Volume 3, Issue 7, July 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Outlier Detection in WSN- A Survey

Ritika* | *M.M.E.C, Mulana* | *India*
Tarun Kumar | *G.I.M.T, Kanipla* | *India*
Amandeep Kaur | *M.M.E.C, Mulana* | *India*

Abstract— In the field of wireless sensor networks, the measurements that deviate from the normal behaviour of sensed data are taken to be as outliers. The potential sources of outliers can be noise and errors, events, and malicious attacks on the network. This paper give an overview of existing outlier detection techniques specifically developed for the wireless sensor networks. Also, a technique-based taxonomy will be discussed with characteristics of outlier data such as data type, outlier type, outlier identity, and outlier degree.

Keywords--

I. INTRODUCTION

The term outlier, also known as anomaly, originally stems from the field of statistics (Hodge and Austin, 2003). The two classical definitions of outliers are:
(Hawkins 1980): "*An outlier is an observation, which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism*".
(Barnett and Lewis, 1994): "*an outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data*".
Additionally outliers can be defined as, "*those measurements that significantly deviate from the normal pattern of sensed data*" [21]. This definition is based on the fact that in WSN, SNs are assigned to monitor the physical world and thus a pattern representing the normal behavior of sensed data may exist. Potential sources of outliers in data collected by WSNs include noise & errors, actual events, and malicious attacks.
Recently, the topic of outlier detection in WSNs has attracted much attention. According to potential sources of outliers as mentioned earlier, the identification of outliers provides data reliability, event reporting, and secure functioning of the network. Here, we exemplify the essence of outlier detection in several real-life applications.

- Environmental monitoring
- Habitat monitoring
- Health and medical monitoring
- Industrial monitoring
- Target tracking
- Surveillance monitoring

It should be noted that several research topics have been developed for identifying sources of outliers occurred in WSNs.

A. Types of outliers

Compared to a centralized approach where all the outliers are determined at the central node, outlier detection in a distributed approach can be done at the network nodes individually as well as at the sink node. This is the concept of multi-level outlier detection [22]. In multilevel outlier detection each node can determine the outliers locally using the sensed data stream. A simple classification of different types of outliers is given below.

- Local Outliers or First Order Outliers: The existence of these types of outliers means that some of the observations at a SN are anomalous with respect to the rest of data. The Local Outliers are also known as First Order Outliers. The First order outliers are further classified into following categories:
  - Type 1
  - Type 2
  - Type 3
  - Type 4

Figure 1 shows various types of local outliers that may be encountered in a data set collected from a WSN.

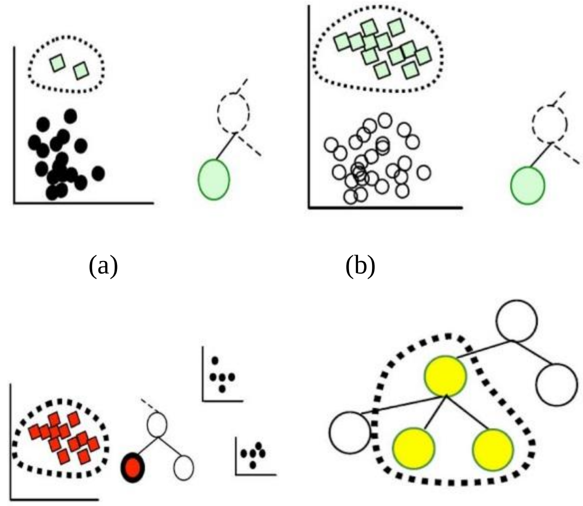
- Global Outliers or Higher order Outliers: There are two different types of global outliers [1, 3, 2]. First, all of the data at a SN may be anomalous with respect to neighboring nodes. These types of outliers are called second order external outliers. In this case, a SN will be identified as an anomalous node. Second, a set or a subtree of SNs in the network may be anomalous.

These are known as third order external outliers [1]. Second and third order outliers are collectively known as Higher Order (HO) external outliers [2]. Identification of global outliers can be performed at different levels in a hierarchical network, depending upon the network architecture.

B. Outlier Identity

There are three potential outlier sources in WSNs:

- Noise and errors
- Events
- Malicious attacks



(a) (b)

(c) (d)

Fig. 1 (a) First order outliers. Some of the measurements are anomalous with respect to others. In the plot, squares represent the abnormal measurements. (b) First order epoch outliers (Type-4 local outliers). (c) Second order external outliers. All measurements of a SN are anomalous with respect to neighboring nodes. (d) Third order external outliers. A subset/subtree of nodes is anomalous with respect to neighboring nodes in the network.

As illustrated in Figure 2, these topics include fault detection [21], event detection [25] and intrusion detection.

