Distributed Anomaly Detection Models for Industrial Wireless Sensor Networks

A thesis submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy

Heshan Dhanushka Kumarage

B.Sc. Engineering

School of Computer Science and Information Technology

College of Science, Engineering, and Health

RMIT University

March, 2015

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Heshan Dhanushka Kumarage
30^{th} March, 2015

Acknowledgement

The preceding work would not have been possible without the kind help, guidance and support from many people. Here, I express my deep appreciation towards them all.

First, I would like to thank my supervisors Prof. Zahir Tari and A/Prof. Ibrahim Khalil. Their dedication, unreserved support and efforts ensured a smooth candidacy and the ultimate completion of this research. Many times their advices helped me in not losing track of what I was trying to accomplish. Endless encouragement of my supervisor A/Prof. Ibrahim Khalil boosted my spirits and kept me motivated throughout the duration of the Ph.D. All their efforts are greatly appreciated.

I would also like to thank my family. Foremostly to my father who has passed away. It is with deep gratitude and love that I dedicate this thesis to him, my father, Lankathilake Kumarage. I would not have been in this position if not for you. To my mother, Chithrani Senanayake, who always helped me on everything and provided emotional strength often while being thousands of kilometres away. My wife, Madhuwanthi who has been a constant companion on this journey as we shared and endured many different challenges together. Thanks for the limitless love and support that I continuously receive. To my little daughter, Jinuli Akithra for the constant joy she brings to my life. You make the darkest of days shine with the illumination of the brightest of stars.

Finally, I would like to thank all my friends in and outside of RMIT that directly or indirectly supported me throughout this time.

Credits

Portions of the material in this thesis have previously appeared in the following publications:

- H. Kumarage, I. Khalil, Z.Tari, and Albert Zomaya. "Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling." *Journal of Parallel and Distributed Computing* 73, no. 6 (2013): 790-806. **Accepted,** February 2013.

- H. Kumarage, I. Khalil and Z.Tari. "Granular Evaluation of Anomalies in Wireless Sensor Networks using Dynamic Data Partitioning with an Entropy Criteria." *IEEE Transactions on Computers.* **Accepted,** November 2014.

- Yaakob, N., I. Khalil, H. Kumarage, M. Atiquzzaman, and Z. Tari. "By-Passing Infected Areas in Wireless Sensor Networks using BPR." *IEEE Transactions on Computers.* **Accepted,** July 2014.

- H. Kumarage, I. Khalil and Z.Tari. "Point-of-View (POV) Entropy Evaluations for Real-time Decision Support in Sensor Network Data Streams" *IEEE Transactions on Information Forensics and Security.* **Submitted,** March 2015.

Abstract

Wireless Sensor Networks (WSNs) are firmly established as an integral technology that enables automation and control through pervasive monitoring for many industrial applications. Consisting of a large number of spatially distributed and autonomous sensors they collaborate to collect process and deliver information. Usually deployed unattended in an area of interest with little or no associated infrastructure their application domains span many environments. These range from environmental or earth sensing applications and healthcare applications to major industrial monitoring applications such as machine use, infrastructure and structural monitoring. Streamlined sensing capabilities and the potential for aggregated intelligence via parallel processing offer unique opportunities for distributed control in these applications. The key features that are common to such applications can be noted as involving large amounts of data, consisting of dynamic observation environments, non-homogeneous data distributions with evolving patterns and sensing functionality leading to data-driven control and decision making. Also in most industrial applications a major requirement is to have near real-time decision support over streaming functionality that produces a large amount of data. Accordingly there is a vital need to have a secure continuous and reliable sensing mechanism in integrated WSNs where the integrity of the data is assured.

However, in practice WSNs are vulnerable to different security attacks, faults and malfunction due to their inherent resource constraints, openly commoditised wireless technologies employed and often, naive modes of implementation. Misbehaviour resulting from such threats manifest as anomalies in the sensed data streams in critically compromising the systems through wrong operational and control decisions. Therefore, it is vital that effective techniques are introduced in accurately detecting anomalies and assuring the integrity of the

data. Considering these factors this research focuses on investigating effective anomaly detection models for large scale industrial wireless sensor networks. Particular attention is given to devising distributed models that are adaptable and scalable, works in an unsupervised manner without prior training, achieves non-parametric and non-probabilistic detection over dynamic data streams and supports near real-time decision making.

Focusing on achieving an anomaly detection framework that is both adaptable and scalable, a hierarchical data partitioning approach with fuzzy data modelling is introduced first. In this model unsupervised data partitioning is performed in a distributed manner by adapting fuzzy c-means clustering in an incremental model over a hierarchical node topology. It is found that non-parametric and non-probabilistic determination of anomalies can be done by evaluating the fuzzy membership scores and inter-cluster distances adaptively over the node hierarchy. Robust thresholds for anomaly differentiation are derived using only second order statistical knowledge that is locally available. The viability of the model is demonstrated through sensitivity and specificity analysis performed for a variety of data distributions. Scalability of the proposed model is highlighted from the reduced communication costs that accompany high detection accuracies that are achieved when compared to existing approaches.

Specifically, considering the heterogeneous data distributions with evolving patterns, a granular anomaly detection model that uses an entropy criterion to dynamically partition the data is proposed next. This successfully overcomes the issue of determining the proper number of expected clusters with regard to an unsupervised cluster based anomaly detection process. In this approach the data is partitioned on to different cohesive regions using cumulative point-wise entropy directly. The effect of differential density distributions when relying on an entropy criterion is mitigated by introducing an average relative density measure to segregate isolated outliers prior to the entropy based partitioning. The combination of these two factors is shown to be significantly successful in determining anomalies adaptively in a fully dynamic manner. The model is also implemented in-network over a hierarchical topology with reduced communication costs in offering granular anomaly detection over different network levels. Robustness of the proposed model is highlighted through the higher level of detection accuracy attained for a variety of data distributions representing dynamic observation domains

The need for near real-time anomaly evaluation is focused next on this thesis. Building upon the entropy based data partitioning model that is also proposed, a Point-of-View (PoV) entropy evaluation model is developed next. This employs an incremental data processing model that is performed locally over the different nodes as opposed to batch-wise data processing. Three unique points-of-view are introduced as the reference points over which point-wise entropy is computed in evaluating its relative change as the data streams evolve. It is shown that each of these points are capable in offering different lenses that evaluate relative entropy change with regard to identifying anomalies in an online fashion. In order to identify instances of sensor drift and level shifts that occur as part of normal behaviour, a secondary analysis stage is also incorporated where potential anomalies are again subjected to entropy evaluation. This works to significantly reduce false alarms that are common with an otherwise unsupervised process. Experiments indicate higher detection accuracies with low false alarm rates for a variety of evolving behaviour.

Overall this thesis proposes efficient unsupervised anomaly detection models that employ distributed in-network data processing for accurate determination of anomalies. The introduced models particularly cater to dynamic observation domains consisting of heterogeneous and non-homogeneous data with evolving patterns. The resource constrained environment is taken in to account in each of the models with innovations made to achieve non-parametric and non-probabilistic detection. Therefore, the research contributions in this thesis presents effective unsupervised models to solve the critical issue of determining data anomalies in scalable and adaptive frameworks that is robust with regard to dynamic data domains in large scale industrial wireless sensor networks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Wireless Sensor Networks (WSNs) are a comparatively new technology that is increasingly gaining traction over a wide spectrum of different applications in the context of pervasive monitoring and automation [Yick et al., 2008]. They consist of a large number of spatially distributed and autonomous sensors that collaborate to collect, process and deliver information. Each individual sensor node is equipped with a radio tranceiver, microcontroller an integrated memory, an energy source (e.g battery or form of energy harvesting) and application sensors. These sensors are used to jointly measure physical or environmental conditions such as temperature, humidity, pressure, voltage, sound, vibration and motion. Most modern sensor networks are designed and configured to support bi-directional functionality that allows for sensor activity control and querying in addition to normal data communication originating from source nodes.

Initially developed for military applications, such as battlefield surveillance their current usage today spans a wide range of industrial and consumer applications where the need for more automation and continuous monitoring is increasing [Gungor and Hancke, 2009]. To date, WSNs have been successfully applied to many such domains and include environmental or earth sensing applications (forest fire/landslide detection, water and air quality assessment, land use/irrigation monitoring), healthcare applications (patient monitoring, body sensor networks) and industrial applications such as machine use, infrastructure and structural health monitoring [Buttyan et al., 2010], [Bertocco et al., 2008], [Luo et al., 2012], [Guevara et al., 2012].

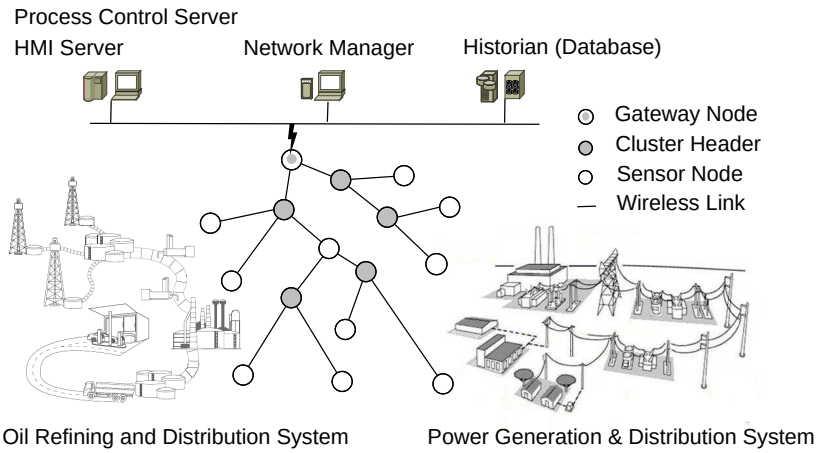

Figure 1.1: Wireless Sensor Networks in Distributed Industrial Infrastructure

WSNs are usually deployed unattended in an area of interest which may either be a homogeneous or heterogeneous environment. Their deployment typically involves little or no infrastructure and can be classified as a structured or unstructured network [Yick et al., 2008]. If the node deployment is done in an ad-hoc manner it is said to be unstructured where as if it is done in a pre-planned manner (typically in some form of a topological hierarchy) it is said to be a structured network. Each node may perform a variety of services such as localization, synchronization, data aggregation or compression as well as security and coverage depending on the needs of the application in keeping effective network performance. The communication among sensor nodes is achieved through the radio transceiver and following the five layer communication protocol stack that include the physical layer, data-link layer, network layer, transport layer and the application layer.

