RECOGNITION AND CLASSIFICATION OF TOLERANT NODES IN WIRELESS SENSOR NETWORK

Professor Siddarth Tiwari*

Dimit Vishwakarma*

Abstract:

This Research addresses the issue of dispersed oddity recognition in Wireless Sensor Networks (WSN). Programming such sensor nodes at a substantial scale can be a dull occupation if the framework is not deliberately composed and identify tolerant nodes in WSN. This kind of conveyed sensor displaying can be utilized as a part of a wide assortment of sensor systems, for example, distinguishing the vicinity of gatecrashers, identifying sensor disappointments, etcetera. The benefit of this methodology is that the human fashioner does not need to describe the odd marks ahead of time. The commitments of this methodology include:

1.Giving a route to a WSN to autonomously demonstrate sensor information with no former learning of the earth;

2.Empowering a disseminated framework to identify peculiarities in both sensor signs and fleeting occasions online:

3.Giving an approach to consequently separate semantic marks from fleeting groupings;

4.Giving an approach to WSNs to spare correspondence power by transmitting packed temporal arrangements;

5.Empowering the framework to identify time-related irregularities without earlier learning of strange occasions;

6.And,giving a novel missing information estimation

System that uses transient and spatial data to supplant missing qualities. The calculations have been composed, created, assessed, and approved tentatively in integrated information, and in certifiable sensor system applications.

Keywords: WSN, anomaly detection, framework

* Shree ram institute of science & technology Jabalpur(m.p)

I INTRODUCTION

An emerging class of Wireless Sensor Network (WSN) applications involves the acquisition of large amounts of sensory data from battery-powered, low computation and low memory wireless sensor nodes. Examples of such applications include volcano monitoring [6], animal habitat monitoring [8], structural monitoring [4], etc. These systems share the same goal of detecting "interesting" events in an unknown environment over a period of time. The systems must acquire data at a constant rate and transfer high-fidelity

data across a network. Moreover, the systems are also subject to unpredictable constraints on radio bandwidth, computational power, and energy usage over long periods of time. Given these constraints, it is typically not feasible to send all collected sensory data to a central location for processing and decision making. As a result, the sensor nodes should strive to process the raw sensor signal lo- cally and perform local decision making to determine the most "interesting" signals/events, such as detecting anomalous events. Local processing and decision making avoids wasting resources on "uninteresting" data, such as avoiding sending normal raw sensor readings to a human operator for interpretation. Currently, most research in the WSN area has focused on hardware design, self-organization, various routing algorithms, or energy saving patterns [2]. Practical distributed decision making algorithms for anomaly detection in a natural environment are still lacking.

In anomaly detection applications, a wireless sensor node in the network can monitor its local region and communicate through a wireless channel with other nodes to collaboratively produce a high-level representation of the environment. Using such a network, a large area can be monitored at a relatively low cost. The fundamental challenge is how to use the limited resources of wireless sensor nodes to deliver the most valuable information. The goal of this research was to develop a practical, scalable, autonomous and robust anomaly detection system that is able to detect time-related anomalies and impute missing data in an unknown environment using a WSN.

1.1 Wireless Sensor Network characteristics

The term *Wireless Sensor Network* refers to a network of small, low-cost, low-power devices that can sense, actuate, and communicate information about their environment. Below is a summarized list of some fundamental characteristics of a WSN [2], [6], [8]:

Application specific: WSNs are conceivable for various applications with different spatial deployments that range from being very sparse to very dense. In sparse deployment with non-overlapping sensing ranges, the sink (sensor node to which in- formation should be delivered) needs to collect sensory data from all sensor nodes in order to monitor the entire environment. On the other hand, if the sensor deploy- ment is dense with overlapping sensing ranges, approximations of neighboring nodes' sensory data can be used to save communication costs.

Responsive to the environment: WSNs generally have to respond to the environment; their traffic characteristics can be expected to be very different from traditional wired networks, mobile ad hoc networks, etc. WSNs are likely to have long periods (e.g., months) of inactivity that can alternate with short periods (e.g., seconds or minutes) of very high activity when an event of interest occurs.

Wireless ad-hoc in nature: WSNs are often required to self-configure into connected networks. One basic scenario includes a fixed topology of sensor nodes, together with a limited number of more powerful base stations, in which no maintenance or recharging is allowed after deployment. Therefore, cost minimization and autonomous behavior are desirable.