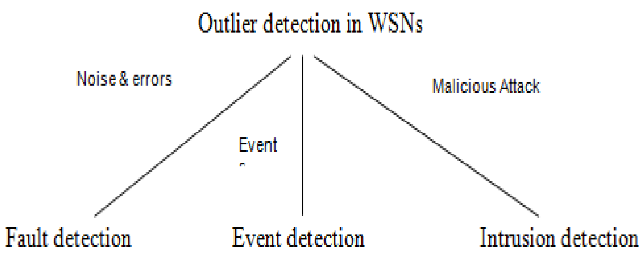


Fig. 2 Outlier sources in WSNs and their corresponding detection techniques

C. Number of determined outliers

Techniques can also be classified based on the number of outliers they determine.

- Single Outliers
- Multiple Outliers

These topics include fault detection [21], event detect- ion [25] and intrusion detection.

D. Degree of being an outlier

In WSNs, outliers are measured in two scales.

- Scalar
- outlier score [8]

II. Literature Survey

A. Techniques Designed For WSNs

Recently, many outlier detection techniques specifically developed for WSNs have emerged. In this section, we provide a technique-based taxonomy framework to categorize these techniques. As illustrated in Figure 3, outlier detection techniques for WSNs can be categorized into different categories.

B. Classification of the state-of-the-art outlier detection

Outlier detection techniques can be classified based on the following approaches used for outlier detection:

- Distance based
- Normal state model based (Machine Learning based)

C. Analysis of Distance based Outlier Detection techniques for WSNs

Distance based techniques, as shown in Figure 3, use some distance measure from statistical distribution, nearest neighbor or clusters to detect the outliers. We classify the distance based techniques into the following types:

- Statistical based
- Nearest Neighbor based
- Clustering based

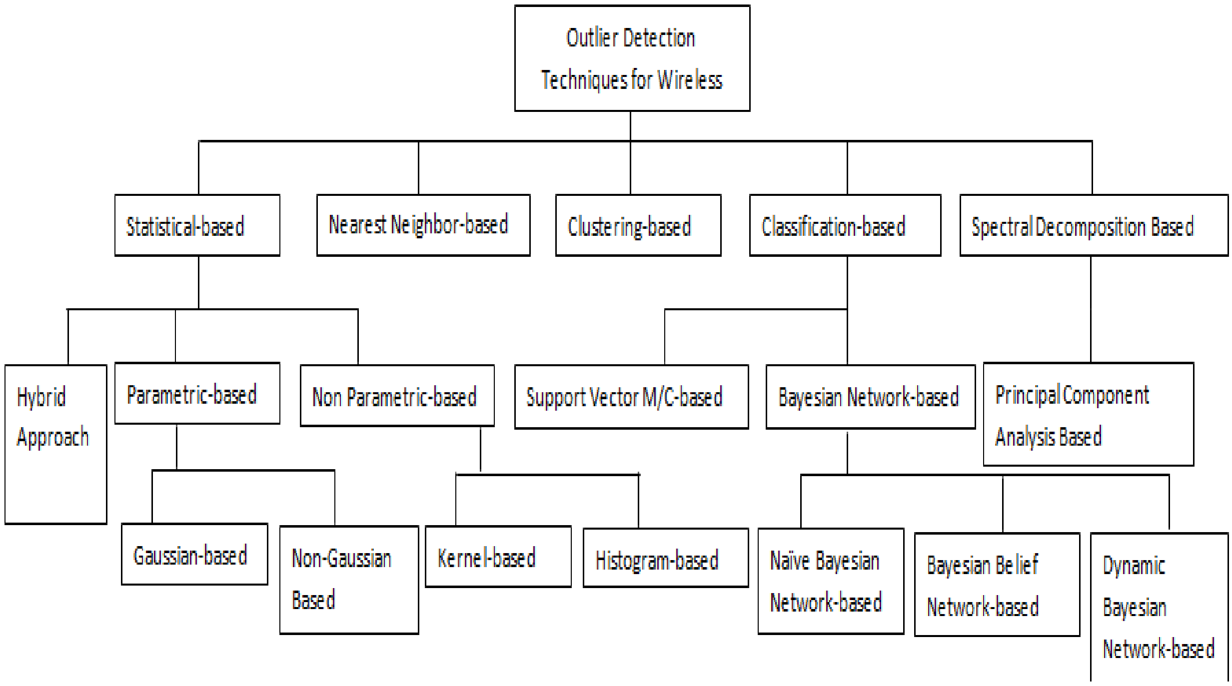


Fig 3: Taxonomy of outlier detection techniques for WSNs

1) Statistical Outlier Detection Techniques

These techniques require an underlying data distribution model for the detection of outliers. They assume or estimate a statistical (probability distribution) model which captures the distribution of the data and evaluate data instances with respect to how well they fit the model. A data instance is declared as an outlier if the probability of the data instance to be generated by this model is very low, based on the distance measure. These techniques can be classified as parametric or non-parametric techniques.

a) Parametric Outlier Detection Techniques

Parametric techniques assume availability of the knowledge about underlying data distribution, i.e., the data is generated from a known distribution. Distribution parameters are then estimated from the available data.