The area that has gained the most popularity with regard to adaption of WSNs for pervasive monitoring is the industrial monitoring applications. Streamlined sensing capabilities and the potential for aggregated intelligence (via parallel processing) also offer unique opportunities for distributed control as required in this area [Buttyan et al., 2010, Puccinelli and Haenggi, 2005]. However, these applications also provide the most challenges due to their very dynamic states of monitored environments and the critical decision making that is enabled through sensed data in contrast to most land use and environmental monitoring applications. Moreover these applications feature continuous automation and control requirements as well as in some cases

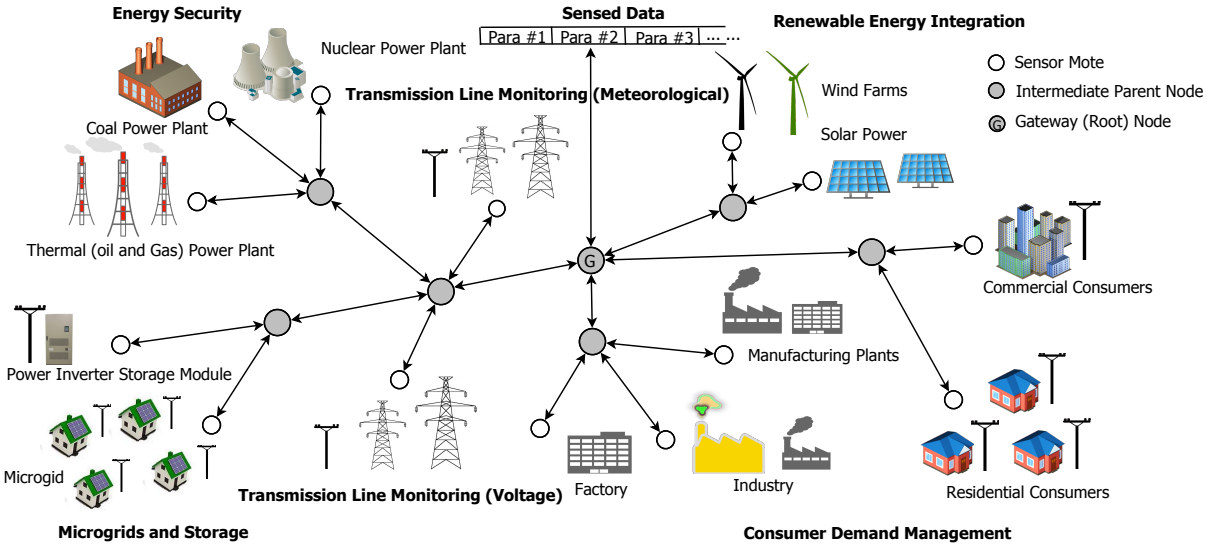

Figure 1.2: Wireless Sensor Networks (WSNs) in a Smart Grid Environment: WSNs are used for distributed sensing and communication regarding a variety of applications in Smart Grid. From ensuring energy security and renewable energy integration to dynamic power allocation and management at micro-grid levels through real time consumer demand management they form a vital information core that enables smart functionality. The different applications with regular time critical sensing paves the way for *very large amounts* of multi dimensional data of which ensuring integrity is a core concern in achieving reliable and optimum performance. Such data demographics are *dynamic* and *non-homogeneous* presenting a challenge for traditional security measures. The Sensor networks are often dynamically ordered on to a hierarchical topology offering sensing granularity at different levels in this large scale heterogeneous environment.

the need for near real-time decision support. Examples include blockage and leakage detection in pipeline systems, machine health monitoring, structural stability monitoring and predictive maintenance to detect potential failures in large scale industrial facilities such as oil and gas industries and power generation and distribution systems. Figure 1.1 shows an example of a WSN connecting to an industrial process control system with the application environments in power and oil sectors. Here, the HMI (Human Machine Interface) Server presents the updated information collected over the WSN while the Historian Database keeps a record of all aggregated observations. The WSN is typically modelled as a hierarchical topology of sensors with the Network Manager controlling communication and processing the sensed information in between.

Another key industrial platform where the use of WSNs is currently gaining prominence is *smart grids* in the power industry. This application involves most features as discussed

above and can be used as a key example that enables to understand the issues and challenges involved in an industrial WSN application for pervasive monitoring. The essence of Smart Grids lies in the integration of communication and networking technologies into traditional electrical power grids, with the overarching goal being an efficient and timely communications platform. Doing so unlocks critical capabilities, such as pervasive monitoring, fault diagnosis and automation over all aspects of power generation, transmission and distribution [Bose, 2010, Aggarwal et al., 2010]. Issues pertaining to voltage sags, blackouts, congestion and overloads can be addressed through tight integration between monitoring and communication technologies over what is currently an inefficient, fragile and aged electricity infrastructure [Gao et al.]. Major advantages in the use of WSNs include scalability, manageability and extensibility within the context of a large-scale and geographically distributed infrastructure, as well as a cost effective and interoperable method to perform multi-granularity monitoring with a variety of sensing options [Gungor et al., 2010, Erol-Kantarci and Mouftah, 2011a].

A multitude of different smart-grid applications provide opportunities for the effective use of WSNs in the area of distributed sensing and communications. For example, at the end-user/consumer level, there are general energy management and appliance co-ordination issues to address. From an industry perspective, there is the need to monitor transmission lines for voltage sags and the balancing of power loads across districts. To address such needs, WSNs provide streamlined sensing capabilities and an aggregated intelligence (via parallel processing) for distributed control [Erol-Kantarci and Mouftah, 2011b]. Through flexible and pervasive communications between consumer and utility controllers, WSNs are shown to effectively reduce energy expenditure, lessen contribution to both peak loads and carbon emissions [Erol Kantarci and Mouftah, 2011]. In [Liang et al., 2012], they are used for decentralized economic dispatch and an optimal decision-making process for power generation in micro-grids. They have also proven key in ensuring reliability across large-scale distributed transmission lines, through voltage sensing and meteorological monitoring [Yang et al., 2011]. The application context of WSNs in smart grid is depicted in Figure 1.2.

Therefore, underlying aspects of major industrial scale applications as represented by the example context of the smart grid application can be identified as follows: (i) involves large

amounts of sensed data; (ii) observes dynamic environments where the observation domain consists of unpredictable and dynamic distributions; (iii) heterogeneity of data with differing types in unmatched distributions; and (iv) sensing functionality leading to data-driven control and core decision making. Accordingly, there is a vital need for a secure, continuous, reliable and effective sensing mechanism through integrated WSNs, one that places utmost importance on the integrity of the data. However, in practice the sensor nodes of a WSN are **severely limited** by power, computation, memory and communication. This makes them vulnerable to faults, node malfunction and different security threats both internal and external [Djenouri et al., 2005, Shi and Perrig, 2004]. Furthermore, the physical security of the nodes cannot be guaranteed as they are often deployed in large scale environments that allow physical access and often without any tamper resistance measures due to cost considerations. The use of publicly accessible communication channels and openly commoditised wireless technologies often employed exacerbates the situation where the information exchange can be captured by any parties with malicious intent. Consequently wireless sensor networks are threatened by multiple factors which can be noted as follows [Lopez and Zhou, 2008].

- Node faults and malfunctions
- Communication attacks
- Denial of service attacks
- Node compromise and impersonation attacks
- Protocol specific attacks

Misbehaviour resulting from such occurrences will result in *anomalies* that are manifested in sensor network data streams. If left undetected these data anomalies will lead to wrong operational and control decisions that can impact wider society through severe economic losses, environmental damage and possible human harm. Therefore, it is vital that proper techniques are introduced in **accurately detecting** these anomalies given the critical role of the sensed data in core decision making.

This thesis focuses on how to effectively detect these anomalies in integrated wireless sensor network environments and mitigate their impact on key decision support systems with an aim to ensuring the security of the overall system. Of particular concern is the specific resource limitations in sensor nodes that mandate efficient procedures and models with reduced computational and communicational complexity. Novel anomaly detection models are investigated with innovative data processing frameworks that are uniquely suitable for anomaly detection in the application context of large scale industrial wireless sensor networks. Each of the presented anomaly detection models are specifically designed to overcome the challenges of unpredictable and dynamic environments in a robust manner that is both adaptable over evolving data streams and scalable with regard to large scale implementation. The first part of the research focuses on developing models that act in a completely unsupervised manner without any prior knowledge in differentiating anomalous behaviour from normal behaviour. Next, investigations are done and models introduced to classify data in a fully dynamic manner with non-parametric and non-probabilistic detection of anomalies. The final concern is to present an incremental approach for fully unsupervised anomaly detection that critically supports near real-time decision making. In the next section the specific problems associated with anomaly detection in WSNs are discussed and the challenges presented leading to the definitions of the research questions in the following section.

1.1 Research Challenges

Securing the integrity of wireless sensor network data streams is an imperative task with many significant challenges. This is especially critical when the underlying application is part of the process control system of various critical infrastructures as in oil, gas and power sectors. A successful security attack against such systems could result in potential threats to public health and safety, environmental damage and massive economic losses effecting wide aspects of society due to loss of power, manufacturing and transmission [Group, (2005, Marsh, (1997]. These systems have been designed with specific attention given to reliability and availability requirements. However, most of the time little or no attention is given to the security aspects on the system's design criteria. With process control systems increasingly

moving away from traditional wired and closed systems into more heterogeneous, open and wireless sensor network based environments, new risks and threats are appearing that cannot be countered with existing security mechanisms. Therefore, it is an extremely important issue to develop new solutions considering the specific threats and vulnerabilities from the integrated WSNs and ensure security of the overall system.

Malicious attacks exploiting the wireless communication medium on WSNs [Djenouri et al., 2005] enable for eavesdropping, illegal modification and data fabrication resulting in confidential information being available to unauthorized parties [Shi and Perrig, 2004]. These critical data could then be used against the system for further process disruptions or inflicting wide scale damage. Furthermore, an attacker with inside information can cause what are generally termed as insider attacks. These will be much more difficult to defend against, and detecting them reliably is vital. Inside attackers can modify data used in operational decisions, programs controlling industrial equipment or data reported to control centres. They could result in widespread damage, premature system shutdown or interfere with existing safety systems [Group, (2005, Marsh, (1997]. Therefore, security mechanisms that accurately and effectively counters intrusions consisting of insider attacks as well as other routing (selective forwarding, wormhole, sybil), jamming/spoofing and eavesdropping attacks in preventing critical situations is extremely vital.

The nature of openly commoditised wireless technologies often employed, and the inherent limitations of the sensor nodes themselves in naive implementations also expose the application environment to a variety of threats and vulnerabilities that are both internal and external [Djenouri et al., 2005, Shi and Perrig, 2004]. Such threats can: (i) affect the integrity of the network through path and node configurations; (ii) alter routing processes; (iii) introduce illegitimate network operations; (iv) perform illegal modifications or feed falsified data and (v) induce process monitoring blackspots [Djenouri et al., 2005, Shi and Perrig, 2004, Luo et al., 2006b, Phipatanasuphorn and Ramanathan, 2004]. Such threat vectors are easily realised through attacks on WSNs deployed for industrial equipment monitoring purposes [Mo and Sinopoli, 2012, Wang et al., 2010]. The need to deploy sensor networks unattended over a large geographic area in most applications exacerbates the situation further, with physical

security of individual nodes not guaranteed [Di Pietro et al., 2009]. This can lead to instances of side-channel attacks [Ravi et al., 2004, Bar-El et al., 2006] and compromised security keys that open windows of intrusion.

Most of the existing literature on this area focus on *prevention-based* measures. These are mostly built around cryptography with particular focus on secure key-management, data encryption and authentication [Eschenauer and Gligor, 2002, Du et al., 2004]. Some others has focused on defining more secure routing protocols [Datta, 2005, Villas et al., 2013] and managing link quality, security keys and reliability [Luo et al., 2006a, Yu and Guan, 2008]. Recent studies have focused on both probabilistic and deterministic key protocols with symmetric key cryptography [Rahman and El-Khatib, 2010, Gupta et al., 2007]. However, as the level of protection gained from these measures increase, so does the communication and computational complexity. This is not feasible in large distributed environments of industrial WSNs with their limited resources. Furthermore, as these networks cover a large geographic extent it is not always possible to guarantee the physical security of each sensor node. This can lead to physical tampering on the sensor nodes as with side channel attacks and the security keys being compromised [Ravi et al., 2004, Bar-El et al., 2006]. Secure routing methods [Boukerche et al., 2006, Ben-Othman and Yahya, 2010, Gandham et al., 2008, Mun and Shin] propose mitigating misbehaviour through stimulation of correct routing with node co-operation. These involve path rating, detecting non forwarding nodes as well as management of a shared reputation system for node behaviour. Again, it is costly in terms of energy to implement such solutions in a large scale while only offering protection for known attacks. There is always a non negligible probability for new attacks exploiting unknown vulnerabilities.

Another key concern is the unavoidable occurrence of faults in the sensor nodes that can lead to wrong measurements and impact key process parameters [Sharma et al., 2007, Sharma et al.]. In practice, isolated node failures through faults can bring down the entire network through communication disruptions. The limited resources and capability may also degenerate the quality of the data being produced. Especially when battery power is low the probability of erroneous data may grow [Subramaniam et al., 2006]. Furthermore, in instances where the sensor nodes are left unattended in harsh environments the nodes are frequently susceptible to

adverse environmental effects. In such cases the sensor nodes may malfunction and produce noisy, faulty, missing and redundant data.