Physical distributed: WSNs are composed of a large number of nodes, each of which has an autonomous computational unit, and communicates with its neigh-bors via data packets. Data is also distributed throughout the nodes of the network and can be gathered at a central station only with high communication costs. Con-sequently, algorithms requiring global information from the entire network become very expensive.Scable to large numbers of nodes: WSNs have to scale to large numbers (hundreds or perhaps thousands) of nodes in order to cover a large geographical area. This requires scalable learning and communication structures.

Energy restricted: WSNs are usually powered by batteries; therefore, energy is a scarce resource. When there are no "interesting" events, it is better to reduce the activities of the sensor nodes.

II. RELATED WORK

Intrusion and outlier detection in WSNs

Many intrusion and outlier detection systems that are implemented in the area of WSNs focus on detecting network intrusion instead of detecting intruders in the physical environ- ment, e.g., [6],[7], [2], [5],. Existing detection systems either use a statistical based detection technique or a swarm intelligence-based technique. The works of [6] and [10] presented Intru- sion Detection Systems (IDS) for a sensor network that is based on the network activities (e.g., number of success and failure of authentications). The system compares event data with signature records to find harmful attacks from an intruder. Additionally, the authors of [7] applied the detection system in a cluster-based sensor network very much like the developed system in this dissertation. This type of detection system can only identify the anomalies that it has seen before. However, this research is interested in detecting anomalies in unknown environments, in which there are no abnormal prototypes available for the system to learn.

The authors of [5], [6], and [2], presented intrusion detection schemes that build a model of normal traffic behavior, and then use this model of normal traffic to detect abnormal traffic patterns. Their approaches are able to detect attacks that have not been previously seen. The detection system also has a feature selection phase where the features are specific to network traffic activities. The anomaly detection system presented in this dissertation takes a similar approach by building a normal model of the environment; sensor signals that do not match the normal models are considered as anomalies. However, this prior research does not address the issue of detecting time-related anomalies.

In [5], an ant colony based intrusion detection mechanism that could keep track of the intruder trails is presented. This technique can work in conjunction with conventional machine learning based intrusion detection techniques to secure sensor networks. This work tracks the paths of intrusion after anomalies are detected. Tracking the path of an intruder is one of the future directions of this research. However, it is not included in the scope of this dissertation.

Various machine learning techniques are used in WSNs. However, many of them have focused on improving the communication protocols of the WSNs. In general, they usually use one of the following three techniques: swarm intelligence, reinforcement learning or statistical-based learning.

The most widely used routing protocol that uses ant colony optimization is AntNet [2], which is an online Monte Carlo technique. There are many variations of this work, such as AntHocNet [3], ARS [9], MANSI [5], [7], etc. All of these algorithms use the notion of stigmergy, which is indirect communication that takes place among individuals through modifications induced in their environment. Moreover, attempts to solve classification or detection problems have been carried out using this optimization technique [4]. However, the classifiers based on ant- colony optimization generally extract a set of class rules from the data. This process requires knowledge of the environment, and needs to be carried out offline.

Time-related analysis in WSNs

Various regression models have been presented for time-related analysis WSNs, such as autoregressive models [6], Least-Square-Error based Linear forecast method [11], and the Non-Seasonal Holt-Winters Linear forecast method [12]. Most of these systems are linear regression models; these time series models have been widely used outside the wireless sensor network domain as a way to approximate and summarize time series with applications in finance, communica- tion, weather prediction, and a variety of other areas. However, as mentioned earlier, using regression models to detect time-related anomalies normally involve a complex parameter estimation process. They may suffer from model mismatch and bias-variance trade-off problems.

There has been some work on the use of probabilistic time-series models in WSNs, e.g., Kalman Filter-based models. These systems rely on a combination of local and global prob- abilistic models, which are kept in synchronized to reduce communication between sensor nodes and the network sink [2], [4], [6]. Moreover, Kalman Filter-based models are sophisticated and require significant computation.

The fixed length Markov model is another commonly used technique for time series analysis [2]. Examples of fixed order Markov models include the Markov Chain, Hidden Markov Model, etc. Due to the limited resources in WSNs, building high and fixed order Markov models is not feasible. Mazeroff et al., [8] implemented Variable Memory Markov model (VMM) in the forms of Probabilistic Suffix Tree (PST) models and Probabilistic Suffix Automata (PSAs) to build models of benign application behavior with the goal of detecting malicious applications in Windows XP. Note that in practice, the VMM is usually implemented in the form of either a PST or a PSA model. The two models are proven to be equivalent [5] and a PSA model can be

inferred directly from a PST model by using the algorithm described in [7]. PST models depend on a fixed number of random variables; in PST models this number of conditioning random variables may vary based on the specific observed realization.

