International Conference on Intelligent Technologies & Science - 2021
(ICITS-2021)

A Review on Outlier Detection using ANN For WSN Application

Brijesh Yadav[1], Ravi Prakash Chaturvedi[2], Sameer Asthana[3]
Computer Science Engineering,
United College of Engiineering & Research, Greater Noida., India
brijeshmailon@gmail.com, rpchaturvedi51@gmail.com, sam.asthana@gmail.com

Abstract: Wireless sensor networks (WSNs) are composed of a large number of tiny sensor nodes deployed in an environment for monitoring and tracking purposes. Sensor nodes use ad-hoc communications and collaborate with each other to sense different phenomena that may vary in time and space, and send the sensed data to a central node for further processing and analysis. An anomaly is an observation in a data set, which appears to be inconsistent with the remainder of that data set.

Keywords: Anomaly Detection, Entropy, Outliers Detection, Wireless Sensor Networks.

1. Introduction:

The complex and dynamic characteristics of WSNs have made them vulnerable to anomalies. Anomalies are observations that do not correspond to a well defined notion of normal behaviours. [1].In WSNs, anomalies can occur in the nodes, networks, transmission channels and application data and can be caused by systematic errors, random errors and malicious attacks. For instance, WSNs may be deployed in a hostile and inaccessible location, maintenance on the network components is impossible. These nodes usually operate unattended over a long period of time until the battery depleted. Node failure can cause the networks to be unavailable. The networks are also susceptible to systematic hardware failure, random hardware and communication errors, and malicious attacks.
Anomalies in WSNs can be [2] classified into three broad categories, Node, Network and Data anomaly. Node Anomalies occur due to fault at single node. Main reason behind this anomaly is battery issue, i.e. battery failure or depletion. The node fault occurs due to deployment of nodes in harsh environment Unlike node anomalies, the network Anomalies can occur at group of nodes. These are mainly communication related problem. The sensor nodes communicate with each other and if that communication is interrupted due to some reasons then network anomaly occurs. Malicious attacks such DOS, sinkhole, black hole, selective forwarding and wormhole attacks contributes to the occurrence of network anomalies. Data Anomaly occurs when there are some irregularities are present in the sensed data. Some security breaches can also lead to anomalous data. Data anomalies are of three types, viz. Temporal, Spatial and Spatial temporal. Temporal anomaly at a single node location due to changes in data values over time. Spatial anomaly at a single node location due to comparison with neighbouring nodes. Spatiotemporal anomaly detected through a number of node location due to changes in data value over time and space.

2. Related Work:

Outlier detection techniques designed for WSNs can be categorized into statistical based, nearest neighbour-based, clustering based, classification-based, and spectral decomposition-based approaches [3] Classification-based approaches are important systematic approaches in the data mining and machine learning community. Classification-based techniques learn a classification model using a set of data instances in the training phase and classify an unseen instance into one of the learned (normal/outlier) class in the testing phase. SVM-based techniques are from family of classification-based approaches and have the following three main advantages: (a) have a simple geometric interpretation. (b) provide an optimum solution for classification by maximizing the margin of the decision boundary. (c) avoid the problem of the curse of dimensionality.
The fact that in many WSNs applications, pre-classified normal/anomalous data is neither always available nor easy to obtain implies that unsupervised classification techniques suit the WSNs the best. Further, 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 environment, Existing detection systems either use a statistical based detection technique or a swarm intelligence-based technique.

A) Architectural Structure: Existing anomaly detection techniques mainly use either centralized or distributed or local approach for detecting the anomalies.

i) Centralized Approach: In centralized detection, the anomaly detection is performed at the base station. WSNs collect information from the sensor nodes and send it to the base station to be processed and analyzed. The anomaly detection techniques can utilize this information to detect any missing data or data anomalies collected. [7] used 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. [13] described an approach that is based on distributed non-parametric anomaly detection and requires sensors to maintain a tree communication network topology. Here, each sensor clusters its sampled measurements using a fixed-width clustering algorithm, then extracts statistics of the clusters (i.e., the centroid and number of contained data vectors), and then sends them its parent node.