Such situations as described of security attacks and faults often manifest as anomalies in data streams and pose grave implications to decision-making processes at the heart of process monitoring implementations, and can easily threaten key components of the underlying application. If left undetected these data anomalies will lead to wrong operational and control decisions on industrial infrastructure that impact wider society through power blackouts, congestion and production loss, and can ultimately lead to severe economic losses, human harm and environmental damage. Urgent focus is required for more data-centric security approaches that ensure the integrity of the sensed data. Therefore, methods for accurate and efficient detection of potential anomalies within the sensed data is of severe importance in WSN implementations. These factors have led to the design of *detection-based* mechanisms as an indispensable next level of security. Anomaly detection is defined as the process of comparing the observed behaviour with that of a built-up normal profile in order to identify any deviations. An anomaly in this context is defined as an observation that is inconsistent in significantly deviating from that of the majority normal data.

The key challenge for any evolving anomaly detection model for WSNs is to identify anomalies with high accuracy but with minimal associated energy cost. In sensor networks the majority of the energy is consumed on data communication than computation implying that a main focus should be on reducing communication overheads [Raghunathan et al., 2002]. For example it is shown that in Sensoria sensors and Berkeley motes the ratio between communication and computation energy costs range from 10^3 to 10^4 [Zhao et al., 2003]. Therefore, a relative overhead on computational costs can be favoured when coupled with a much higher reduction in communicational complexity. This leads to distributed approaches that minimize the amount of data that is communicated as well as divide the required computational cost among the different nodes in contrast to that of a centralized approach. Development of such distributed models mostly rely on an in-network data processing framework that involves performing analysis on different individual nodes and then sharing information as required to achieve the anomaly detection task.

Therefore, considering these aspects of the current state in wireless sensor network applications for large scale industrial monitoring the following core challenges are identified with regard to achieving an effective anomaly detection model.

- The resource constrained environment requiring high data processing efficiency with reduced communication and computational overheads.
- The heterogeneous observation environments with dynamic data distributions, requiring an unsupervised, non-parametric and non-probabilistic data classification method.
- The Unavailability of labelled training data sets requiring a fully unsupervised anomaly classification process.
- Non-homogeneous and multi-density data distributions within observation domains, that can result in both individual and correlated clusters of normal and anomalous behaviour.
- The need to detect anomalies in an online fashion in facilitating real-time decision support

1.2 Research Questions

In order to overcome the afore-mentioned challenges the following research questions are defined with the aim of achieving distributed anomaly detection for large scale industrial wireless sensor networks. The anomaly detection models that are developed to address these research questions are robust and adaptable with efficient and scalable data processing frameworks.

- First: How to detect anomalies accurately and efficiently in a manner that is both adaptable and scalable for large scale industrial wireless sensor networks? How to achieve that task in an unsupervised process without prior knowledge in a non-parametric and non-probabilistic fashion?
- Second: How to make the anomaly detection process fully dynamic and robust with regard to data streams with heterogeneous distributions and evolving patterns?
- Third: How to detect anomalies in an incremental model that supports near real-time decision making in large scale industrial wireless sensor networks?

1.3 Limitations of Existing Solutions

There exists a wide range of anomaly detection methods [Chandola et al., b], [Chandola et al., 2012], [Patcha and Park, 2007] that have been proposed specific to different application areas. Most of these techniques are not presented in relation to WSNs and assume the availability of significant computational resources as well as centralized access to all data under the interested domain [da Silva et al., 2005]. Such computationally and communicationally complex algorithms cannot directly be applied in WSNs due to their inherent resource constraints [Xu et al., 2010, Liu et al., 2011]. Therefore, distributed data processing frameworks are mandated that achieves anomaly detection through localized processing on individual sensor nodes with intermediate information sharing. While other distributed data clustering frameworks [Kargupta et al., 2001, Tasoulis and Vrahatis, 2004] exist regarding different applications, these only attempt to solve the problem of distributing/parallelizing the clustering process for *large data sets* that exist in a single server and is largely homogeneous and static. The aim is only to reduce computational loads and doesn't address the fundamental issue of data that is generated distributively and in a dynamic environment as in WSNs. The very limited number of such methods for WSNs [Nowak, 2003, Forero et al., 2011] are also severely handicapped as they assume an *independent and identicaly distributed (i.i.d)* condition for all data and only creates a global clustering considering the whole data domain as a single entity while ignoring any dynamicity in the environment.

Previous attempts at anomaly detection aimed specifically towards WSNs exist [Zhang et al., 2010, Chatzigiannakis and Papavassiliou, 2007, Onat and Miri, 2005, Xie et al., 2011]. However, almost all of them either focus on detecting/preventing specific attacks or consider sensor networks in a generalized context with assumptions made on major limitations. Existing approaches can be classified onto different classes as (i) Distribution based, (ii) Distance based, (iii) Density based and (iv) Clustering based approaches. Of these, approximating distribution models require significant prior knowledge for data fitting as well as expensive testing/refitting. This is not practical in the sense of dynamically changing data and the low resource capacity environment of WSNs. On the other hand, distance based approaches [Knorr et al., 2000] identify outliers based on the fraction (β) of the data that is further than a defined distance

(r) to the majority. However, this can lead to misclassifications and detection problems where the observation domain consists of both sparse and dense distributions within a single time window.

Density based techniques mine outliers depending on the local density of each observation [Breunig et al., 2000, Xie et al., 2012]. The local neighborhood is typically defined by a distance (r) that encompasses the K nearest data points. While this approach handles multi-density distributions with minimum assumptions/parameters, it suffers in the context of anomalies comprising of uniform density micro-clusters similar to normal data. In such instances the parameter K will have to be defined as large as the size of these clusters. Thus, sensitivity to this single parameter brings unreliability and reduced detection rates in the context of observations that comprise of isolated outliers and uniform density micro-clusters of abnormal behaviour. In [Papadimitriou et al., 2003], these issues are addressed using a *local correlation integral* based on statistical measures derived through local neighborhood densities. This identifies outliers non-parametrically and is robust considering dynamic distributions with different densities. However, it cannot detect instances of abnormal behaviour (consisting of micro-clusters) that has same density as normal data while being significantly different in individual values.

Data partitioning/clustering approaches stand out as the most viable in this context when implemented with a non-parametric approach. They offer fully unsupervised classification without any prior knowledge/assumptions and can be adapted over an incremental model that offers different granularity. This is best attained in a hierarchical network topology that performs the partitioning distributively and evaluates anomalies incrementally over different levels. However, existing data clustering approaches are not optimized for anomaly detection and are mostly centralized processes with high complexity and resource costs. Furthermore, inferring outliers through an integrated formula cannot easily be achieved as the explicit goal is segregating data groups with only implicit attention on the effects of noise/outliers [Xu and Wunsch, 2005, Rokach, 2010].

Considering these limitations this thesis focuses on advance anomaly detection models that work in an unsupervised process in order to be non-parametric and non-probabilistic. The

models are specifically implemented over distributed in-network data processing frameworks that are uniquely suitable for the resource constrained environment of large scale industrial wireless sensor networks.

1.4 Research Contributions

To address the research questions defined previously three different frameworks are proposed each with its own unique models that enable efficient in-network anomaly detection for WSNs.

1.4.1 Contribution #1

The approach proposed in the first framework uses distributed anomaly detection to achieve high efficiency on the WSN, in accurately identifying intrusions and faults that compromise the security of the overall system. The issues presented in the first research question are addressed here with an innovative approach that effectively utilize in-network processing in a distributed anomaly detection framework. Fuzzy c-means clustering [Bezdek et al., 1984] is adapted in an incremental manner for unsupervised data partitioning over several analytical stages with regard to a hierarchical node topology. Thresholds are introduced adaptively at each of these stages in differentiating local anomalies from the fuzzy membership scores and inter-cluster distances. These thresholds are determined adaptively using second order statistical knowledge that is available at each analytic stage. Global correlations are evaluated incrementally over the node hierarchy in identifying global anomalies as the network domain expand. The approach is extensively evaluated using different data distributions with spatially sparse and dense representations of normal and anomalous behaviour. These data distributions are derived from the publicly available sensor data distributions from Intel Research Laboratories [Bodik et al., 2004] and the Australian Research Council's research network on Intelligent Sensors, Sensor Networks and Information Processing (*ISSNIP*) [Suthaharan et al., 2010]. Experimental investigations are performed in calculating the detection accuracy and complexity in terms of communication overheads. Detailed comparisons are done with regard to a centralized data processing approach and the existing fixed-width based data clustering

approach in [Rajasegarar et al., 2006]. The major contributions of this research are highlighted below.

- *A distributed data processing algorithm for efficient anomaly detection*: This uses distributed in-network processing over a hierarchical sensor node topology. Centroidal-based fuzzy data clustering is performed at each node while communicating only the locally identified cluster centroids and corresponding outliers to the next hierarchical level in reducing communication overheads. The use of a centroid based data clustering approach enables an efficient incremental model where data can be processed as they are sensed, and evaluated for anomalies at different stages on the hierarchical topology. Experimental results demonstrate significant communication reductions compared to existing centralized solutions [Chatzigiannakis and Papavassiliou, 2007, an Huang et al., 2003].

- *An unsupervised classification method for heterogeneous and dynamic data distributions in WSNs*: This is a distributed adaptation of the fuzzy c-means algorithm [Bezdek et al., 1984], where fuzzy membership scores are computed by evaluating global correlations over an incremental model. This is performed at several analytical stages over a hierarchical topology enabling the model to accurately detect both local and global anomalies. This reduces misclassifications compared to other binary classification methods [Onat and Miri, 2005, Eik Loo et al., 2006], which are deterministic as well as restricted to local data correlations. Experimental results clearly demonstrate reduced false positives with the sensitivity range increasing compared to a non-fuzzy fixed-width clustering scheme [Rajasegarar et al., 2006].

- *Non-parametric and non-probabilistic detection for local and global anomalies*: The thresholds for identifying local and global anomalies, are derived based on second order statistical knowledge of *mean* and *standard deviation*. These are introduced to fuzzy membership scores and inter-cluster distance distributions adaptively at different analytical stages. Therefore, unlike in [Onat and Miri, 2005] and [Rajasegarar et al., 2006], no arbitrary definitions of parameters or probabilistical assumptions are made, leading to a

robust and more *accurate* detection of anomalies. Experimental results show an average sensitivity in detection accuracy of (83.44 – 95.1%) and an average specificity in detection accuracy of (99.73 – 99.98%) compared to (12 – 48%) and (14 – 72%) for the existing data clustering approach in [Rajasegarar et al., 2006].

1.4.2 Contribution #2

The challenges with regard to the second research question are addressed here with an innovative data processing framework that detects anomalies at different levels on a hierarchical topology. In-network data processing is performed over several analytical stages, working to reduce the amount of data communicated as well as offer granular detection. The data space on each sensor node with regard to a specific time window ΔT, is partitioned in to different cohesive regions using an entropy criterion that captures data characteristics above second order statistics. This is performed in a non-parametric and completely unsupervised manner. Unlike in the previous model the number of data partitions are determined dynamically. Data correlations are therefore, compared through a *point-wise entropy* measure that is further complemented by a *relative density* factor for a defined local neighbourhood at each data point. The effect of noise and isolated outliers on the partitioning process is effectively curtailed through their prior segregation using the relative density measure. Thresholds are then applied adaptively on each analytical stage in identifying anomalous data regions as well as individual outliers. The resultant local anomalies together with average entropy and corresponding density information in identified regions, are then communicated to the next hierarchical level, where more global correlations are evaluated and representative anomalies at that level of granularity identified. This is performed at all the different network levels until the root of the hierarchy is reached for the considered WSN. Experimental investigations are performed to calculate the detection accuracy for both normal and anomalous data points through *sensitivity* and *specificity* analysis as well as communication complexity compared to a centralized data processing approach. The major contributions of the research are summarised below:

- *A dynamic and unsupervised data partitioning method for heterogeneous and dynamic data*: This partitions the data into different cohesive regions, based on correlations iden-

tified through a point-wise entropy measure using euclidean distance similarity. The effects of noise and outliers is mitigated through prior segregation of such through the use of average relative density calculations defined over two localised neighbourhood levels. Unlike in [Gokcay and Principe, 2002, Temel and Aydin, 2007], entropy is used directly in determining the partition formations and significantly the number of partitions are determined dynamically without prior knowledge in contrast to most existing methods [Rokach, 2010]. This is performed at each node for its local observation domain over a hierarchical topology, enabling different levels of granularity. The resulting partitions are representative for anomaly detection while being sensitive to local density variations and robust for dynamic distributions. Experimental results demonstrate high classification accuracy of more than 94% in both sensitivity and specificity metrics consistently over different data distributions.