- Gaussian based models: A semi-supervised and localized outlier detection and event boundary detection technique that is based on two algorithms is presented in [18]. Each sensor is able to locate its physical position using GPS or GPS-less techniques. This algorithm is different from the related work [15], in that it uses numeric input data, as compared to 0/1 decision predicates, where 1 indicates the occurrence of some phenomena and 0 indicates a normal status. Another semi-supervised technique [34] focuses on making the aggregation operations reliable under the presence of outliers.
- Non-Gaussian based models: Jun et al. [31] presents a statistical-based technique, which uses a symmetric alpha stable distribution to model outliers being in form of impulsive noise.
- Hybrid Outlier Detection Techniques: Hybrid statistical [15] is a technique which presents a, semi-supervised, localized local outlier detection method to identify errors and detect events in ecological applications of WSN.

b) Non-Parametric Outlier Detection Techniques

Non-parametric techniques do not assume availability of data distribution. They typically define a distance measure between a new test instance and the statistical model and use some kind of thresholds on this distance to determine whether the observation is an outlier.

- Histogram based approaches: [16] presents a non-parametric, histogram based and distributed approach to the global outlier detection. Two algorithms have been given based on two different definitions of outliers (O(n, k) and O(d, k)). Both of these algorithms are based on the distance of a data point to its k-nearest neighbors.

Outliers can be identified by collecting more histogram information from the network. Various approaches to the approximation of density using histograms have been discussed in the literature like [33].

- Kernel function based approaches: Like histogram based approaches, [37] proposes a kernel based, localized approach for online identification of outliers in streaming sensor data.

The shortcomings of [37] are overcome in [23]. In this technique, Kernel density function has been used to approximate the data distribution, like [37]. This approach is based on efficient, in-network approximation of the input data distributions and can be efficiently extended to more than one dimension. Various approaches to hierarchical decomposition have been given in the literature [32, 42, 38]. These approximations can also serve other purposes, such as online estimation of range queries. This approach is versatile, in that, it presents two different algorithms for identifying outliers, the first algorithm is distance based and hence fast, whereas the second algorithm is more robust and employs a local metrics based approach [39]. "Chain-sample" algorithm has been used for producing a uniform random sample of the data set from which to estimate the density and histogram approach (along the time axis) has been used for the approximation of standard deviation of the distribution, as done in [40]. In multi granular local metrics based approach, it is not sufficient for the parent node to only test for the outliers declared by the leaf nodes, as evident from the definition in [39].

2) Nearest Neighbor based Outlier Detection Techniques

Nearest neighbor-based approaches are the most commonly used approaches to analyze a data instance with respect to its nearest neighbors in the data mining and machine learning community. Work done in [24] proposes an unsupervised (non-parametric), non-hierarchical and distributed system architecture based outlier detection technique that is generic, suitable for many outlier heuristics like distance to k^{th} nearest neighbor , average distance to k nearest neighbors , inverse of the distance to k nearest neighbors etc, works in-network with a communication load that is proportional to the outcome, is robust with respect to data and network change and the outcome of this method is revealed to all the sensors. The technique discussed in [64] has a large communication overhead and does not define an ending condition properly. Zhuang [22] presents an unsupervised, non-parametric, distance based approach to perform in-network outlier cleaning of outliers.

3) Clustering based Outlier Detection Techniques

Grouping similar data instances into clusters with similar behavior is known as clustering. Clustering algorithms can be either centralized or distributed. In centralized clustering algorithms each node transfers its data vectors to the gateway/ central node which then performs data clustering.

A distributed anomaly detection technique based on clustering is presented in [25]. This technique assumes hierarchical network structure, consisting of parent, children and sibling nodes and one global sink node. Each node then uses this global data to condition its local data. Each node then runs fixed width clustering algorithm, which creates a set of clusters with fixed width. Clustering is then performed based on Euclidean distance between the data vectors. Two clusters can be merged if the centroids of these clusters are less than the chosen width apart. Each cluster then calculates its average inter cluster distance based on only the k nearest neighbors, as compared to the conventional approach that considers all the neighbors for computing this distance. If the average inter-cluster distance of any cluster is greater than one standard deviation of mean of inter-cluster distances of all the clusters, it is declared as an outlying cluster. This technique greatly reduces communication and computation overheads as well as the memory requirements as the data values do not have to be remembered by the nodes, but a major drawback of this technique is the fixed width threshold, which is not easy to choose.