ii) Distributed Approach: In distributed approach, the detection agent is installed in every node. It monitors the behaviour of neighbouring node within its transmission range locally to detect any abnormal behaviour. To perform a real time anomaly detection, some rule based detection techniques are used in a node. Node listens promiscuously to neighbouring nodes within its transmission range to collect data necessary for anomaly detection. The collected data will be analyzed to detect any deviation from normal behaviour using neighbouring historical data stored in the memory. Once anomalies have been detected, an alarm is sends to alert the base station or neighbouring nodes.[4] presented Intrusion 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. [5] 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.[6] presented intrusion detection schemes that build a model of normal traffic behaviour, 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. [8] presented an outlier detection algorithm based on Bayesian Belief Networks (BBN). The system is also able to estimate missing values in the sensor data. The BBNs are able to capture the relationship between the attributes of sensor nodes as well as spatial temporal correlations that exist among the sensor nodes.[9] developed a framework for the discovery of k-nearest-neighbour based outliers: points whose distance to their k-nn exceeds a fixed threshold or the top n points with respect to the distance to their k-nns. Each sensor maintains a histogram-type summary of pertinent information over a sliding window of its data points. The sink node collects these summaries and queries the network for any additional information needed to correctly determine the outliers over the whole network. The use of summaries allows less communication than a naive, centralized approach. Their approach differs from ours in several ways. First, they only detect outliers over one dimensional data, and difficulty of building compact, multi-dimensional histograms will hinder any extension beyond that. Second, they only consider the two k-nn based outlier definitions described above, while our approach encompasses these and more. Thirdly, their approach only applies in settings where spatial proximity is unimportant while our approach can, if needed, to accommodate spatial proximity ("semi-local" outlier detection). [10] requires the sensors to maintain a tree communication topology and compute outliers using an estimate of the underlying probability distribution from which the data arises. Such an estimate is computed by each sensor maintaining a random sample of its data observations. [11] developed a framework based on a Bayesian Belief Network (BBN) that has been constructed over the WSN (and distributed to each sensor). Using this, each sensor can estimate the likelihood of an observed tuple and, therefore, detect outliers. [12] use a wavelet-based technique for correcting large isolated spikes from single sensor data streams. A dynamic time warping (DTW) distance-based technique is also used to identify more steady intervals of erroneous sensor data by comparing the data streams of spatially close sensors assumed to produce similar streams. [14] studied the anomalies in WSN, desirable properties of anomaly detection techniques and analyze the various anomaly detection techniques for wireless sensor networks.[15] described the types of WSNs and the possible solutions for tackling the listed problems and solution of many other problems. This paper will deliver the knowledge about the WSN and types with literature review so that a person can get more knowledge about this emerging field.

[16] discussed the wireless sensor network which is a modernistic addition in the fields of radio networks and also provides new applications with a new prototype for sensing and broadcasting information from various environments with potential to serve many and desired applications at very low cost.[17] targeted to categorize the routing troubles and examines the routing-related optimization problems.

Subsequently, the literature is analyzed based at the simulation surroundings and experimental setup, attention over the Quality of Service (QoS) and the deployment towards diverse programs.[18] presented a literature review on WSN networks, in which the capacity of network nodes are limited with respect to energy supply, restricted computational capacity and communication bandwidth. It analysed the physical relation between the power consumption and the link utilization of wireless sensor networks. ***iii) Local Approach:*** In some methods, node at their own level within their scope detect the anomalies. [19] proposed a hybrid anomaly detection technique with the application of k-means clustering. The analysis of the network data set consists of traffic data and end to end delay data is performed. The data set is clustered using weka 3.6.10. During the experimentation, it has been observed that two types of anomalies are activated in the network causing misdirection and black hole attacks.

B) Data Dimension: In WSNs, data streams can be univariate or multivariate.

i) Univariate: Univariate streams are
represented by a set of values read by a unique type of sensor, e.g., a sensor node that monitors only environmental temperature. This approaches are in paper [4].

ii) Multivariate : On the other hand, multivariate streams are represented by a set of values coming from different sensors of the same sensor node, e.g., a node that monitors temperature, pressure and humidity simultaneously, or by a set of measurements coming from the same sensor type located in different sensor nodes, e.g., a node that processes data from different nodes monitoring only temperature. The transmission of multivariate data causes excessive network delay, node energy consumption and reduces the network lifetime resulting in node and network anomalies. This approaches are in paper [6].