- *Non-parametric and non-probabilistic detection of anomalies that is robust and adaptive over different granular levels*: This identifies anomalies using adaptive thresholds derived from second order statistical information for observed mean entropy measures and corresponding average relative densities over different analytical stages. Each stage focuses only on domain knowledge available at that network level in avoiding any arbitrary definitions of parameters or probabilistic assumptions. This leads to a more robust and accurate detection for dynamic data with differential density compared to existing methods as in [Chandola et al., a, Hodge and Austin, 2004]. Experimental results show higher averages in the detected *true positives* and *true negatives* with a corresponding reduction in *false positives* and *false negatives*.

- *A distributed data processing framework for efficient and scalable anomaly detection*: This capitalizes on in-network data processing performed in parallel on each sensor node, where local data is aggregated over a defined sliding window. Entropy-based data partitioning is performed locally while communicating only the identified local anomalies, and representative information (of mean entropy and related average relative density)

on defined partitions, to the next hierarchical level. Thus, communication overheads are dramatically reduced leading to enhanced efficiency, while the use of a data-partitioning method enables an incremental model where data is processed as they are sensed and evaluated for anomalies on different granular levels. Experimental results demonstrate an average communication reduction of more than 85% compared to a centralized approach where analysis is performed globally.

1.4.3 Contribution #3

The third contribution addresses the issue of enabling near real-time anomaly detection in an incremental approach in relation to the concerns in research question three. This is again performed through the effective use of an entropy criterion to differentiate anomalies. The proposed model is implemented in-network with data processing performed in an online fashion. This supports the real time nature of a typical WSN monitoring application in mitigating negative impacts of batch models. Specifically, in the proposed model a small buffer is maintained in relation to incoming data streams and according to memory constraints at each node. This data buffer is continuously updated with the addition of the latest data point and the removal of the oldest. With the arrival of each new data point the point-wise entropy is calculated according to three unique Point-of-View (POV) approaches. These are introduced as (i) PoV of the mean μ, (ii) PoV of the incoming data point η and (iii) PoV of the historic mean μ' (without the influence of latest data point). Then the change in entropy with the advent of a new data point is is evaluated in relation to the current buffer elements of each sensor node in identifying anomalous points.

However, in order to accommodate instances where sudden sensor drifts or level shifts make the observed data distribution jump to a new range before continuing normal behaviour (in that range), a secondary analysis is performed for potential anomalies identified in the first stage. Therefore, a secondary buffer is introduced and any element that is designated as an anomaly in the first stage is temporarily stored there. If the potential anomalies are continuously identified the secondary analysis will be performed on this buffer to evaluate entropy change

within the context of these new measurements. This gives the option to identify a major change in measurement range that can be normal within a particular application environment such as voltage monitoring where the measured voltage may jump between ranges as part of normal behaviour. The accuracy of detection is extensively investigated and performance evaluated over different data distributions using sensitivity and specificity analysis. This is performed with regard to classifying normal data when the data stream evolves with evident sensor drifts and with regard to detecting anomalies while the normal data stream continues to evolve. A detailed comparison to other existing approaches is also done. The major contributions of this research can be summarised as follows.

- Dynamic evaluation of anomalies in an online fashion for evolving data streams. This uniquely detects outliers and abnormal behaviour in data streams while accommodating for dynamic and evolving patterns of normal behaviour. A dual buffer model is used in facilitating the detection of sudden behaviour changes in the data streams that may happen as part of normal activity within the observation environment and enabling realtime decision support. Experiment results reveal high accuracy of more than 98% in average sensitivity in detecting abnormal behaviour while still being robust with reduced misclassifications as the different data streams evolve dynamically.

- PoV approach in evaluating relative change in entropy for dynamic detection of abnormal behaviour. This uses point-wise entropy and its relative change as compared to the mean entropy value observed over the buffered data space over evolving data streams to uniquely capture the abnormalities and outliers that are present. As such three PoV approaches are proposed as different lenses providing insight for more accurate detection. Therefore, the perspectives of the mean μ, the median η and the historic mean μ' with regard to the maintained data buffer is used. The experimental results show improved accuracy in sensitivity (more than 98% in average) and specificity (more than 96% in average) metrics with regard to these PoV's in facilitating anomaly detection in an online fashion.

1.5 Thesis Organization

The remaining chapters of this thesis are organized as follows.

- **Chapter 2: Hierarchical Data Partitioning with Fuzzy Data Modelling for Scalable Anomaly Detection**. This chapter presents the unsupervised data partitioning approach that is developed using fuzzy c-means clustering in an incremental model towards scalable anomaly detection over a hierarchical sensor network topology. Non-parametric and non-probabilistic detection methods are introduced in a distributed and in-network data processing framework. Detailed evaluation of the proposed model is presented with quantitative and qualitative comparisons to existing work.

- **Chapter 3: Dynamic Data Partitioning with an Entropy Criterion for Multi-granular Anomaly Detection**. The entropy based dynamic data partitioning approach for WSN anomaly detection is presented in this chapter. The unique advantages of an entropy criterion to partition data in the context of anomaly detection is discussed in relation to limitations in existing work. The model is evaluated over different data sets extensively with regard to accuracy and communication complexity studies.

- **Chapter 4: Point-of-View (POV) Entropy Evaluations for Real-time Decision Support in Evolving Data Streams**. An in-network data processing model that uses a point-wise entropy criterion to identify data anomalies in an online fashion is presented here. Three unique reference points are introduced as part of a Point-of-View (PoV) approach that computes entropy and offer different lenses to evaluate its relative change as the data stream evolve on each sensor node. In order to identify instances where the data stream may suddenly shift its spatial distribution as part of normal behaviour a secondary analysis is performed employing a secondary data buffer that stores potential anomalies identified in the primary phase of detection. The approach is evaluated for classifying normal data as the distribution evolve as well as anomalous data that is aberrant in different degrees. Detailed comparisons to existing work are also presented

- **Chapter 5: Conclusion**. This chapter concludes the thesis with a summary on major contributions and key findings of the performed investigations. Potential areas for future research along with some limitations of the proposed models are also presented.

Note: The three core chapters (chapters 2 - 4) are presented in a standalone and self-contained manner as much as possible in enhancing accessibility to major theoretical contributions. Therefore, the relevant context including discussions on related work, descriptions on data sets used and evaluation metrics are presented in each of these chapters separately in the relevant discussion vein specific for that chapter

Chapter 2

Hierarchical Data Partitioning with Fuzzy Data Modelling for Scalable Anomaly Detection (HDP-FM)

As discussed in Chapter 1, Wireless Sensor Networks (WSNs) have increasingly gained prominence over a variety of applications with regard to providing an efficient and cost effective platform for pervasive monitoring. However, due to the nature of the WSNs themselves including resource limitations and modes of implementation they are vulnerable to different kinds of security attacks, malfunctions and faults. Taken together these issues impact the integrity of the sensed and communicated data in undermining the decision making processes that rely on the continuous availability of accurate monitoring data.

In addressing this issue and in relation to the first research question as presented in Section 1.2, this chapter introduces a new data partitioning model based on fuzzy data modelling in order to detect sensed data anomalies in a robust and scalable framework. Data processing is performed in a distributed manner with efficient in-network procedures over a hierarchical node topology that is adapted as the network model. Unsupervised data partitioning is performed distributively by adapting *fuzzy c- means* clustering in an incremental model. Non-probabilistic anomaly detection is then performed through fuzzy membership evaluations over the resulting

data clusters and through thresholds on observed inter-cluster distances. Thresholds are determined adaptively in a dynamic manner using second order statistical knowledge available at each analysis stage on the hierarchy. The approach is made to be as non-parametric as possible with such adaptive measures taken where any fixed thresholds are used.

The rest of the chapter is organized as follows. First, Section 2.1 presents the background for the study consisting of the main research problems and the challenges that are to be overcome within the relevant context. Limitations of existing solutions are also discussed. Further, we specifically highlight the motivation for the research and the contributions of the proposed model in addressing the identified research issues. Section 2.2 extensively analyses and discusses existing approaches and their drawbacks with regard to anomaly detection in the current context as well as specific limitations in identified related work. The proposed data partitioning approach based on fuzzy data modelling for distributed anomaly detection is then presented in detail in Section 2.3. It focuses separate sections in soft data partitioning with fuzzy clustering that include distributed and local phases of clustering, anomaly detection where detection of both individual outliers and anomalous clusters are presented as well as a section presenting the overall algorithmic framework. Extensive experiments are performed using a variety of data distributions representing different aspects of observed behaviour in Section 2.4. Here the different sub sections focus on the nature of the data sets used, evaluation criteria, detailed discussions on results for detection accuracy and communication efficiency and finally on a comparative analysis. The chapter is summarised in Section 2.5

2.1 Motivation and Contributions

Modern industrial processes in the sectors of energy, utilities and manufacturing increasingly depend on Supervisory Control and Data Acquisition (SCADA) systems for process control. Most of these systems are large and distributed, with the geographic range extending in the order of kilometres for typical power, water and oil distribution applications. In such systems wireless sensor networks (WSNs) provide a low cost and flexible solution to sensing and monitoring. They provide a large number of low cost sensors in a variety of sensing options enabling more fine grained process monitoring. It is more feasible in terms of economical cost to deploy

them flexibly, while increasing the sensing ability compared to their wired counterparts [Roman et al., 2007, Ye and Heidemann, 2006]. However, due to their inherent limitations in terms of power and communication bandwidth, as well as vulnerabilities pertaining to the wireless nature of communication, they add to the security risks of process control (SCADA) systems [Djenouri et al., 2005].

Ensuring the security of industrial process control systems is vital given the critical aspects of the underlying infrastructure. A successful security attack against such systems could result in potential threats to public health and safety, environmental damage and massive economic losses effecting wide aspects of society due to loss of power, manufacturing and transmission [Group, (2005, Marsh, (1997]. These systems have been designed with specific attention given to reliability and availability requirements. However, most of the time little or no attention is given to the security aspects on the system's design criteria. With process control systems increasingly moving away from traditional wired and closed systems into more heterogeneous, open and wireless environments, new risks and threats are appearing that cannot be countered with existing security mechanisms. Therefore, it is an extremely important issue to develop new solutions considering the specific threats and vulnerabilities from the integrated WSNs and ensure security of the overall system.

Malicious attacks exploiting the wireless communication medium on WSNs [Djenouri et al., 2005] enable for eavesdropping, illegal modification and data fabrication resulting in confidential information being available to unauthorized parties [Shi and Perrig, 2004]. These critical data could then be used against the system for further process disruptions or inflicting wide scale damage. Furthermore, an attacker with inside information can cause what are generally termed as insider attacks. These will be much more difficult to defend against, and detecting them reliably is vital. Inside attackers can modify data used in operational decisions, programs controlling industrial equipment or data reported to control centres. They could result in widespread damage, premature system shutdown or interfere with existing safety systems [Group, (2005, Marsh, (1997]. Therefore, security mechanisms that accurately and effectively counters intrusions consisting of insider attacks as well as other routing (selective forwarding, wormhole, sybil), jamming/spoofing and eavesdropping attacks in preventing critical situations

is extremely vital.

Most of the research focus on securing WSNs have been on intrusion prevention systems. These involve designing more secure communications through encryption and authentication as well as secure routing protocols. Recent studies have focused on both probabilistic and deterministic key protocols with symmetric key cryptography [Rahman and El-Khatib, 2010, Gupta et al., 2007]. However, as the level of protection gained from these measures increase, so does the communication and computational complexity. This is not feasible in large distributed environments of industrial WSNs with their limited resources. Furthermore, as these networks cover a large geographic extent it is not always possible to guarantee the physical security of each sensor node. This can lead to physical tampering on the sensor nodes as with side channel attacks and the security keys being compromised [Ravi et al., 2004, Bar-El et al., 2006]. Secure routing methods [Boukerche et al., 2006, Ben-Othman and Yahya, 2010, Gandham et al., 2008, Mun and Shin] propose mitigating misbehaviour through stimulation of correct routing with node co-operation. These involve path rating, detecting non forwarding nodes as well as management of a shared reputation system for node behaviour. Again, it is costly in terms of energy to implement such solutions in a large scale while only offering protection for known attacks. There is always a non negligible probability for new attacks exploiting unknown vulnerabilities. These factors have led to the design of intrusion detection mechanisms as an indispensable next level of security

Contributions

The approach proposed in this chapter uses distributed anomaly detection over a hierarchical node topology to achieve high efficiency on the WSN, in accurately identifying intrusions and faults that compromise the security of the overall system. The following are some of the major challenges addressed by the proposed approach, which are not satisfactoraly addressed in existing solutions [Zhang et al., 2010, Chatzigiannakis and Papavassiliou, 2007, Onat and Miri, 2005, Xie et al., 2011]:

- Resource constrained environment requiring high data processing efficiency with reduced communication and computational overheads.