D. Outlier Detection Techniques for WSNs based on Normality Model (Machine Learning based)

Machine learning based approaches involve the estimation or inference of the next state of the system based on previous measurements. These techniques either learn or predict a normality model of data, to classify the newly arrived measurements as normal or anomalous. Three types of such techniques exist:

1) Prediction based Outlier Detection Techniques (Kalman Filter)

These techniques predict the normal model of the data based on filtering methods. A Kalman filer based approach is explained below. A kalman filter approach for outlier detection is presented in [9]. This technique exploits spatial and temporal dependencies among sensor data for in-network outlier detection. Kalman filter is an efficient recursive filter that estimates the state of a dynamic system, such as the temperature of habitat environment, from a stream of noisy measurements. No more history of observations is required. KF achieves optimal estimation of the state (the state can be considered as sample of a particular sensory attribute of a system with white noise disturb). The accuracy of outlier detection can be expected to be higher than other methods and does not require much computation and large storage, and is suitable for implementation on resource constrained SNs. The common least squares method is based on the Yule-Walker equations [11].

2) Classification based Outlier Detection Techniques

Classification based techniques learn a classification model using the set of data instances and then classify the data instance to one of the training classes. This is called training phase of the classifier. These techniques can be either supervised or unsupervised. The one-class unsupervised techniques learn the boundary around normal instances during training while some anomalous instance may exist and declare any new instance lying outside this boundary as an outlier.

a) Support Vector Machine based

A support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. A kernel function is used to approximate the dot product between the mapped vectors for identifying the hyperplane.

A brief review of SVM based techniques is given below. However, SVM based techniques are a major portion of this work and will be treated specially in the forth-coming chapters. Rajasegarar [17] presented a quarter-sphere SVM based approach for distributed anomaly detection in hierarchical or non-hierarchical network structures. The major motivation of this work is the fact that communication cost is much higher than computation cost.An online and local outlier detection technique with low resource consumption based on an unsupervised (one-class) centered quarter-sphere Support Vector Machine (SVM) for environmental monitoring applications of WSNs is presented in [10]. This approach takes advantage of spatial correlations that exist in sensor data of adjacent nodes to reduce the false alarm rate and to distinguish between events and errors in real-time and overcomes the disadvantages of [17]. An online and adaptive local outlier detection technique is presented in [7]. This technique is identical to [10], in that, it identifies outliers based on spatiotemporal deviations.

b) Bayesian based

These techniques are actually a probabilistic graph model. Bayesian approaches involve three phases, learning, testing and inference. Learning phase involves the modeling of all the states of data via a training data. There are different approaches, centralized as well as distributed to train data in the learning phase. The trained data is tested until correct results are obtained and then it is used for inference.

- Naive Bayesian models: Elnahrawy [35] presented an online, in-network and distributed approach to identify local outliers and missing values. The technique uses contextual information learned by the network model to predict the correct class of the next sensor reading and then compare it with the sensed reading. This technique is also applicable to centralized approach; however a centralized approach will pose a lot of computation and communication overhead for non-stationary conditions.
- Bayesian Belief Network models: The technique presented in [26] does not take into account attribute correlations and multivariate data. Moreover it is inefficient because of the naive Bayes assumption. Janakiram [26] proposed a technique that takes into consideration multivariate data and attribute correlations as well as the conditional dependency of the feature vectors over current sensor reading. Spatial data is considered in this technique as well as in, by considering Markov chain assumptions [44, 20] and time series data by Markov random fields [41]. This technique determines local outliers via collaboration of neighboring nodes.
- Dynamic Bayesian Network models: The Bayesian techniques discussed in above do not take into account the dynamic nature of network. Hill considered a technique [19] which takes into account the dynamic changes in a network. Rao-Blackwellized particle filtering [36] is used to sequentially infer the posterior distributions of the hidden and observed states as new measurements become available from the sensors. Classification based approaches provide an exact set of outliers. However, a main drawback of these techniques is that their computational complexity is high.

3) Spectral decomposition based

These techniques aim at finding normal modes of behavior by using Principal Component Analysis (PCA) [43]. PCA is used to reduce the dimensionality before outlier detection.

An approach that fuses data gathered from different nodes in distributed WSNs is presented in [27]. It provides an integrated methodology of taking into consideration and combining effectively correlated sensor data, in a distributed fashion, in order to reveal anomalies that span through a number of neighboring nodes. A PCA based technique is used to solve the data integrity and accuracy problem caused by compromised or malfunctioning nodes. The technique has an offline phase and an online phase.