2.1 Related Work:

The dynamic nature of wireless sensor networks (WSNs) and numerous possible cluster configurations make searching for an optimal network structure on the- fly an open challenge. To address this problem, **Xiaohui Yuan, (2017) [1]** proposed a genetic algorithmbased, self-organizing network clustering (GASONeC) method that provides a framework to dynamically optimize wireless sensor node clusters. In GASONeC, the residual energy, the expected energy expenditure, the distance to the base station, and the number of nodes in the vicinity are employed in search for an optimal, dynamic network structure. Balancing these factors is the key of organizing nodes into appropriate clusters and designating a surrogate node as cluster head. Compared to the state-of-the-art methods, GASONeC greatly extends the network life and the improvement up to 43.44 %. The node density greatly affects the network longevity. Due to the increased distance between nodes, the network life is usually shortened. In addition, when the base station is placed far from the sensor field, it is preferred that more clusters are formed to conserve energy. The overall average time of GASONeC is 0.58 s with a standard deviation of 0.05. Forming network clusters is an effective way of improving the scalability and longevity of WSNs. A pre-determined communication structure or a randomized clustering scheme is far from fulfilling the critical need of maximizing the network life. Despite the great efforts in automatic organizing nodes, the dynamic nature of sensor network and numerous possible cluster configurations make searching for an optimal network structure on-the-fly an open challenge. To address this problem, we propose a GA-based, self-organizing network clustering method that provides a framework to integrate multiple factors and optimize dynamic node clustering. In the GASONeC method, we devise a concise way of encoding nodes and propose fitness functions that include residual energy, expected energy expenditure, distance to the base station, and local node density in search for an optimal, dynamic network structure. Balancing these factors is the key of organizing nodes into appropriate clusters and designating a surrogate node as the cluster head. Compared with state-of-the-art methods, the GASONeC method greatly extended the network life and the improvement is up to 43.44 %. The results showed that as the distance between nodes and the base station increases, the average network life is shortened. This is due to the extra energy required to forward data to the base station. Moreover, when the base station is placed far from the sensor field, it is preferred that more clusters are formed to conserve energy. The node density greatly affects network longevity. Due to the increased distance between nodes, the network life is shortened. The average running time of GASONeC is very close for all cases. It is evident that the efficiency of GASONeC is mostly independent from field size and number of nodes. The overall average time across all experiments is 0.58 s with a standard deviation of 0.05. The efficiency of GASONeC is satisfactory. In future work, we plan to explore the effectiveness of GASONeC in heterogeneous network structures. In addition, the parallel programming will be investigated to reduce the optimization time for constructing the network clusters.

A cluster-based model is preferable in wireless sensor network due to its ability to reduce energy consumption. However, managing the nodes inside the cluster in a dynamic environment is an open challenge. Selecting the cluster heads (CHs) is a cumbersome process that greatly affects the network performance. Although there are several studies that propose CH selection methods, most of them are not appropriate for a dynamic clustering environment. To avoid this problem, several methods were proposed by **Mohamed Elhoseny, (2017), [2]** based on intelligent algorithms such as fuzzy logic, genetic algorithm (GA), and neural networks. However, these algorithms work better within a single-hop clustering model framework, and the network lifetime constitutes a big issue in case of multi-hop clustering environments. This paper introduces a new CH selection method based on GA for both single-hop and the multi-hop cluster models. The proposed method is designed to meet the requirements of dynamic environments by electing the CH based on six main features, namely, (1) the remaining energy, (2) the consumed energy, (3) the number of nearby neighbors, (4) the energy aware distance, (5) the node vulnerability, and (6) the degree of mobility. We shall see how the corresponding results show that the proposed algorithm greatly extends the network lifetime. In a clustering WSN, each cluster usually consists of at least one surrogate node, often referred to as the CH. The CH may be dynamically chosen or preassigned by the network designer. Communication between the cluster and the base station is facilitated by this CH. The problem of choosing the CH in a multi-hop clustering model is more complex than in a single-hop model. The multi-hop clustering model is a special case of the clustering model in which a CH cannot transmit the data directly to the BS. There are six main significant factors for selecting a CH node in a multi-hop cluster model in WSNs. These factors are: the distance from cluster center, the vulnerability index, the degree of mobility, the degree of mobility, the remaining battery power, the number of nearby neighbors, and the consumed energy. All these factors are related when choosing CH and ignoring one of them will affect the network lifetime. The degree of priority differs for each factor. In the present contribution, we have proposed a new CH selection method based on GA, called DCH-GA, for both single-hop and the multi-hop cluster models. The procedure introduced here is designed to match the requirements of the dynamic environments by electing the CH based on six main features. During examination, we repeated simulations and reported the average performance. Comparison studies were conducted with some methods reported in the literature. In the future work, we shall focus on secure data transfer between each node and the CH node taking into consideration the dynamic environment inside the cluster.

