here, we report the results for the global radius computation and the centralised detection at the top most parent node (S1) of the topology (Figure 1(a)).

Figures 3(a) and 3(b) show graphs of the false positive rate and detection rate with varying σ values for the centralised (Central) and distributed detection scenarios. For this simulation the ν value is fixed at 0.10. All four strategies of the global radius computation are shown (Max, Min, Median, Mean) for the distributed detection scenario.

The results using the global radius computed using the Median and the Mean closely follow the results for the centralised approach, while Max and Min show considerable deviations from the centralised results. The global radius computed using Max has the largest value. In this case, there is a greater chance that most of the data vectors will fall inside the sphere, thus giving the minimum false positive rate amongst all the strategies. Similarly, the global radius computed using Min will have the smallest radius. This causes most of the data vectors to fall outside the sphere, resulting in higher false positives. The Median is more robust to extreme radius values, and hence gives better performance than the Mean, which is more biased towards extreme radius values and thus results in more false positives than when using the Median.

Further, Figure 3(b) shows the sensitivity of the detector with the kernel width parameter σ. Better detection performance is observed for the σ values between 0.2 and 0.6. Hence, it is important to select the kernel parameter to attain acceptable performance from the detector. In practice, these values can be selected by training the system before deployment. In addition, this parameter setting can be refined by updating the parameter based on values obtained in previous time windows. This is a topic to be investigated in our future research.

Figures 3(c) and 3(d) show graphs of the false positive rate and detection rate with varying regularisation parameter values ν for the centralised and distributed detection scenarios. For this simulation the σ value is fixed at 0.25. Here also similar trends are observed for Max, Min, Mean and Median as explained above. In Figure 3(c), for the centralised case, the false positive rate increases in proportion to ν. This is expected as the parameter ν indicates exactly the fraction of the data vectors that can be outliers [17]. In Figure 3(d), the sensitivity of the detector with ν can be observed. Better performance is observed for values beyond 0.08.

The communication overhead for distributed detection in

the network is 12 times the cost of communicating one radius value between a pair of sensor nodes. For the centralised case, 1429 data vectors (each with 3 attributes) are transmitted to the gateway node. Therefore, there is a 357 fold reduction in communication overhead achieved in the distributed case. These significant savings are achieved while obtaining comparable accuracy with the centralised case for the distributed detection.

B. Evaluation for Different Kernel Functions

In this evaluation, we used the Median strategy for the global radius computation. We considered three kernel functions in our evaluations: (1) A distance based RBF kernel function k_{rbf} with the width parameter $\sigma = 0.25$; (2) A polynomial kernel function $k_{poly} = (x_i.y_j + 1)^p$, where p is the degree of the polynomial and is set to 3; (3) A linear kernel function $k_{linear} = (x_i.y_j)$.

We performed distributed detection for varying ν values in the range from 0.01 to 1.0 in intervals of 0.005. Kernel centering (7) is performed for all three kernel matrices. The results are shown as a receiver operating characteristics (ROC) curve in Figure 3(e).

The graph shows that the RBF kernel produces the best results for the data while the linear and polynomial kernels give the worse performance. These results are consistent with the observations made by Tax et al. [13]. They observed that the performance of the linear and polynomial kernels is heavily influenced by the norms of the data vectors, i.e., larger norm values of data vectors dominate other values in the data set. Therefore, larger radius values are produced for the hypersphere in the system, resulting in lower detection rates. Also, the better performance of the RBF kernel is due to its independence from the norms of the data vectors, as it only depends on the distance between them.

V. Conclusion

In this paper, we have proposed a distributed anomaly detection approach with low communication overhead, based on a one-class quarter sphere SVM, for wireless sensor networks. We evaluated our approach in a multi-level hierarchical topology, using real data gathered from a sensor network deployment in Great Duck Island. Our evaluation reveals that the distributed scheme achieves significant energy savings in terms of communication overhead in the network, while achieving a comparable performance to that of the centralised case. Our future research includes periodically adjusting the parameters in the system based on statistics from previous time windows, and evaluating detector performance in identifying a variety of sensor network attack scenarios.