III PROPOSED WORK

In WSN, the dominating communication pattern is that a large number of sensor nodes deliver their sensed information to one or a few data sinks through multi-hop transmission. This kind of communication pattern causes a drastic imbalance to the traffic load distribution across the network in which the nodes close to a sink experience heavy traffic loads. Since communication is believed to dominate the energy consumption of a sensor node and sensor nodes are usually provided with limited energy resources, the imbalanced traffic load distribution is very harmful and it could cause the nodes close to a sink to die at an earlier stage which thus renders the remainder of the network to be useless. To counter or alleviate the harm resulting from an uneven traffic load distribution, many researchers have turned their attention to the problem of load balancing. The authors of realize that the imbalanced traffic load distribution can cause one part of nodes to die earlier than the others, thus degrading the network performance. To counter the negative effect of the imbalanced traffic load distribution on network performance, new routing algorithms which resort to the measure of the remaining energy reserves and other kinds of path capacity measurements are proposed. consider the load balancing problem of uniformly distributed traffic demands in a unit disk.

By deliberately routing traffic along slightly longer paths instead of the shortest paths, the highly congested links are avoided and a particularly flat traffic load distribution is achieved. The problem of balancing the traffic load in multi-hop wireless networks with uniformly distributed point-to-point communication. They develop a routing algorithm called Curveball Routing which can avoid the crowded center and provide a performance which is not significantly worse than that of the optimum for malicious traffic is detected on the bases of routing table.

Our proposed work has following key point to be solved-

a. Evaluation and Detection of intrusion in the routing protocols (Table driven and dynamic) in sensor network.
b. Proposed solution based on three parameters for best routing selection (less overhead of energy)

i. **Minimum Delay of the path**

ii. **Maximum packet delivery ratio**

iii. **Max residual energy remains**

These 3 parameters will be the core metrics for proposed routing algorithm and for isolation of malicious traffic and node will be done by using SVM (support vector machine).

Machine Learning based Anomaly detection

Machine learning can be defined as the ability of a program and/or a system to learn and improve their performance on a certain task or group of tasks over time. Machine learning aims to answer many of the same questions as statistics or data mining. However unlike statistical approaches which tend to focus on understanding the process that generated the data, machine learning techniques focus on building a system that improves its performance with experience.

The proposed algorithm as follow:

Steps of Implementation

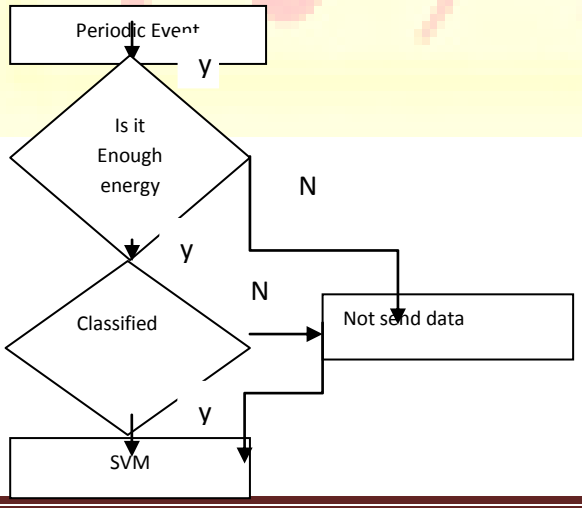
Step1: create the Wireless sensor network using NS-3 simulator and start simulation.

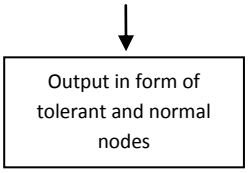
Step2: Generate routing table of each node in WSN with respect to energy consumed and remaining energy of each node.

Step3: input the routing table of each node into SVM for detection of tolerant nodes on the bases of PDR, PMIR, PMR parameters.

Step 4: Generate training data for SVM to perform classification of nodes as tolerant and non tolerant

Step 5: SVM generate classified nodes for better isolation on linear plane.





The primary problem with network based anomaly detection is that a globally accepted norm for what constitutes a normal profile does not exist. Maxion et al. [11] characterized the normal behavior in a network by using different templates that were derived by taking the standard deviations of Ethernet load and packet count at various periods in time. An observation was declared anomalous if it exceeded the upper bound of a predefined threshold.