NitinMittal (2016), [3] worked on nature-inspired algorithms are becoming popular among researchers due to their simplicity and flexibility. The nature-inspired metaheuristic algorithms are analysed in terms of their key features like their diversity and adaptation, exploration and exploitation, and attractions and diffusion mechanisms. The success and challenges concerning these algorithms are based on their parameter tuning and parameter control. A comparatively new algorithm motivated by the social hierarchy and hunting behavior of grey wolves is Grey Wolf Optimizer (GWO), which is a very successful algorithm for solving real mechanical and optical engineering problems. In the original GWO, half of the iterations are devoted to exploration and the other half are dedicated to exploitation, overlooking the impact of right balance between these two to guarantee an accurate approximation of global optimum. To overcome this shortcoming, a modified GWO (mGWO) is proposed, which focuses on proper balance between exploration and exploitation that leads to an optimal performance of the algorithm. Simulations based on benchmark problems and WSN clustering problem demonstrate the effectiveness, efficiency, and stability ofmGWO compared with the basicGWO and some well-known algorithms. This paper proposed a modification to the Grey Wolf Optimizer named mGWO, inspired by the hunting behavior of grey wolves in nature. An exponential decay function is used to balance the exploration and exploitation in the search space over the course of iterations. The results proved that the proposed algorithm benefits from high exploration in comparison to the standard GWO. The paper also considered the clustering problem inWSN in which the CH selection is performed using the proposed mGWO algorithm, which is a challenging and NP hard problem. The results show that the proposed method is found to be very effective for real-world applications due to fast convergence and fewer chances to get stuck at local minima. It can be concluded that the proposed algorithm is able to outperform the current well-known and powerful algorithms in the literature. The results prove the competence and superiority of mGWO to existing metaheuristic algorithms and it has an ability to become an effective tool for solving real word optimization problems.

Vehicular Ad hoc NETworks (VANETs) are a major component recently used in the development of Intelligent Transportation Systems (ITSs). VANETs have a highly dynamic and portioned network topology due to the constant and rapid movement of vehicles. Currently, clustering algorithms are widely used as the control schemes to make VANET topology less dynamic for Medium Access Control (MAC), routing and security protocols. An efficient clustering algorithm must take into account all the necessary information related to node mobility. In this work, **Mohamed Hadded,(2015) [4]** proposed an Adaptive Weighted Clustering Protocol (AWCP), specially designed for vehicular networks, which takes the highway ID, direction of vehicles, position, speed and the number of neighboring vehicles into account in order to enhance the stability of the network topology. However, the multiple control parameters of our AWCP, make parameter tuning a nontrivial problem. In order to optimize the protocol, we define a multi-objective problem whose inputs are the AWCP's parameters and whose objectives are: providing stable cluster structures, maximizing data delivery rate, and reducing the clustering overhead. We address this multi-objective problem with the Nondominated Sorted Genetic Algorithm version 2 (NSGA-II). We evaluate and compare its performance with other multi-objective optimization techniques: Multi-objective Particle Swarm Optimization (MOPSO) and Multi-objective Differential Evolution (MODE). The experiments reveal that NSGA-II improves the results of MOPSO and MODE in terms of spacing, spread, ratio of non-dominated solutions, and inverse generational distance, which are the performance metrics used for comparison. Because of the rapidly changing topology and the lack of infrastructure, it is very challenging to deploy clustering methods in vehicular networks. In this paper, we focus on designing an adaptive and optimized clustering algorithm for vehicular networks, called AWCP, that takes into consideration the highway ID, direction, position, and speed information, in order to maximize cluster stability. However, due to the high number of feasible configurations of AWCP and the conflicting nature of its performance metrics, we defined a multi-objective optimization problem where the non-dominated sorted genetic algorithm NSGA-II is coupled with the ns2 simulator to find the optimal parameter values for the AWCP QoS metrics. The NSGA-II optimized configuration is validated by comparing it with the optimized MODE and MOSPO configurations on realistic VANET scenarios taken from the metropolitan area of Tunis (Tunisia). The experimental results show that the NSGAII algorithm obtains well-distributed solutions over the Pareto front and presents the best results in terms of performance metrics. Thus, NSGA-II algorithm is more suitable for the AWCP parameter tuning problem. Since the computational time required to perform 15 independent runs for all MOEAs in the S4 scenario is about 32 days, a parallel version of MOEAs running on multiple processors would allow larger populations and more generations to be used in this multi-objective optimization method while reducing the computational time required for very large scale VANET scenarios. Moreover, channel efficiency in VANETs could be improved by the development of a crosslayer architecture (MAC/AWCP) in which each cluster head is responsible for assigning bandwidth to all the members of its cluster.