- Unpredictable dynamic changes in the monitored environment requiring non-parametric anomaly detection without prior knowledge of the data distribution.

- Unavailability of labelled training data sets requiring a fully unsupervised anomaly classification process.

These challenges have been addressed here with an innovative data partitioning approach that works in an unsupervised manner. Data is processed in-network in a distributed anomaly detection framework that effectively reduce associated communication costs. Fuzzy c-means clustering [Bezdek et al., 1984] is adapted incrementally for unsupervised data partitioning, on several analytical phases over a hierarchical node topology. Non-parametric anomaly detection is performed based on a robust thresholding technique defined using second order statistical knowledge. The thresholds are introduced adaptively at each data processing stage in evaluating local outliers from the calculated fuzzy membership scores and inter-cluster distances. Global anomalies are identified through the evaluation of global correlations in an incremental model over the node hierarchy. The proposed approach is evaluated using different data distributions with both sparse and densely concentrated representations of normal and abnormal behaviour. Two major data sets are derived based on the publicly available sensor data distributions from Intel Research Laborataries [Bodik et al., 2004] and the Australian Research Council's research network on Intelligent Sensors, Sensor Networks and Information Processing (*ISSNIP*) [Suthaharan et al., 2010]. Experiments are performed to calculate both the detection accuracy (for normal/abnormal data instances) as well as complexity in terms of communication overheads. Detailed comparisons are done with regard to a centralized approach and the existing data clustering approach in [Rajasegarar et al., 2006]. The major technical contributions of this research are highlighted below.

- *A distributed and in-network data processing algorithm*: Data is processed distributively at each node over an assumed hierarchical sensor node topology. Centroidal-based fuzzy clustering that is performed at each node communicates only the identified local centroids and corresponding anomalies between two network levels. Communication overheads are reduced by using a centroidal based data clustering approach where the data can be

processed as they are sensed, and evaluated for anomalies over different network levels in a hierarchical topology. Experimental results demonstrate an average communication reduction range of 98 – 99% compared to existing centralized solutions [Chatzigiannakis and Papavassiliou, 2007, an Huang et al., 2003].

- *An unsupervised classification method for heterogeneous and dynamic data distributions*: The fuzzy c-means algorithm [Bezdek et al., 1984], is adapted in a distributed manner to score fuzzy memberships in an incremental model that evaluates increasingly global correlations over a hierarchical network topology. The different analytical stages corresponding to different network levels allow the model to accurately detect both local and global anomalies. Misclassifications are reduced compared to other binary classification methods [Onat and Miri, 2005, Eik Loo et al., 2006], which are deterministic as well as restricted to local data correlations. Experimental results clearly demonstrate reduced false positives with the sensitivity range increasing from (12% - 48%) to (83.44% - 95.1%) compared to a non-fuzzy fixed-width clustering scheme [Rajasegarar et al., 2006].

- *Non-parametric and non-probabilistic detection of anomalies*: Thresholds are introduced on calculated fuzzy membership scores and inter-cluster distance distributions adaptively at different analytical stages in identifying both local and global anomalies. Second order statistical knowledge of *mean* and *standard deviation* is used in deriving thresholds without relying on any probabilitical or parametric methods. Therefore, unlike in [Onat and Miri, 2005] and [Rajasegarar et al., 2006], no arbitrary definitions of parameters or probabilistical assumptions are required. This leads to a *robust* and more *accurate* detection of anomalies that is scalable. Experimental results show an average sensitivity in detection accuracy of (83.44 – 95.1%) and an average specificity in detection accuracy of (99.73 – 99.98%) compared to (12 – 48%) and (14 – 72%) for the existing data clustering approach in [Rajasegarar et al., 2006].

2.2 Related Work & Rationale

The rationale for adapting distributed anomaly detection in an unsupervised and non-parametric approach to identifying anomalous behaviour in WSNs is discussed here. The limitations of existing work are presented and focus is build upon the unique approach to be implemented.

Anomaly detection, as a branch of intrusion detection, identifies abnormal behaviour without prior knowledge on the nature of that behaviour. Therefore, anomalies are identified as measurements that significantly deviate from an established profile for normal behaviour within a particular spatio-temporal domain. This enables the detection of new types of attacks and emergent abnormal behaviour in the system through an incremental model that profiles the normal behaviour. This is preferable in the sense of industrial WSNs due to scalability and flexibility in adaption for dynamically changing data distributions and large scale implementations [Raghunathan et al., 2002, Pottie and Kaiser, b].

Existing anomaly detection techniques can generally be identified under two major classes. The first class uses *unsupervised learning* to build a normal profile that is used to identify anomalous outliers without any prior knowledge on observed data. The second class is based on *supervised learning* that makes use of prior knowledge in building up a normal profile. This method requires the use of a training data set with labeled annotations. However, in the proposed application context, such training data sets are unavailable with dynamic data distribution changes requiring the classifier to be retrained each time with new labeled data. Therefore, no prior knowledge can reliably be assumed, requiring an unsupervised classification with the identification of outliers being both non-parametric and non-probabilistic regarding the observed distribution.

Most of the existing anomaly detection techniques assume the availability of significant computational resources as well as centralized access to all data under the interested domain [da Silva et al., 2005]. Such computationally and communicationally complex algorithms cannot directly be applied in WSNs due to their inherent resource constraints [Xu et al., 2010, Liu et al., 2011]. Therefore, distributed data processing frameworks are mandated that achieves anomaly detection through localized processing on individual sensor nodes with intermediate information sharing. While other distributed data clustering frameworks [Kargupta et al., 2001,

Tasoulis and Vrahatis, 2004] exist regarding different applications, these only attempt to solve the problem of distributing/parallelizing the clustering process for *large data sets* that exist in a single server and is largely homogeneous and static. The aim is only to reduce computational loads and doesn't address the fundamental issue of data that is generated distributively and in a dynamic environment as in WSNs. The very limited number of such methods for WSNs [Nowak, 2003, Forero et al., 2011] are also severely hadicapped as they assume an *independent and identicaly distributed (i.i.d)* condition for all data and only creates a global clustering considering the whole data domain as a single entity while ignoring any dynamicity in the environment.

Previous attempts at anomaly detection aimed specifically towards WSNs exist [Zhang et al., 2010, Chatzigiannakis and Papavassiliou, 2007, Onat and Miri, 2005, Xie et al., 2011]. However, almost all of them either focus on detecting/preventing specific attacks or consider sensor networks in a generalized context with assupmtions made on major limitations. Chatzigiannakis et al [Chatzigiannakis and Papavassiliou, 2007] proposed a scheme based on PCA (Principal Component Analysis) and the subspace method. However, lacking a distributed approach, data is processed centrally leading to high communication overheads. Onat et al [Onat and Miri, 2005] introduced a method based on predefined statistical models for neighbouring nodes. A normal profile is built upon average recieve power and average packet arrival rates. This is heavily limited by the assumptions on normal behaviour probabilities that define the thresholds for deviations from the normal profile. Huang et al [an Huang et al., 2003] proposed a cross feature analysis measure but limits itself to only identifying routing anomalies based on feature correlations that can be selective. A clustering technique is used by Loo et al in [Eik Loo et al., 2006] for detecting routing attacks by creating fixed width clusters on routing records. However, the determination of cluster boundaries for outliers is not clear when dynamic changes occur in the observed data. Rajasegarar et al [Rajasegarar et al., 2006] provided a distributed clustering approach that overcomes most of the limitations mentioned above. The clustering provides a means of unsupervised anomaly detection while the distributed framework largely reduces the communication costs. However, the use of fixed width clusters has an impact on accuracy when defining an appropriate width for the clusters. Anomalies are

also evaluated based on a threshold for inter-cluster distance on a K-nearest neighbourhood algorithm. The parameter K cannot easily be defined and can be arbitary given the dynamic changes to data that can be expected.

2.3 Distributed Anomaly Detection Based on Incremental Fuzzy Cluster Evaluation

Considering the limitations in existing methods, a novel anomaly detection framework is required that deals with the specific constraints of sensor nodes and challenges pertaining to the dynamic nature of the sensed data itself. The proposed approach takes these aspects into consideration to identify both standalone and distributed attacks through local and global data correlations in a co-operative framework. It considers dependencies between attributes of locally sensed data as well as the spatial and temporal correlations among neighbouring nodes in a distributed environment.

The proposed approach can be summarised as follows. A WSN is modelled as a hierarchical topology of sensor motes with several tiers offering different levels of granularity (see Figure 3.4). This organization of the network with the selection of parent nodes on each level can be achieved using any of the techniques proposed in the literature [Ganesan et al., b, Malpani et al., 2000]. Each node collects multidimensional data measurements from the observed environment and aggregates them locally within a fixed window of time (Δt). The nodes are considered to be time synchronised and maybe deployed in either a homogeneous or heterogeneous environment. Therefore, observations can constitute either same or different data distributions which are unknown and cannot be predetermined.

The framework identifies both *local* and *global* anomalies using an unsupervised and non-parametric process: *l*ocal anomalies are identified using correlations/similarities among data within a single node, whereas *g*lobal anomalies are identified considering correlations/similarities on the union set of measurements representing multiple sensor nodes on the network. Detection of global anomalies require the knowledge of all data observations sensed within that specific spatial and temporal domain, which incurs large energy intensive communication overheads,

when adapted in traditional centralized schemes [Zhang et al., 2010, Onat and Miri, 2005]. The proposed framework overcomes this through a distributed data clustering approach that utilize in-network processing, in an incremental model for a hierarchical WSN topology.

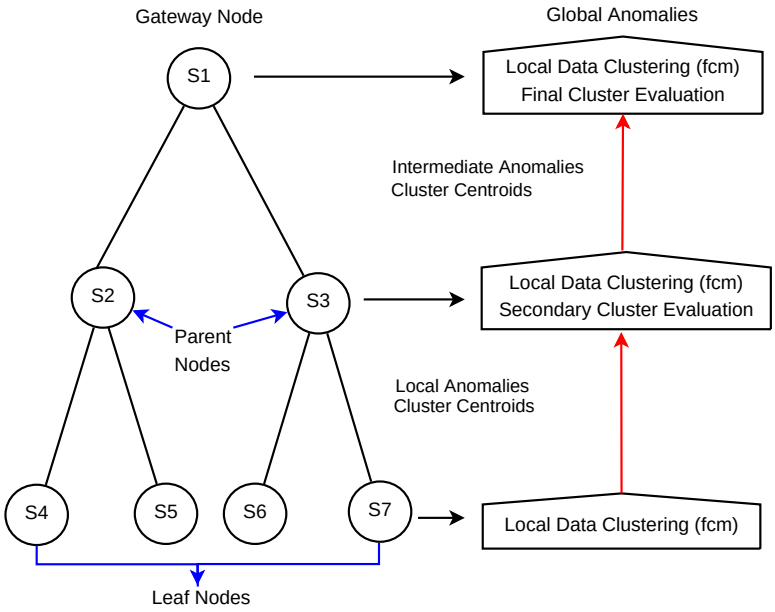

Figure 2.1: Distributed Anomaly Detection Architecture on a 2 Level WSN Hierarchy

The core of the framework consists of a distributed data clustering algorithm, adapting fuzzy logic and fuzzy set theory concepts to accurately model the normal behaviour of the sensed data space. First, local data partitioning is performed using *fuzzy c-means (fcm)* clustering, and statistical thresholds are adaptively introduced to precisely classify the data points (observations) into clusters and identify local outliers using local correlations at each node. Later, the resulting cluster centroids and outlying data points are communicated over the network (node) hierarchy to the next level. Here, the data points are again evaluated for cluster memberships considering more global correlations on the union set of data points at each level. This is performed recursively until the gateway node enabling global anomalies to be detected using the thresholds on membership values computed at this (global) level.