On the basis of Table 1 calculations we will classify nodes as energy less and normal node it will improve the overhead in classification approach in this table we summarized the result for node id 1 PMIR ratio is 0 indicate number of miss route packets are less this will conserve the energy and we have categorize nodes if PDR is greater and PMR and PMIR should be less for energy less and normal node classification based on SVM.

Node id	PDR	PMR	PMIR	DELAY
1	90%	10%	0	50ms
2	2%	0	0	15ms
3	30%	60%	8%	80ms
4	5%	0	0	10ms
5	10%	0	90%	60ms

Table 1

Table 2 indicates the number of parameter used by base paper and our approach for increasing lifetime and improving QOS parameter for Wireless sensor network

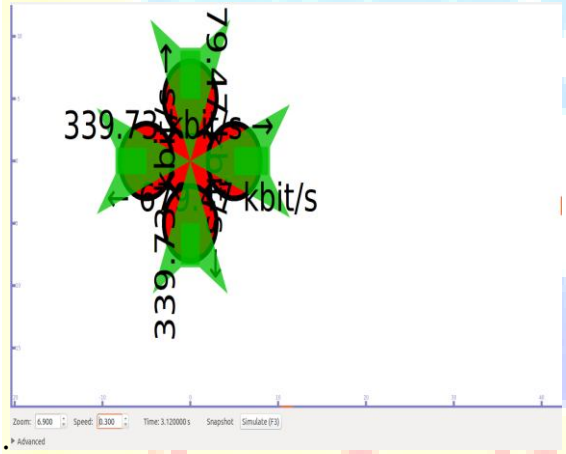
Parameters	PDR	PMR	PMIR	DELAY	Energy	Classification Technique
Base Paper	No	No	No	No	Yes	No
Our Approach	Yes	Yes	Yes	Yes	Yes	Yes (SVM)

Table 2

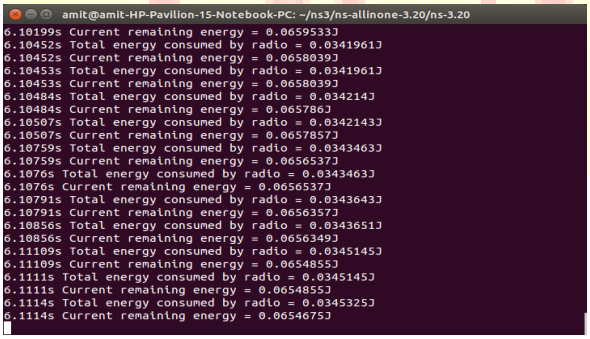
IV SIMULATION AND RESULT

We have done simulation in NS-3 and *ns-3* has been developed to provide an open, extensible network simulation platform, for networking research and education. In brief, *ns-3* provides models of how packet data networks work and perform, and provides a simulation engine for users to conduct simulation experiments. Some of the reasons to use *ns-3* include to perform studies that are more difficult or not possible to perform with real systems, to study system behavior in a highly controllled, reproducible environment, and to learn about how networks work. Users will note that the available model set in *ns-3* focuses on modeling how Internet protocols and networks work, but *ns-3* is not limited to Internet systems; several users are using*ns-3* to model non-Internet-based systems.

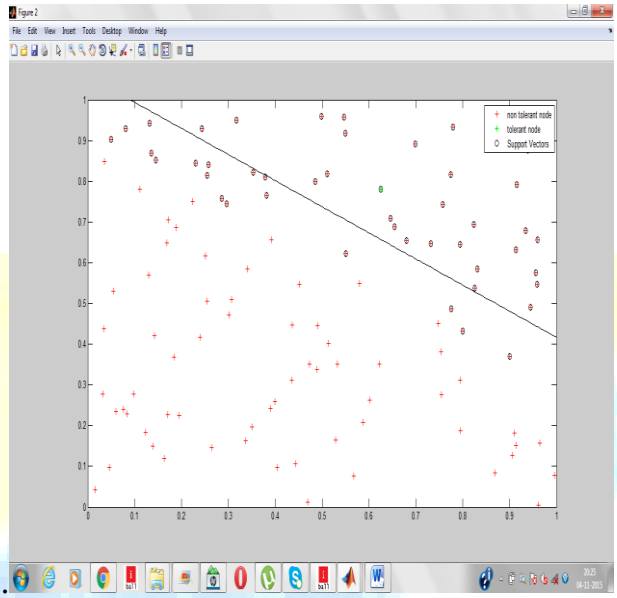
I.This shows number of sensor nodes deploying in simulation area for communication and making topology between nodes shown in green color and nodes in red color



II. this shows the generation of routing table during simulation



III. This is the final result of Support vector Machine which shows linear classification of tolerant and non tolerant nodes based on PDR,PMIR,PMR parameters.