Every type of network, be it wired or wireless, will be influenced by several key factors for its efficient functioning. Routing issue, applicable to all types of networks, is one among the several such key factors. Wireless Sensor Networks (WSN) has not been exception to this. Moreover, such issues are very critical due to severe resource constraints like efficient energy utilization, lifetime of network, and drastic environmental conditions in WSNs. Neither hop-by-hop or neither direct reach ability is possible in case of WSNs. In this regard, many routing protocols have been proposed **by Geetha. V.(2012) [5]** to optimize the efficiency of WSNs amidst of above mentioned severe resource constraints. Out of these, clustering algorithms have gained more importance, in increasing the life time of the WSN, because of their approach in cluster head selection and data aggregation. LEACH (distributed) is the first clustering routing protocol which is proven to be better compared to other such algorithms. This paper elaborately compares two important clustering protocols, namely LEACH and LEACH-C (centralized), using NS2 tool for several chosen scenarios, and analysis of simulation results against chosen performance metrics with latency and network lifetime being major among them. The paper will be concluded by mentioning the observations made from analyses of results about these protocols. Wireless Sensor Networks, which may be spread over vast geographical area, are finding applications in many areas. In this context, there is need of approaches which can manage these WSNs in better way. In this regard, this paper, presented need for clustering to overcome several limitations of WSNs. Detailed discussion about existing work is provided. Brief working of chosen clustering protocols, namely LEACH & LEACH-C, is presented. We also presented the simulation results and analyses of these protocols. As a conclusion of observation from results, it can be mentioned that LEACH can be preferred if localized coordination of nodes in clustering without involving BS is of high priority than other factors like assurance over desired number of clusters etc.; and LEACH-C can be chosen when centralized and deterministic approach covering entire network is expected still bringing in increased network lifetime and desired number of clusters.

Wireless sensor networks are composed of a large number of sensor nodes with limited energy resources. One critical issue in wireless sensor networks is how to gather sensed information in an energy efficient way since the energy is limited. The clustering algorithm is a technique used to reduce energy consumption. It can improve the scalability and lifetime of wireless sensor network. In this paper, **Fuad Bajaber,(2011) [6]** introduce an adaptive clustering protocol for wireless sensor networks, which is called Adaptive Decentralized Re-Clustering Protocol (ADRP) for Wireless Sensor Networks. In ADRP, the cluster heads and next heads are elected based on residual energy of each node and the average energy of each cluster. The simulation results show that ADRP achieves longer lifetime and more data messages transmissions than current important clustering protocol in wireless sensor networks. We introduce an adaptive clustering scheme ADRP, for electing cluster heads and next heads in wireless sensor networks. The selection of cluster heads and next heads are weighted by the remaining energy of sensor nodes and the average energy of each cluster. The sensor nodes with the highest energy in the clusters can be a cluster heads at different cycles of time. By means of the former, the role of cluster heads can be switched dynamically. Simulations results show that ADRP has extended the lifetime of the network and reduced the communication overhead. Hence, the performance of the proposed protocol is better in terms of lifetime, data delivery and communication overhead, when compared with LEACH-C and CDC. When the sensor nodes use single hop communication to reach the base station, the sensor nodes located farther away from the base station have the highest energy load due to long range communication. When the sensor nodes use multihop communication to reach the base station, the sensor nodes closer to the base station have a higher load of relaying packets. As for future work, we will design an adaptive and energy efficient protocol to determine the optimum mode of communication in each cluster single hop or multihop.