Figure 2.1 graphically summarises the architecture of the proposed distributed anomaly detection framework considering a two level hierarchical topology with seven sensor nodes. Nodes S_4, S_5, S_6 and S_7 are the leaf nodes with nodes S_2 and S_3 as their immediate parents

with node S_1 making up the final gateway node for the assumed hierarchy. Each of the main algorithmic steps that are performed at each node as well as the communicated data between each level are shown with respect to the relavant node and hierarchical level.

2.3.1 Soft Partitioning and Fuzzy Clustering

A sensor network data space consisting of multidimensional attributes cannot directly be partitioned into a set of disjoint clusters. Therefore, any hard assigning of individual observations into specific clusters will be arbitrary. In such cases a soft-partitioning of the data space is preferable using weights to each observation and cluster indicating the degree to which that observation belongs to any particular cluster. The value w_{ij} is then used as the weight that any observation (say X_i) belongs to a particular cluster C_j. Fuzzy clustering provides a non-probabilistic assignment of weights w_{ij}, which represent membership scores based on *fuzzy logic* and *fuzzy set theory* through distance dissimilarity. The main concept in fuzzy set theory in this context is that a particular observation is allowed to belong to a particular cluster with a membership value between 0 an 1, while a pertaining statement derived through fuzzy logic can be true to a degree between 0 and 1.

The fuzzy partitioning of a concerned multidimensional data space for a WSN can be explained as follows. Let us consider a set of observations $X = [X_1, X_2, ..., X_n]$, where each data point X_i is a d-dimensional observation, where $X_i = (x_{i1}, x_{i2}, ..., x_{id})$. Then a group of fuzzy clusters $C_1, C_2, ..., C_k$ is a subset of all possible fuzzy subsets of X where

- The summation of weights for a particular point add up to 1, i.e. $\sum_{j=1}^{k} w_{ij} = 1$.
- Each cluster consists with non-zero weight atleast one data point, and doesn't consist with a weight of 1, all the points. We have: $0 < \sum_{i=1}^{n} w_{ij} < n$.

In this context the fuzzy c-means (*fcm*) algorithm [Bezdek et al., 1984] is versatile in defining a fuzzy psuedo partition for a WSN. It minimizes an objective function (denoted as J_m), which is the weighted sum of squared errors,

$$J_m(U, C; X) = \sum_{j=1}^{k} \sum_{i=1}^{n} w_{ij}^m ||x_i - c_j||_A^2, 1 < m < \infty$$

where $C = (c_1, c_2, ..., c_k)$ is a vector of unknown cluster prototypes and $c_j \in R^p$. Here w_{ij} is the membership degree for data point X_i in the j^{th} cluster. The inner product given by the norm matrix A defines the similarity measure between a particular data point and the cluster centroid (center). The Euclidean Distance is chosen as the similarity measure, as it provides an effective similarity score with comparitively low computational complexity. Therefore a nondegenerate fuzzy psuedo partition of X is derived by the representing matrix $U = [w_{ij}]$.

Focusing on centroid-based clustering to effectively capture the structure of the data is vital to having an efficient distributed approach. Therefore, cluster centroids are used to represent the set of individual data points for that cluster at different stages in the proposed model. This reduces both the amount of data communicated among different levels on the hierarchy as well as computational costs involved with their analysis. The main stages in our data clustering approach can be represented in two phases as follows.

2.3.1.1 Local Clustering Phase

Soft partitioning of the data space is done using fuzzy clustering considering a local neighbourhood of one hop distance at each analysis level over the concerned hierarchy. Accordingly, all of the sensor nodes in the considered sensor network will locally cluster its own data space for a specific temporal window of time (Δt). This is the initial step in the proposed algorithm and is performed using *fuzzy c-means* with a user defined number of expected clusters. Therefore, each node produces a set of fuzzy values w_{ij} with regard to the degree of membership towards the determined number of clusters C_K. Next, de-fuzzification of the soft clustering is performed through the introduction of statistically derived thresholds for the cluster membership values of each data point. These (thresholds) will be non-probabilistic and non-parametric in the sense that they will not be based on any prior knowledge of the data distribution. The thresholds are derived from the statistical features of the observed data itself in truly representing the currently observed distribution in an incremental model, rather than relying on hard coded values that are arbitrary in a dynamic environment. Therefore, we use the value of one *stan-*

dard deviation away from the *mean* as the threshold (T) for defining cluster memberships. T is defined as follows:

$$T(T_1, T_2) = 1/n \sum_{i=1}^{n} X_i \pm \sqrt{1/n \sum_{i=1}^{n} [X_i - (1/n \sum_{i=1}^{n} X_i)^2]}$$

T_1, T_2 is applied on the set of maximum values of w_{ij} computed for each data point. Therefore, each node identifies the relevant set of clusters, denoted as $C_1, C_2, \cdots, C_k$, and their members (with membership between thresholds). It also identifies the set of *l*ocal outliers that cannot be attributed to a particular cluster, with a satisfactory degree (membership outside of thresholds).

2.3.1.2 Distributed Clustering Phase

After the local clustering phase, every node sends the representative cluster centroids and the identified local outliers to each of their immediate parent nodes on the hierarchy. The number of these anomalies and centroids will be largely lesser in extent compared to the number of total observations. Therefore, the overall communication cost compared to centralized methods [Chatzigiannakis and Papavassiliou, 2007, an Huang et al., 2003], where all the observations are transmitted, is significantly lower. Each parent node then combines the received information with the results of its own local clustering in incrementally refining clusters. This helps in evaluating global correlations (across multiple nodes) as well as reducing any misclassification at the local level.

Therefore, each parent node calculates new membership scores and updates the clustering with respect to the the union set of outliers and centroids (local and recieved). Therefore, the outliers are evaluated again for membership scores in a refined set of clusters at each of the parent nodes via the evaluation of global correlations at that level. The parent level results of intermediatte outliers are identified by applying the statistical thresholds T_1, T_2 to current degrees of calculated membership scores. The computed centroids and anomalies are then sent to their parent nodes in the hierarchy. This process is recursively repeated until the root of the hiearchy (*g*ateway node) is reached, and the final results obtained with global anomalies identified.

2.3.2 Anomaly Identification

As stated earlier, anomalies can be defined as either local or global. However, they can also be either individual data points or a group of data points in a cluster, which can be reffered to as *i*ndividual anomalies and *a*nomalous clusters respectively. The identification of these anomalies is performed via two mechanisms within the distributed fuzzy clustering framework described above. The first mechanism involves the identification of specific outlying data points representing individual anomalies, while the second mechanism identifies outlying clusters where all member points of that cluster are considered to be anomalous.

2.3.2.1 Outliers as Individual Anomalies

Outlying data points are identified by applying the thresholds T_1, T_2 on the highest membership degrees $[max\, U(w_{ij})values]$ resulting from the soft partitioning of the data by *fcm*. This value, (one standard deviation away from the mean) is introduced adaptively at each hierarchical stage over the network. Therefore, the final remaining outliers at the gateway node provides the set of individual anomalies. These data points have now been evaluated for cluster membership at all the existing hierarchical levels of the sensor network, in evaluating more global correlations and showing no satisfactory degrees of cluster membership. These outliers corresponding to both local and global anomalies is therefore, declared as the set of *i*ndividual anomalies identified for the concerned observation domain.

2.3.2.2 Anomalous clusters

The second mechanism computes the intra-cluster distances between each of the representing clusters generated at each hierarchical level until the gateway node. The statistical thresholds (T_1, T_2) of one standard deviation away from the mean with respect to the distribution of observed intra-cluster distances will define the clusters being considered as normal or anomalous. Once this is performed accurately, the result is a set of normal and anomalous clusters with representative centroids at the gateway node. With the thresholds being applied adaptively over different topological levels for the network, both local and global anomalies are identified and isolated with respect to the formed clusters on observed data distributions at each level.

Furthermore, the accumulated results of globally evaluated anomalies at the gateway node is shared down the sensor network hierarchy and used for anomaly detection locally. This will provide the foundation for detecting global anomalies on a local basis in detecting distributed anomalous behaviour in the network.

2.3.3 The Distributed Anomaly Detection Algorithm

The individual steps of the framework (for anomaly detection) is summarised as follows.

1. Each sensor node S_j aggregates its local observations over a specific time period Δt.

2. Each sensor node S_j uses *fuzzy c-means* clustering on the locally observed data domain as below in minimizing the objective function (J) of weighted sum of squared errors with m determining weights influence.

$$J_m(U, C; X) = \sum_{j=1}^{k} \sum_{i=1}^{n} w_{ij}^m ||x_i - c_j||_A^2, 1 < m < \infty$$

initiate fuzzy pseudo partition with exponent $m > 1$

repeat

// compute centroids

$$c_j = \sum_{i=1}^{n} w_{ij}^m \frac{X_i}{\sum_{i=1}^{n} . w_{ij}^m}$$

// score fuzzy membership values

$$w_{ij} = \frac{(1/||x_i - c_j||^2)^{1/(m-1)}}{\sum_{q=1}^{n} (1/||x_i - c_q||^2)^{1/(m-1)}}$$

until c_j *doesn't change*

3. Each sensor node S_j then calculates the euclidean distance (i.e. $dist(c_i, c_j), i \neq j$) between the identified cluster prototypes (i.e. $\{c_1, c_2, c_3, ...c_k\} \in C$) in step 1 together with the *mean inter-cluster distance* for the set of inter cluster distances (i.e. $D_c =$

$dist(c_i, c_j) : j = 1...(|C| - 1))$ at each cluster prototype.

4. The cluster memberships are then evaluated and the local anomalies identified as both individual outlying data points and anomalous clusters through derivation of the statistical thresholds (T_1 and T_2) on the highest membership value in the fuzzy partition $[maxU(w_{ij})]$ as well as the inter-cluster distance between cluster centroids.

$$T(T_1, T_2) = 1/n \sum_{i=1}^{n} X_i \pm \sqrt{1/n \sum_{i=1}^{n} [X_i - (1/n \sum_{i=1}^{n} X_i)^2]}$$

for T_1, T_2 where $X = max[U_{ik}]$ *and* $X = mean[D_c]$

if $maxU_{ik} < T_1$ —— $maxU_{ik} > T_2$ **then**

X_i= local anomaly

end if

if mean$(distc_{ij}) > T_2$ **then**

$\forall X_i \in C, X_i$= local anomaly

end if

5. Each sensor node communicates the resulting anomalous data points (X_a) and the cluster prototypes(C) (both normal and anomalous) to its immediate parent node.

6. The parent node S_u collects the received information and combines them with its own local clustering results in merging and refining the clusters as well as the anomaly classification in secondary evaluation of fuzzy membership scores, as in Step 3 - 4.

7. Recursively follow Steps 1 - 6 up until gateway node S_g.

8. Results at the gateway node represent final anomalies comprising and representing both local and global anomalies on the considered WSN domain.

9. Gateway node S_g sends the global anomalies back to individual nodes through its children.

10. Children nodes compare the conceding local clusterings with the global anomaly values and detects them locally.

2.4 Results & Evaluation

The proposed anomaly detection algorithm is evaluated using both synthetic and real data sets in this section. This is performed in terms of classification accuracy for both normal and anomalous data points, as well as communication overheads compared to a centralized approach. A comparitive analysis is also performed between the proposed approach and the work in [Rajasegarar et al., 2006], which is the only distributed algorithm that propose an unsupervised approach to detect anomalies. Therefore, experimental evaluations on sensitivity and specificity analysis is performed between the two algorithms on different data distributions.