Acknowledgment

We thank Dr. Adil Bagirov of University of Ballarat for advice on the complexity of linear programming, and Mr. Robert Szewczyk of University of California at Berkeley for giving access to the Great Duck Island data. We also thank ARC Research Network on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), and DEST International Science and Linkage Grant. This work was supported by the Australian Research Council (ARC).

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] A. Perrig, J. Stankovic, and D. Wagner, "Security in wireless sensor networks," *CACM*, vol. 47, no. 6, pp. 53–57, 2004.

[3] N. R. Prasad and M. Alam, "Security framework for wireless sensor networks," *Wireless Personal Communications*, vol. 37, no. 3-4, 2006.

[4] V. Barnett and T. Lewis, *Outliers in Statistical Data*. John Wiley and Sons, 3rd ed., 1994.

[5] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Commun. ACM*, vol. 43, no. 5, pp. 51–58, 2000.

[6] V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava, "Energy aware wireless microsensor networks," in *IEEE Signal Processing Magazine*, March 2002.

[7] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative signal and information processing: an information-directed approach," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1199 – 1209, 2003.

[8] T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, "Distributed deviation detection in sensor networks," *SIGMOD Rec.*, vol. 32, no. 4, pp. 77–82, 2003.

[9] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, "Online outlier detection in sensor data using non-parametric models," in *VLDB*, pp. 187–198, VLDB Endowment, 2006.

[10] I. Onat and A. Miri, "An intrusion detection system for wireless sensor networks," in *Wireless And Mobile Computing Networking And Communications*, vol. 3, pp. 253–259, August 2005.

[11] C. L. C. Loo, M. Ng and M. Palaniswami, "Intrusion detection for routing attacks in sensor networks," vol. 2, pp. 313–332, October-December 2006. ISSN 1550-1329.

[12] B. Scholkopf and A. Smola, *Learning with Kernels*. MIT Press, 2002.

[13] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[14] A. Navia-Vazquez, D. Gutierrez-Gonzalez, E. Parrado-Hernandez, and J. Navarro-Abellan, "Distributed support vector machines," *IEEE Trans. on Neural Networks*, vol. 17, no. 4, pp. 1091–1097, 2006.

[15] B. B.-L. K. Flouri and P. Tsakalides, "Training a svm-based classifier in distributed sensor networks," in *EUSIPCO*, (Florence), 2006.

[16] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, "Distributed anomaly detection in wireless sensor networks," in *IEEE International Conference on Communications Systems (IEEE ICCS)*, (Singapore), October 2006.

[17] P. Laskov, C. Schafer, and I. Kotenko, "Intrusion detection in unlabeled data with quarter sphere support vector machines," in *Detection of Intrusions and Malware & Vulnerability Assessment*, (Dortmund), 2004.

[18] V. N. Vapnik, *Statistical Learning Theory*. John Wiley & Sons, 1998.

[19] S. G. Nash and A. Sofer, *Linear and nonlinear programming*. McGraw-Hill, 1996.

[20] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[21] B. Scholkopf, A. J. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[22] M. Epelman and R. M. Freund, "A new condition measure, preconditioners, and relations between different measures of conditioning for conic linear systems," *SIAM J. on Optimisation.*, vol. 12, no. 3, pp. 627–655, 2002.

[23] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin, "Habitat monitoring with sensor networks," *CACM*, vol. 47, no. 6, pp. 34–40, 2004.

[24] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in *Intl. Conference on Embedded Networked Sensor Systems*, pp. 214–226, 2004.

[25] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, and D. Tax, "Prtools4, a matlab toolbox for pattern recognition, delft university of technology"',' 2004. http://www.prtools.org/.

[26] D.M.J.Tax, "Ddtools, the data description toolbox for matlab,version 1.5.4," Sept 2006. http://ict.ewi.tudelft.nl/davidt/dd_tools.html.