Some applications of Wireless Sensor Networks (WSNs) to the automobile are identified, and the use of Crossbow MICAz motes operating at 2.4 GHz is considered together with Tiny OS support. These WSNs are conceived in order to measure, process and supply to the user diverse types of information during an automobile journey. Examples are acceleration and fuel consumption, identification of incorrect tire pressure, verification of illumination, and evaluation of the vital signals of the driver. A brief survey on WSNs concepts is presented by **Jorge Tavares,(2008) [7]**, as well as the way the wireless sensor network itself was developed. Calibration curves were produced which allowed for obtaining luminous intensity and temperature values in the appropriate units. Aspects of the definition of the architecture and the choice/implementation of the protocols are identified. Security aspects are also addressed. This work addressed the conception of a WSN capable of measuring, processing and supplying diverse types of information to the user during an automobile journey. The examples are acceleration and fuel consumption, identification of incorrect tire pressure, failures of illumination, and evaluation of the vital signals of the driver. Beside a survey on the concepts, the wireless sensor network itself (transmitter/ receiver/control board) was configured, and aspects of the architecture and protocols were addressed. By using the calibration curves for the light and temperature sensors, precise experimental values were extracted. Security aspects were also identified, and the difficulties and solutions were discussed. Competition cars in a controlled environment constitute a suitable scenario for experimental work. Besides, the evolutions in this field promise a lot in the automobile industry, e.g., for cooperation among cars for road safety purposes.

Wireless sensor networks consist of sensor nodes with sensing and communication capabilities.
Ramesh Rajagopalan,(2006) [8] focused on data aggregation problems in energy constrained sensor networks. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. In this paper, we present a survey of data aggregation algorithms in wireless sensor networks. We compare and contrast different algorithms on the basis of performance measures such as lifetime, latency and data accuracy. We conclude with possible future research directions. We have presented a comprehensive survey of data aggregation algorithms in wireless sensor networks. All of them focus on optimizing important performance measures such as network lifetime, data latency, data accuracy and energy consumption. Efficient organization, routing and data aggregation tree construction are the three main focus areas of data aggregation algorithms. We have described the main features, the advantages and disadvantages of each data aggregation algorithm. We have also discussed special features of data aggregation such as security and source coding. The trade-offs between energy efficiency, data accuracy and latency have been highlighted. Most of the existing work has mainly focused on the development of an efficient routing mechanism for data aggregation. However, the performance of the data aggregation protocol is strongly coupled with the infrastructure of the network. There has not been significant research on exploring the impact of heterogeneity and mode of communication (single hop versus multi-hop) on the performance of the data aggregation protocols. Although, many of the data aggregation techniques presented look promising, there is significant scope for future research. Combining aspects such as security, data latency and system lifetime in the context of data aggregation is worth exploring. A systematic study of the relation between energy efficiency and system lifetime is an avenue of future research. Analytical results on the bounds for lifetime of sensor networks is another area worth exploring. Existing work has provided bounds on lifetime for networks with specific network topologies and source behaviors. It would be interesting to extend this work to more general network topologies such as cluster based sensor networks. Security is another important issue in data aggregation applications and has been largely unexplored. Integrating security as an essential component of data aggregation protocols is an interesting problem for future research. Data aggregation in dynamic environments presents several challenges and is worth exploring in the future. Another interesting domain of research is the application of source coding theory for data gathering networks. The sensor data are usually highly correlated and energy efficiency can be achieved by joint source coding and data compression. Although some research has been pursued in this direction [20], there is significant scope for future work.