2.4.1 Data Sets

Two data sets representing different data distributions for normal and anomalous behaviour are used. These data sets consist of observations from real sensor motes as well as artificially created data points that are introduced to act as anomalies. The first data set is based on the publicly available Intel Lab Data consisting of real measurements collected from 54 sensors deployed at the Intel Berkeley Research Lab [Bodik et al., 2004]. The data collected includes timestamped topology information along with temperature, humidity, light and voltage values as the measured parameters from Mica2Dot sensors with weatherboards implemented. The format of the data is presented in the form shown in Table 2.1. As this data isn't annotated and doesn't contain any labelled information on observed anomalies, it is used in evaluations as follows. Four different data distributions are created by considering the available node data of the first four sensor motes (mote 01 - mote 04). The four attributes of temperature, humidity, light and voltage together are used as data vectors corresponding to individual observations with 4700 observations per single node.

First, the data is cleaned manually by identifying extreme values and removing them with the use of scatter plots. The rest of the data is labeled as normal for evaluations. Once this is

Table 2.1: Intel Data Set Data Format; Each observation consists of 4 mote attributes and 4 measured parameters

Attr. 1	Attr. 2	Attr. 3	Attr. 4	Para. 1	Para. 2	Para. 3	Para. 4
date:y-m-d	time:h:m:s	epoch	moteid	temperature	humidity	light	voltage

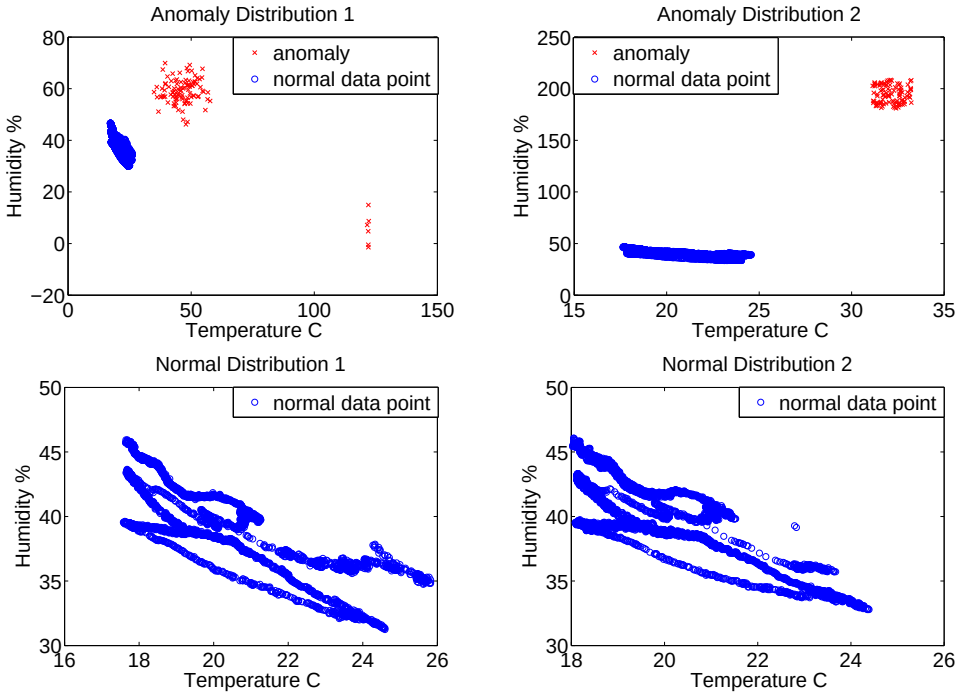

Figure 2.2: Data Distributions based on the Intel Sensor Data Set. Two data distributions representing a spatially dispersed (*top, left*) and a spatially concentrated (*top, right*) set of anomalous vectors are derived together with two complete distrbutions of tightly correlated and spatially focused observations representing normal behaviour (*bottom*).

complete two sets of randomly generated data are introduced in each for two derived normal data distributions to act as labelled anomalies. These data vectors are generated to represent two types of anomaly distributions that may occur in the form of a randomly dispersed set and a more focally concentrated set of anomalous observations. For this purpose, random sets of vectors are drawn from the standard *uniform* and standard *normal* distributions each over the normal measurements of individual attributes. Therefore, this data is simulated using matlab data generation algorithms for standard normal and uniform distributions. The resulting data sets are merged with the earlier derived normal data distributions of mote 01 and mote 03.

The data sets of mote 02 and mote 04 are kept as derived in representing only normal data distributions without any labelled anomalies. The resulting four data distributions are shown with respect to two attributes each in Figure 2.2. Then a three level hierarchical organization of wireless sensor nodes is simulated as depicted in Figure 2.1. The four data distributions created are taken as corresponding to the data distributions of leaf nodes S4 - S7 according to this hierarchical architecture.

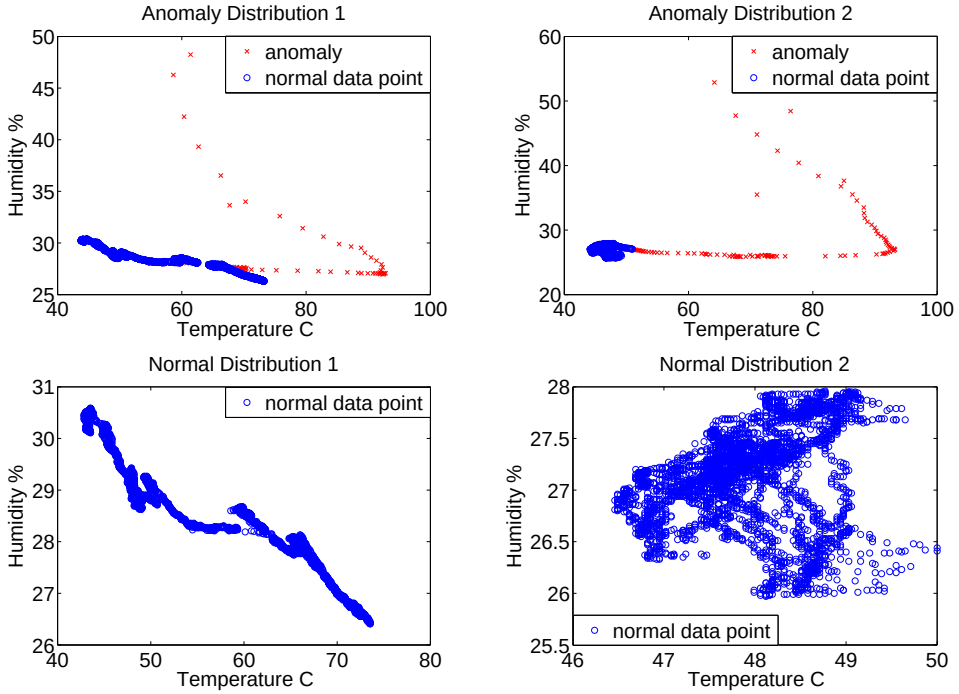

Figure 2.3: Data Distributions based on the ISSNIP Data Set. The two distributions with anomalous data represent a set of sequentially distributed anomalies with two tightly correlated normal data sets where one is is spatially concentrated (*top, right*) and the other more sequentially distributed (*top, left*). The (*bottom*) two distributions consist of similarly corresponding normal data without anomalies.

The second data set that is used is the annotated data set with labelled anomalies available from the Australian Research Council Research Network on Intelligent Sensors, Sensor Networks and Information Processing (*ISSNIP*) [Suthaharan et al., 2010]. This data corresponding to a multi-hop wireless sensor network deployment using TelosB sensor motes is used according to the same three level hierarchy as assumed in Figure 2.1. The data consists of temperature and humidity measurements collected over a period of 6 hours in 5 second in-

tervals with 4690 observations per each node. In this deployment anomalous observations are introduced to the data from node 01 and node 03 by artificially manipulating the measured temperature and humidity values. This is done by introducing a source of hot water and steam within the observed environment. The resulting data distributions available from mote 01-mote 04 is assumed to represent the leaf nodes S4 - S7 according to the simulated three level network hierarchy as stated before. Therefore, the data distributions of node S4 and node S6 will have the introduced labelled anomalies while nodes S5 and S7 has no anomalous data and represents normal observations. The data distributions for each of these sensor nodes are as depicted in Figure 2.3.

2.4.2 Evaluation

The proposed algorithm was implemented on the matlab environment for a simulated wireless sensor node hierarchy based on the two data sets created. Investigations on classification accuracy and communication efficiency were carried out by varying the number of locally expected clusters for each data distribution. Six experiment levels are chosen corresponding to the expected number of clusters from 7 to 12. This range was selected as any number below that will be too small to optimally represent a general sensor mote data distribution to score comparable fuzzy values, while a larger number would increase the computational complexity in making the process infeasible. Therefore, for each of the three hierarchical stages the *False Positive Rate* (FPR) and the *False Negative Rate* (FNR) were calculated based on the observed number of False Positives (FP) and False Negatives (FN) with the number of clusters ranging from 7 - 12. A fasle positive is identified in the instance where a normal measurement is detected as anomalous and a false negative identified in the instance an anomalous measurement is detected as normal. The false positive rate gives the ratio between the detected false positives and the actual normal measurements while the false negative rate gives the ratio between the detected false negatives and the actual number of anomalies respective to the underlying data distribution.

$$FPR = \frac{FP}{(FP + TN)} \qquad FNR = \frac{FN}{(FN + TP)}$$

The observed instances of True Poitives (TP) and True Negatives (TN) is also determined at each stage. Then the *Sensitivity* and *Specificity* values are calculated as follows to be the main evaluation metrics for data classification accuracy.

$$sensitivity = \frac{TP}{(TP + FN)} \qquad specificity = \frac{TN}{(TN + FP)}$$

Sensitivity measures the probability that a statistical test is positive for a given true positive statistic, while specificity measures the probability that a statistical test is negative for a given true negative statistic. These parameters are calculated in three stages as detailed next. The first set of results are computed based on the local *fcm* clustering for each of the four sensor nodes S4 - S7 in the assumed three level hierarchy depicted in Figure 2.1. Next, the results are calculated for the intermediate fuzzy scoring and cluster refining at nodes S2 and S3 and the final results are computed at the gateway node S1. The number of individual data points and the number of cluster centroids that are communicated on each wireless communication link are also recorded between the three topological levels for the node hierarchy. These are recorded in order to calculate the reduction in communication overhead for the algorithm compared to a centralized approach where all observed data points are communicated to the gateway node as is the case in most existing frameworks. The values of the objective function of the local *fcm* clustering is plotted against the number of iterations in investigating the complexity of arriving at a stable minimum with respect to the expected number of clusters. *fcm* works by minimizing the weighted sum of squared errors as the objective function, and therefore, the minimum value it reaches without any observed subsequent change produces the optimal clustering with the number of iterations representing time complexity. The scatterplots highlighting the change in the identified clusters and their distribution as well as isolated anomalies corresponding to the increase in the number of clusters in each data distribution are also presented.

2.4.2.1 Classification Accuracy - Intel Data Set

Performance analysis on data distributions based on the Intel Research Lab Data is evaluated first. The accuracy of data classification is investigated with respect to identifying anomalous and normal data points by calculating the values for sensitivity and specificity. Considering the

assumed three level hiararchical topology, results are given for each of the four local clustering phases at nodes S4 - S7, the two intermediate clustering phases at nodes S2 - S3 and the final analysis at the gateway node S1 in the sequence of tables from Table 2.2 to Table 2.5.

2.4.2.1.1 Local Clustering Results

The first data distribution consists of a spatially dispersed set of random vectors as the introduced anomalies at node S4 corresponding to the scatterplot in Figure 2.2(*top,left*). Results in Table 2.2 show the calculated average sensitivity as 0.9050 with an average specificity of 0.9983 over the expected number of clusters from 7-12. Therefore, the proposed approach achieves extremely high accuracy in classifying normal data in detected *True Negatives* and very high accuracy in classifying anomalies in detected *True Positives*. The corresponding objective function plots show the required number of iterations in the *fcm* algorithm slightly increasing with the number of expected clusters in the range of 15 - 20 in Figures 2.4. There's no major impact to accuracy in the choice of the expected number of clusters in the investigated range from 7-12 other than a slight decrease at the lower and higher ends of cluster numbers. This can be explained in two stages corresponding to the increase in the *False Negative* rate regarding a spatially dispersed set of random anomalies. For a higher number of clusters, the calculated fuzzy membership values will be more compact and closer together with every data point getting a score that is closer to the mean in the set of fuzzy values. This is due to the fact that a higher number of clusters for a randomly spread out data distribution provides the opportunity for most data points to belong to a particlular cluster prototype with a high membership score depending on the distance between them. Therefore, the calculated number of outliers will be low from the defined threshold of one standard deviation away from the mean on the membership value, as some of the anomalies also get high membership scores that are closer to the mean with the increase in clusters numbers leading to an increase in the *False Negative* rate.