The online mode applies Subspace method [14] to divide the data into two different spaces: one containing readings that are considered normal and resemble to the modeled data patterns and one containing the residual. When an anomaly occurs the residual vector presents a great variation in some of its variables and the system detects the path containing the anomaly by selecting these variables. Squared Prediction Error (SPE) [4] is used for detecting the abnormal conditions.

When an anomaly occurs SPE exceeds the normal thresholds and the system detects the set of nodes containing the anomaly, by selecting the variables that contribute mostly to large change of the SPE. This technique considers multivariate data and uses spatio-temporal correlations to detect local outliers.

Figure 4 shows the flowchart of different techniques.

- DFD Approach: DFD scheme determines [5] the status of node by testing among neighbor nodes mutually. For two neighbor nodes s_i and s_j, a test result $c_{i,j}$, is produced by the data (such as temperature) sensed by each of them. The data at the moment t should be very close to each other because they are near, and the difference d between this data should not exceed a certain threshold θ_1.

Disadvantage: 1. this scheme works only for large number of neighbor nodes.
2. This scheme does not work when there are large number of faulty nodes.

- Reliable and Energy efficient schedule (REEF): Reliable and Energy efficient schedule in [6] which each CH runs an outlier detection algorithm to detect outliers among readings reported by its members. REEF uses "Spatial Averaging Algorithm" for outlier detection. This algorithm uses the median which is statistically robust to outliers.
  Disadvantage: It is difficult to compute the correct value of threshold.
- History based approach: There is enough number of data entries. It compares current status of node with the previous results. This approach has low false alarm rate. Also, this approach has reliable fault detection rate Examples of History based approach are SVD and PVD.
  Disadvantage: This approach will not give accurate results for less number of data entries

- Non-history based Approach: There is insufficient number of data in motion database. In order to increase in accuracy of posterior probability, we need a sufficient amount of knowledge.
  Disadvantage: It is not easy to find out a right maximum and minimum boundary of a threshold value.

- Modified z-score method: In a Modified z-score test, the z-score is calculated in [13] based on outlier resistant estimators. The Median of Absolute Deviation about the median (MAD) is such an estimator.
  MAD = median {|xi – xm|} (1)
  In z-score calculations, standard deviation is used to detect outliers whereas in Modified z-score, MAD is used in the place of standard deviation.
  Disadvantage: 1. It involves a slightly higher computational overhead i.e. O(nlogn), where n is the number of sensor data involved in the aggregation process.
  2. It should also be noted that significant deviation in the sensed values might also occur due to a transient or persistent change in the phenomenon being monitored. That's why in many techniques we are using spatio-temporal correlation among the nodes to differentiate between the cases where outliers are produced by faulty measurements/nodes and those that are produced due to a fundamental shift in the phenomenon.
- Classical majority voting: Each sensor (e.g., sensor s) in the witness set makes a judgment by comparing its own reading with the unusual reading in [12] sent by the suspected sensor (e.g., sensor s).
  Disadvantage: However, this simple majority voting approach does not work well when the number of faulty sensors increases.
- Distance weighted voting: weighted voting methods have been proposed in the literature [12]. Motivated by an assumption that the closer sensors have more resembled readings, the weighted voting algorithms give more weights to closer neighbors in voting (i.e., the weights are assigned inverse to the distances from a SN to its neighbors). However, they argued that the distance between two sensors does not fully represent the correlation between readings of those two sensors Furthermore, if the nearest sensor is faulty, the voting result may be seriously contaminated by this faulty sensor. They referred to this problem as a domination problem.
  Disadvantage:
  1. It does not precisely capture about the correlation between sensor readings.
  2. It is a good idea to inquire opinions of neighbors, the trustworthiness of neighbors is not considered.

- An Interleaved Hop-by-Hop detection Scheme: This scheme involves the following five phases in [28].
  1. Node initialization and deployment phase.
  2. Association discovery phase
  3. Report endorsement phase
  4. En-route filtering phase
  5. BS verification phase
  Disadvantage: This scheme does not work well when there are more number of outliers node is in the network.
- Statistical En Route Filtering (SEF) Scheme: SEF consists of three components that work in [29] concert to detect and filter out forged messages:
  1. Each legitimate report carries multiple MACs generated by different nodes that detect the same stimulus.
  2. Intermediate forwarding nodes detect incorrect MACs and filter out false reports en-route.
  3. The sink verifies the correctness of each MAC and eliminates remaining false reports that elude en-route filtering.
  In SEF, the sink maintains a global key pool. Each sensor storks a small number of keys that are drawn in a randomized fashion from the global key pool before deployment.