Networking together hundreds or thousands of cheap microsensor nodes allows users to accurately monitor a remote environment by intelligently combining the data from the individual nodes. These networks require robust wireless communication protocols that are energy efficient and provide low latency. In this paper, **Wendi B. Heinzelman (2002) [9]** develop and analyze low-energy adaptive clustering hierarchy (LEACH), a protocol architecture for microsensor networks that combines the ideas of energy-efficient cluster-based routing and media access together with application-specific data aggregation to achieve good performance in terms of system lifetime, latency, and application-perceived quality. LEACH includes a new, distributed cluster formation technique that enables self-organization of large numbers of nodes, algorithms for adapting clusters and rotating cluster head positions to evenly distribute the energy load among all the nodes, and techniques to enable distributed signal processing to save communication resources. Our results show that LEACHcan improve system lifetime by an order of magnitude compared with general-purpose multihop approaches. When designing protocol architectures for wireless microsensor networks, it is important to consider the function of the application, the need for ease of deployment, and the severe energy constraints of the nodes. These features led us to design LEACH, a protocol architecture where computation is performed locally to reduce the amount of transmitted data, network configuration and operation is done using local control, and media access control (MAC) and routing protocols enable low-energy networking. Results from our experiments show that LEACH provides the high performance needed under the tight constraints of the wireless channel.

I.F. Akyildiz,(2002) [10] described the concept of sensor networks which has been made viable by the convergence of microelectro- mechanical systems technology, wireless communications and digital electronics. First, the sensing tasks and the potential sensor networks applications are explored, and a review of factors influencing the design of sensor networks is provided. Then, the communication architecture for sensor networks is outlined, and the algorithms and protocols developed for each layer in the literature are explored. The flexibility, fault tolerance, high sensing fidelity, low-cost and rapid deployment characteristics of sensor networks create many new and exciting application areas for remote sensing. In the future, this wide range of application areas will make sensor networks an integral part of our lives. However, realization of sensor networks needs to satisfy the constraints introduced by factors such as fault tolerance, scalability, cost, hardware, topology change, environment and power consumption. Since these constraints are highly stringent and specific for sensor networks, new wireless ad hoc networking techniques are required. Many researchers are currently engaged in developing the technologies needed for different layers of the sensor networks protocol stack as shown in Fig. 3. A list of current sensor networks research projects is given in Table 5. Along with the current research projects, we encourage more insight into the problems and more development in solutions to the open research issues as described in this paper.

In recent years, advances in energy-efficient design and wireless technologies have enabled exciting new applications for wireless devices. These applications span a wide range, including real-time and streaming video and audio delivery, remote monitoring using networked microsensors, personal medical monitoring, and home networking of everyday appliances. While these applications require high performance from the network, they suffer from resource constraints that do not appear in more traditional wired computing environments. In particular, wireless spectrum is scarce, often limiting the bandwidth available to applications and making the channel error-prone, and the nodes are battery-operated, often limiting available energy. **Wendi Beth Heinzelman (2000) [11]** worked that this harsh environment with severe resource constraints requires an application specific protocol architecture, rather than the traditional layered approach, to obtain the best possible performance. This dissertation supports this claim using detailed case studies on microsensor networks and wireless video delivery. The first study develops LEACH (Low-Energy Adaptive Clustering Hierarchy), an architecture for remote microsensor networks that combines the ideas of energy-efficient cluster-based routing and media access together with application-specific data aggregation to achieve good performance in terms of system lifetime, latency, and application perceived quality. This approach improves system lifetime by an order of magnitude compared to general-purpose approaches when the node energy is limited. The second study develops an unequal error protection scheme for MPEG-4 compressed video delivery that adapts the level of protection applied to portions of a packet to the degree of importance of the corresponding bits. This approach obtains better application-perceived performance than current approaches for the same amount of transmission bandwidth. These two systems show that application-specific protocol architectures achieve the energy and latency efficiency and error robustness needed for wireless networks. Use of the wireless channel is growing at an amazing speed. Advances in energy-efficient design have created new portable devices that enable exciting applications for the wireless channel. While the wireless channel enables mobility, it adds constraints that are not found in a wired environment. Specifically, the wireless channel is bandwidth-limited, and the portable devices that use the wireless channel are typically battery-operated and hence energy-constrained. In addition, the wireless channel is error-prone and time-varying. Therefore, it is important to design protocols and algorithms for wireless networks to be bandwidth- and energy-efficient as well as robust to channel errors. This can be accomplished using cross-layer protocol architectures, that exploit application-specific information to achieve orders of magnitude improvement in bandwidth and energy efficiency and improvements in application-perceived quality. The work described in this dissertation has demonstrated the advantages of application-specific protocol architectures by designing and evaluating protocol architectures for two different application spaces: large-scale, distributed microsensor networks and wireless transport of compressed video.