V. CONCLUSION

This paper makes several contributions to anomaly detection in an unknown environ- ment using Wireless Sensor Networks. The most important contribution is the design of anomaly detection a novel approach that detects both sensor anomalies and time-related anomalies in an online, unsupervised, and distributed fashion. The distributed learning algorithms used are particularly suitable in a hierarchical, resource-constrained sensor net- work for environment monitoring purposes. The system utilizes various machine learning algorithms to model normal sensor data; this model is then used to detect anomalies. In contrast to most learning algorithms, the algorithms that are developed do not require large computational time or space and maintain high quality system performance. In some versions of this work, a mobile robot serves as the root clusterhead of the sensor network. Upon detection of an anomaly, this mobile robot can respond to further investigate the anomaly. In addition, the learning algorithms can take advantage of time and/or space correlations in the sensor data.

VI. REFERENCES.

[1] Intrusion Tolerance Mechanisms Using Redundant Node for Wireless Sensor Networks, 978-1-4799-2/14/2014 IEEE

[2]. Tolerating noisy, irrelevant and novel attributes in instance- based learning algorithms. *International Journal of Man-Machine Studies*, 36:267–287.

[3] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422.

[4] Amelia II (2010). Missing data imputation software. http://gking.harvard.edu/amelia/.

[5] Apostolico, A. and Bejerano, G. (2000). Optimal amnesic probabilistic automata or how to learn and classify proteins in linear time and space. In *Proceedings of the fourth annual international conference on Computational molecular biology*, pages 25–32, Tokyo, Japan. ACM.

[6] Banerjee, S., Grosan, C., and Abraham, A. (2005). Ideas: intrusion detection based on emotional ants for sensors. In *The 5th International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 344–349, Wroclaw, Poland.

[7] Barrenetxea, G., Ingelrest, F., Schaefer, G., and Vetterli, M. (2008). Wireless Sensor Networks for Environmental Monitoring: The SensorScope Ex- perience. In *The 20th IEEE International Zurich Seminar on Communications (IZS 2008)*.

[8] Begleiter, R. and Yona, G. (2004). On prediction using variable order markov models. *Journal of Artificial Intelligence Research*, 22:385–421.

[9] Bentley, J. L. (1975). Multidimensional binary search trees used for asso- ciative searching. *Communications of the ACM*, 18(9):509–517.

[10] Boyan, J. A. and Littman, M. L. (1994). Packet routing in dynamically changing networks: A reinforcement learning approach. In *Advances in Neural Information Processing Systems 6*, pages 671–678. Morgan Kaufmann.

[11] Branch, J., Szymanski, B., Giannella, C., Wolff, R., and Kargupta,. In-network outlier detection in wireless sensor networks. In *Proceedings of the 26th International Conference on Distributed Computing Systems (ICDCS)*, pages 102–111. IEEE Computer Society.

[12] Caro, G. D., Caro, G. D., Ducatelle, F., Ducatelle, F., Gambardella,L. M., and Gambardella, L. M. (2005). Anthocnet: An adaptive nature-inspired algo- rithm for routing in mobile ad hoc networks. *European Transactions on Telecommuni- cations*, 16:443–455.

[13] Carpenter, G. and Grossberg, S. (1991). Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, 4:759–771.

Siddarth Tiwari-He is an asst. professor in the Department of Computer science and Engineering Shri Ram Institute of science and Technology Jabalpur.He recived his M.Tech in computer science Engg. In 2012 from Rajiv Gandhi Technical university Bhopal .He published several research papers in various national and international journals .HE recived "Best Teacher" award in 2013 by MPCST(MP Council of science and technology Bhopal).His area of interest are information security ,big data,WSN,machine learning and compiler.

Dimit Vishwakarma-He is student of M.Tech with computer science and engg branch. From RGTU(Rajiv Gandhi Technical university) Bhopal. He completed his B.E. form RGTU Bhopal in 2011.He has a experience of working as a ASE(Asst. System Engg.) at tcs(TATA Consultancy Services) .His area of interest information security ,big data,WSN,windows,Operating System.