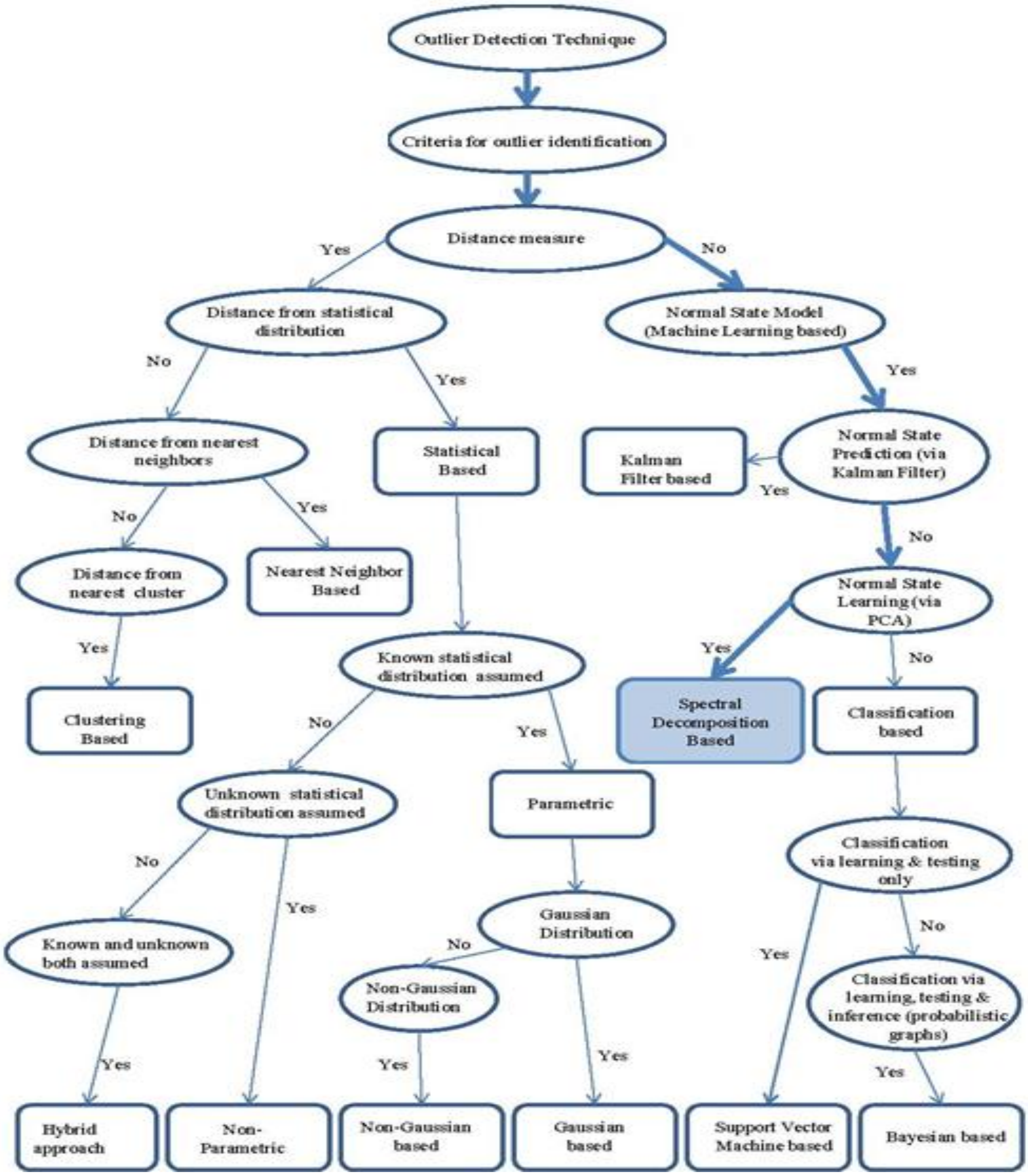


Figure 4: Flowchart of different outlier techniques

- Commutative Cipher Based En-route Filtering (CCEF) Scheme: This scheme in [30] defends against event fabrication attacks without Symmetric key sharing among SNs.

Disadvantage: Computation becomes complex.

- Distribution-based method: These methods are typically found in statistics textbooks. They deploy some standard distribution model (Normal, Poisson, etc.) and flag as outliers those data which deviate from the model.

Disadvantage:
1. We need large sample of data to get accurate results.
2. This approach is uniformly distributed.

- Using SensorRank Scheme: Thus, in the proposed algorithm in [12], each SN is associated with a trustworthiness value (called Sensor-Rank) that will be used in voting.