3. Conclusion:

A novel technique is demonstrated on wireless sensor dataset different prediction model learning methods: Based on the idea of varying the parameters of the model attribute classes and applied it to design a new prediction method. Despite the simplicity of development approach multiple methods have validated when compared with each other in terms of the prediction accuracy of outlier in sensor data including segment entropy as input feature specifically devised for data classification.

References:

[1] Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. 41*, 15, 2009.

[2] Raja Jurdak, X. Rosalind Wang, Oliver Obst, and Philip Valencia,"Wireless Sensor Network Anomalies: Diagnosis and Detection Strategies", 2011.

[3] Y. Zhang, N. Meratnia, and P. J. M. Havinga, "Outlier Detection Techniques for Wireless Sensor Network" A Survey,Technical Report, University of Twente, 2008.

[4] Techateerawat, P. and Jennings, "A Energy efficiency of intrusion detection systems in wireless sensor networks", In Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology (WI- IATW), pages 227–230, Washington, DC, USA. IEEE Computer Society, 2006. [5] Hai, T. H., Khan, F. and nam Huh, E. "Hybrid intrusion detection system for wireless sensor networks", Computational Science and Its Applications ICCSA, 4706:383–396, 2007.

[6] Onat and Miri, Onat, I. and Miri, "An intrusion detection system for wireless sensor networks", In IEEE International Conference on Wireless And Mobile Computing, Networking and Communications, Los Alamitos, IEEE Computer Society Press, 2005.

[7] Banerjee, Banerjee S., Grosan C., and Abraham, "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, 2005.

[8] Janakiram, D., Reddy, V., and Kumar A., "Outlier detection in wireless sensor Networks using Bayesian belief networks". In First International, Conference on Communication System Software and Middleware, pages 1–6, 2006.

[9] Sheng B, Li Q, Mao W, Jin W, "Outlier detection in sensor networks", In Proceedings of the 8th ACM international symposium on mobile and ad hoc networking and computing (MobiHoc), pp 219–228, 2007. [10]Subramaniam S, Palpanas T, Papadopoulos D, KalogerakiV, Gunopulos D., "Online outlier detection in sensor data using non-parametric models" In Proceedings of ACM conference on very large databases (VLDB06), pp 187–198, 2006.

[11] Janakiram D, Reddy VA, Kumar AVUP, "Outlier detection in wireless sensor Networks using Bayesian belief networks" In Proceedings of IEEE conference on communication system software and middleware (Comsware06), pp 1–6, 2006.

[12] Zhuang Y, Chen L "In-network outlier cleaning for data collection in sensor networks", In proceedings of the 1st international VLDB workshop on clean databases (CleanDB06), 2006.

[13] Rajasegarar S, Leckie C, Palaniswami M, Bezdek J, "Distributed anomaly detection in wireless sensor networks", In proceedings of the IEEE Singapore international conference on communication systems, pp 1–5, 2006.

[14] Satish S. Bhojannawar, Chetan M Bulla, Vishal M Danawade, "Anomaly Detection Techniques for Wireless Sensor Networks - A Survey", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 10, pp.3852-3857, 2013.

[15] R. Sharma, Anurag, " Detect Skin Defects by Modern Image Segmentation Approach, Volume 20, Issue 1, 2020.

[16] Anurag, R. Sharma, " Modern Trends on Image Segmentation for Data Analysis- A Review", International Journal of Research and Development in Applied Science and Engineering, Volume 20, Issue 1, 2020.

[17] Young Soo Jang et. al., "Development of the cost-effective, miniaturized vein imaging system with enhanced noise reduction", International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.6, November – December 2019.

[18] Irma T. Plata1, et. al., "Development and Testing of Embedded System for Smart Detection and Recognition of Witches' Broom Disease on Cassava Plants using Enhanced Viola-Jones and Template Matching Algorithm", International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.6, Volume 8, No.5, September - October 2019.

[19] Anurag, R. Sharma, " Load Forecasting by using ANFIS", International Journal of Research and Development in Applied Science and Engineering, Volume 20, Issue 1, 2020.

[20] R. Sharma, Anurag, " Load Forecasting using ANFIS A Review", International Journal of Research and Development in Applied Science and Engineering, Volume 20, Issue 1, 2020.