III. Conclusion

We have studied the problem of outlier detection in WSNs and technique-based taxonomy framework which categorize the current outlier detection techniques designed for WSNs. Also advantages and disadvantages of some of the techniques are also given in the paper.

REFERENCES

[1] Y. Zhang, “Observing the Unobservable -Distributed Online Outlier Detection in Wireless Sensor Networks”, PhD thesis, University of Twente, June 2010.

[2] S. Rajasegarar, J. C. Bezdek, C. Leckie, and M. Palaniswami, “Elliptical anomalies in wireless sensor networks,” ACM Trans. Sen. Netw., vol. 6, pp. 7:1–7:28, January 2010.

[3] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, “Labelled data collection for anomaly detection in wireless sensor networks,” in Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2010 Sixth International Conference on, pp. 269 –274, December. 2010.

[4] J. E. Jackson and G. S. Mudholkar, “Control procedures for residuals associated with principal component analysis”, Technometrics, vol. 21, no. 3, pp. pp. 341–349, 1979.

[5] Peng Jiang, “A New Method for Node Fault Detection in WSNs”, in proceedings of Sensors (14248220), Vol. 9 Issue 2, p1282, February, 2009.

[6] Arun Kumar K.S, Vinay Joseph Ribeiro, “A Reliable and Energy Efficient Framework for Wireless Sensor Networks”, in proceedings of COMSNETS'09, pp. 604-612, 2009.

[7] Y. Zhang, N. Meratnia, and P. Havinga, “Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks,” in Proc. Int. Conf. Advanced Information Networking and Applications Workshops WAINA ’09, pp. 990–995, 2009.

[8] A. R. Ganguly, “Knowledge Discovery from Sensor Data”, CRC Press, 2008.

[9] M. Shuai, K. Xie, G. Chen, X. Ma, and G. Song, “A kalman filter based approach for outlier detection in sensor networks”, in Computer Science and Software Engineering, 2008 International Conference on, vol. 4, pp. 154 – 157, December 2008.

[10] Z. Yang, N. Meratnia, and P. Havinga, “An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine”, in Intelligent Sensors, Sensor Networks and Information. Processing, 2008. ISSNIP 2008. International Conference on, pp. 151 –156, December 2008.

[11] G. J. G. Box and G. Reinsel, “Time Series Analysis: Forecasting & Control”, 3rd Edition, Prentice-Hall, 1994.

[12] Xiang-Yan Xiao, Wen-ChihPeng and Chih-Chieh Hung, “Using sensor rank for in- network detection of faulty nodes in WSNs”, in proceedings of MobiDE’07, June 10, 2007.

[13] L.S.Jayashree, S.Arumugam and K.Vijayalakshmi, “A Robust Outlier Detection Scheme for Collaborative Sensor Networks”, in Journal of Digital Information Management, vol. 5 issue 1, February 2007.

[14] R. Dunia and S. J. Qin, “Subspace approach to multidimensional fault identification and reconstruction”, Aiche Journal, vol. 44, pp. 1813–1831, 1998.

[15] L. M. A. Bettencourt, A. A. Hagberg, and L. B. Larkey, “Separating the wheat from the chaff: Practical anomaly detection schemes in ecological applications of distributed sensor networks”, in Distributed Computing in Sensor Systems (DCOSS 2007), (Santa Fe, NM USA), pp. 223–239, June 2007.

[16] B. Sheng, Q. Li, W. Mao, and W. Jin, “Outlier Detection in Sensor Networks”, Proceedings of the Eighth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), 219-228, 2007.

[17] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. Bezdek, “Quarter sphere based distributed anomaly detection in wireless sensor networks”, in IEEE International Conference on Communications, 2007. ICC ’07, pp. 3864 – 3869, june 2007.

[18] W. Wu, X. Cheng, M. Ding, K. Xing, F. Liu, and P. Deng, “Localized outlying and boundary data detection in sensor networks”, Knowledge and Data Engineering, IEEE Transactions on, vol. 19, pp. 1145 –1157, august 2007.

[19] D. J. Hill, B. S. Minsker, and E. Amir, “Real-time bayesian anomaly detection for environmental sensor data”, in In Proceedings of the 32nd conference of IAHR, 2007.

[20] P. Smyth, “Belief networks, hidden markov models, and markov random fields: a unifying view”, Pattern Recognition Letters, vol. 18, pp. 1261–1268, 1998.

[21] Chandola, V., Banerjee, A. and Kumar, V., “Outlier detection: a survey”, Technical Report, University of Minnesota, 2007.

[22] Y. Zhuang and L. Chen, “In-network outlier cleaning for data collection in sensor networks”, in In CleanDB, Workshop in VLDB 2006, pp. 41–48, APPENDIX, 2006.

[23] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, “Online outlier detection in sensor data using non-parametric models”, in Proceedings of the 32nd international conference on Very large data bases, VLDB ’06, pp. 187–198, VLDB Endowment, 2006.

[24] J. Branch, B. Szymanski, C. Giannella, R. Wolff and H. Kargupta, “In-network outlier detection in wireless sensor networks”, in Distributed Computing Systems, 2006. ICDCS 2006. 26th IEEE International Conference on, pp- 51, 2006.

[25] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, “Distributed anomaly detection in wireless sensor networks”, in Communication systems, 2006. ICCS 2006. 10th IEEE Singapore International Conference on, pp. 1 –5, October 2006.

[26] D. Janakiram, V. AdiMallikarjuna Reddy, and A. Phani Kumar, “Outlier detection in wireless sensor networks using bayesian belief networks”, in Communication System Software and Middleware, 2006. Comsware2006, pp. 1 –6, 2006.

[27] V. Chatzigiannakis, S. Papavassiliou, M. Grammatikou, and B. Maglaris, "Hierarchical anomaly detection in distributed large-scale sensor networks," Computers and Communications, IEEE Symposium on, vol. 0, pp. 761–767, 2006.

[28] Sencun Zhu SanjeevSetiaSushilJajodia, "An Interlçaved Hop-by-Hop Authentication Scheme for Filtering of Injected False Data in Sensoi Networks", in proceedings of Computer Science Dept, George Mason University, 2005.

[29] Fan Ye, HaiyunLuo, Songwu Lu and Lixia Zhang, "Statistical En Route Filtering of Injected False Data in Sensor Networks", in proceedings of IEEE, April 2005.

[30] Hao Yang and Songwu Lu, "Commutative Cipher Based En-route Filtering in Wireless Sensor Networks", in proceedings of Computer Science Department, UCLA Los Angeles, CA 90095-1596, U.S.A., 2005.

[31] M. C. Jun, H. Jeong, and C.-C. J. Kuo, "Distributed spatio-temporal outlier detection in sensor networks", in Proc. SPIE 5819, Digital Wireless Communications VII and Space Communication Technologies, 273, doi:10.1117/12.604764, 2005.

[32] D. Ganesan, B. Greenstein, D. Estrin, J. Heidemann, and R. Govindan, "Multiresolution storage and search in sensor networks", ACM Transactions on Storage, vol. 1, pp. 277–315, 2005.

[33] H. V. Jagadish, V. Poosala, N. Koudas, K. Sevcik, S. Muthukrishnan, and T. Suel, "Optimal histograms with quality guarantees", in In VLDB, pp. 275–286, 1998.

[34] C. P. Yozo, Y. Hida, P. Huang, and R. Nishtala, "Aggregation query under uncertainty in sensor networks", tech. rep., 2004.

[35] E. Elnahrawy and B. Nath, "Context-aware sensors", in European Workshop on Wireless Sensor Networks, pp. 77–93, 2004.

[36] A. Doucet, N. d. Freitas, K. P. Murphy, and S. J. Russell, "Rao-blackwellised particle filtering for dynamic bayesian networks", in Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, UAI '00, (San Francisco, CA, USA), pp. 176–183, Morgan Kaufmann Publishers Inc., 2000.

[37] T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, "Distributed deviation detection in sensor networks", SIGMOD Rec., vol. 32, pp. 77–82, December 2003.

[38] S. Zhao, K. Tepe, I. Seskar, and D. Raychaudhuri, "Routing protocols for self-organizing hierarchical ad-hoc wireless networks", Proc. IEEE Sarnoff 2003 Symposium, March 2003.

[39] S. Papadimitriou, H. Kitagawa, P. Gibbons, and C. Faloutsos, "Loci: fast outlier detection using the local correlation integral", in Data Engineering, 2003. Proceedings. 19th International Conference on Data Engineering, pp. 315 – 326, march 2003.

[40] C. Olston, J. Jiang, and J. Widom, "Adaptive filters for continuous queries over distributed data streams", 2003.

[41] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations", San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 239–269, 2003.

[42] N. Malpani, J. L. Welch, and N. Vaidya, "Leader election algorithms for mobile ad hoc networks", in Proceedings of the 4th international workshop on Discrete algorithms and methods for mobile computing and communications, DIALM '00, (New York, NY, USA), pp. 96–103, ACM, 2000.

[43] I. Jollife, Principal Component Analysis. Springer, ISBN 978-0-387-95442-4, 2002.

[44] D.Hand, H.Mannila, and PSmyth, "Principles of Data Mining", MIT Press, 2001.