On the other hand, for a smaller number of clusters the calculated fuzzy membership scores indicate most of the data points belonging satisfactoraly to the identified clusters without clear outliers. Therefore, the anomalies are being identified through classifying anomalous clusters

Table 2.2: Classification Accuracy (Node S4) Anomalous Data : Distribution 1 INTEL

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	15	0.1402	2	4.3545e-04	0.8598	0.9996	4591	92
8	9	0.0841	11	0.0024	0.9159	0.9976	4582	98
9	9	0.0841	25	0.0054	0.9159	0.9946	4568	98
10	8	0.0748	2	4.3545e-04	0.9252	0.9996	4591	99
11	8	0.0748	2	4.3545e-04	0.9252	0.9996	4591	99
12	12	0.1121	5	0.0011	0.8879	0.9989	4588	95

rather than individual outliers. However, as the number of clusters are small and the observed data distribution is spread out randomly over a wide range the inter-clustster distances between them are also large with little variation. Therefore, introducing a threshold of one standard deviation away from the mean on inter-cluster distance for anomalous clusters provide a higher probability for them to be classified as normal. This results in a higher *False Negative* rate. The scatterplots for the clustering and the isolated anomalies are presented in Figures 2.4.

Table 2.3: Classification Accuracy (Node S6) Anomalous Data : Distribution 2 INTEL

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	0	0	196	0.0426	1	0.9574	4404	100
8	0	0	1	2.1739e-04	1	0.9998	4599	100
9	0	0	1	2.1739e-04	1	0.9998	4599	100
10	0	0	32	0.0070	1	0.9930	4568	100
11	0	0	32	0.0070	1	0.9930	4568	100
12	0	0	1	2.1739e-04	1	0.9998	4599	100

For the second set of results presented in Table 2.3, the sensitivity and specificty of detection is analysed corresponding to the data distribution in mote S6 with a clearly defined set of spatially focused values from a randomly defined *uniform* distribution forming the anomalies. Here, 100% sensitivity (1.0000) is achieved for all stages with an average specificity of 0.9905. Therefore, for a set of tightly correlated and spatially concentrated anomaly range (Figure 2.2(*top,right*)), the proposed approach ideally identifies all of the observed anomalies correctly together with a very high accuracy of 99% in classifying normal data. In such a situation, the woking hypothesis of data clustering that all normal data will be closer together in a tightly correlated and large set of vectors, while anomalous data will be a small set of

vectors significantly away from the normal data and with loose correlation, produces very high accuracy through the classification of anomalies in outlying clusters that significantly standout as deviations from the observed inter-cluster distances corresponding to normal data. The scatterplots for the clustering and the isolated anomalies are presented in Figure 2.6 with the objective function variation showing the required number of iterations below 20.

The data distributions of mote S5 and mote S7 provide two complete distributions of all normal data without any anomalies corresponding to the most probable scenario in a real sensor network deployment (Figure 2.2(*bottom*)). As there are no anomalies in this situation the major influencing metric will be the false positive rate. The results for these two data distributions are identical and as given in Table 2.4 show ideal detection accuracy through specificity values of 100% in all stages with 0 false positive rate. Therefore, the probability of a detected normal data point being actually normal through the proposed detection approach leads to ideal classification accuracy in the normal run of the sensor network deployment. The objective function reaches stability below 15 iterations with the cluster distributions as shown in Figures 2.5 and 2.7.

Table 2.4: Classification Accuracy (Nodes S5 and S7) Normal Data : Distribution 3 and Distribution 4 INTEL

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	0	NAN	0	0	NAN	1	4700	0
8	0	NAN	0	0	NAN	1	4700	0
9	0	NAN	0	0	NAN	1	4700	0
10	0	NAN	0	0	NAN	1	4700	0
11	0	NAN	0	0	NAN	1	4700	0
12	0	NAN	0	0	NAN	1	4700	0

2.4.2.1.2 Distributed Clustering Results

The intermediate step of secondary fuzzy scoring and cluster refining on the hierarchical framework is done at the nodes S2 and S3 via second level evaluation of membership values and inter-cluster distances. Here, the detection accuracy for the combined union of data sets were found to be 0.9050 and 1.0000 in average sensitivity with an average specificity of 0.9992 and

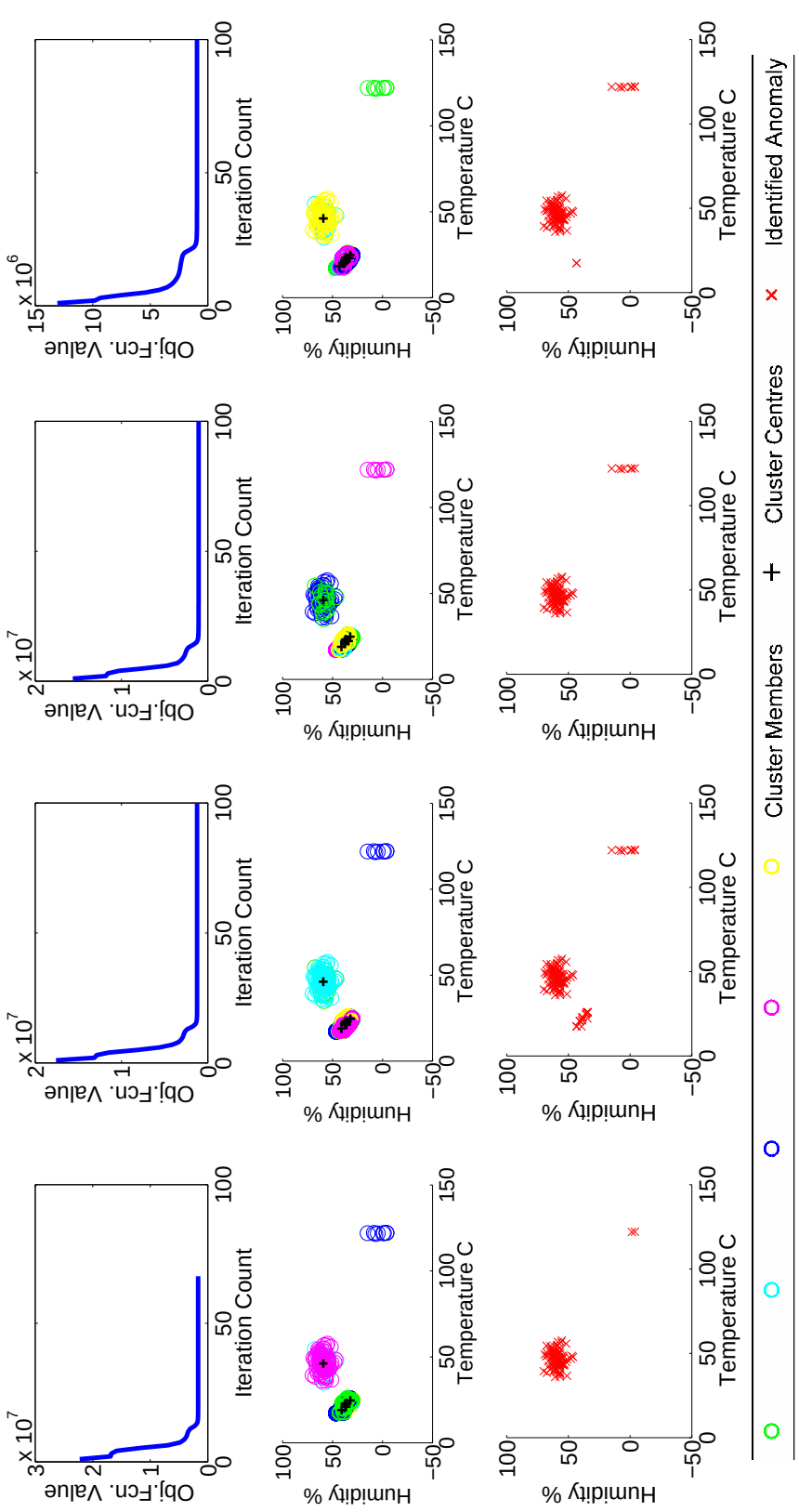

Figure 2.4: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and 12 expected clusters for Intel Anomaly Distribution 1 (*rows 1-3*). For Intel Anomaly Distribution 1, the Obj. Function stabilises with a slightly increased no. of iterations (below 20). Clusters are concentrated on the normal data with 1 cluster in the anomaly range and no discernable difference in the cluster distributions resulting with low impact on isolated anomalies as the no. of clusters increase (from *left* to *right*).

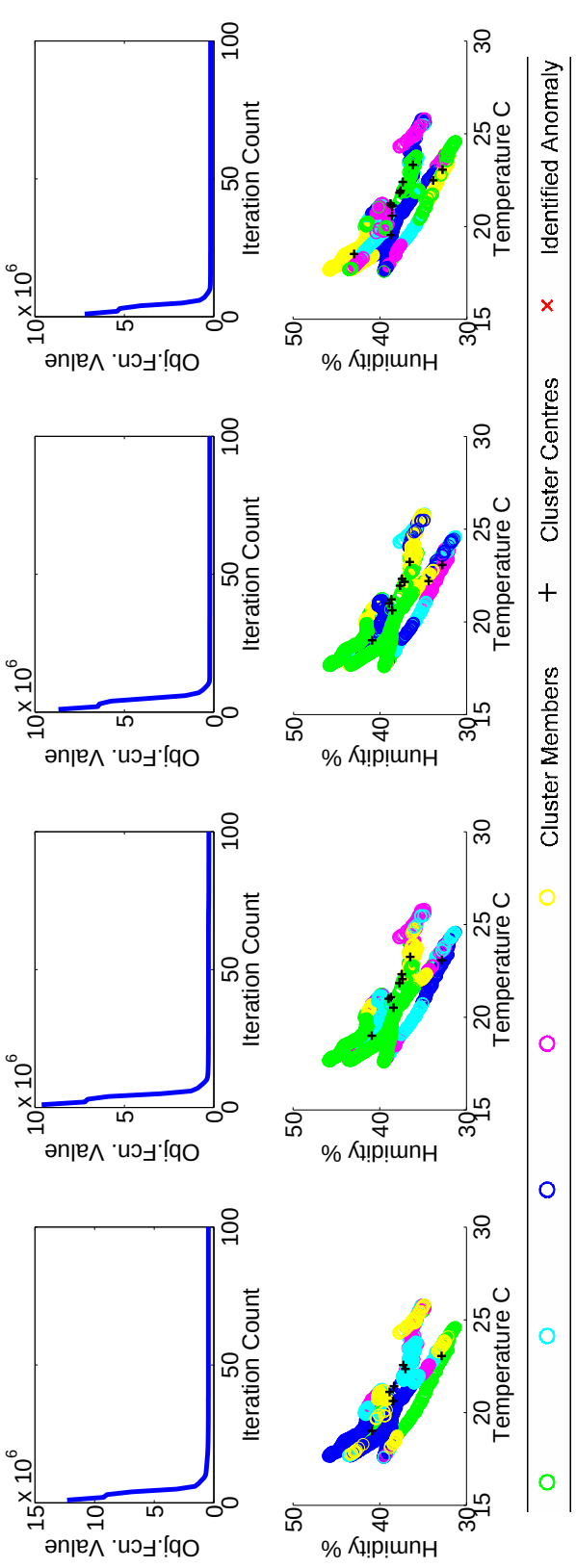

Figure 2.5: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and 12 expected clusters for Intel Normal Data Distribution Distribution 3 (*rows 1-2*). For Intel Normal Data Distribution 3, the Obj. Function stabilises with a slightly increased no. of iterations (below 15). The cluster centroids spread out gradually with the increase in the no. of expected clusters (from *left* to *right*) resulting in more tight cluster memberships for all data points.

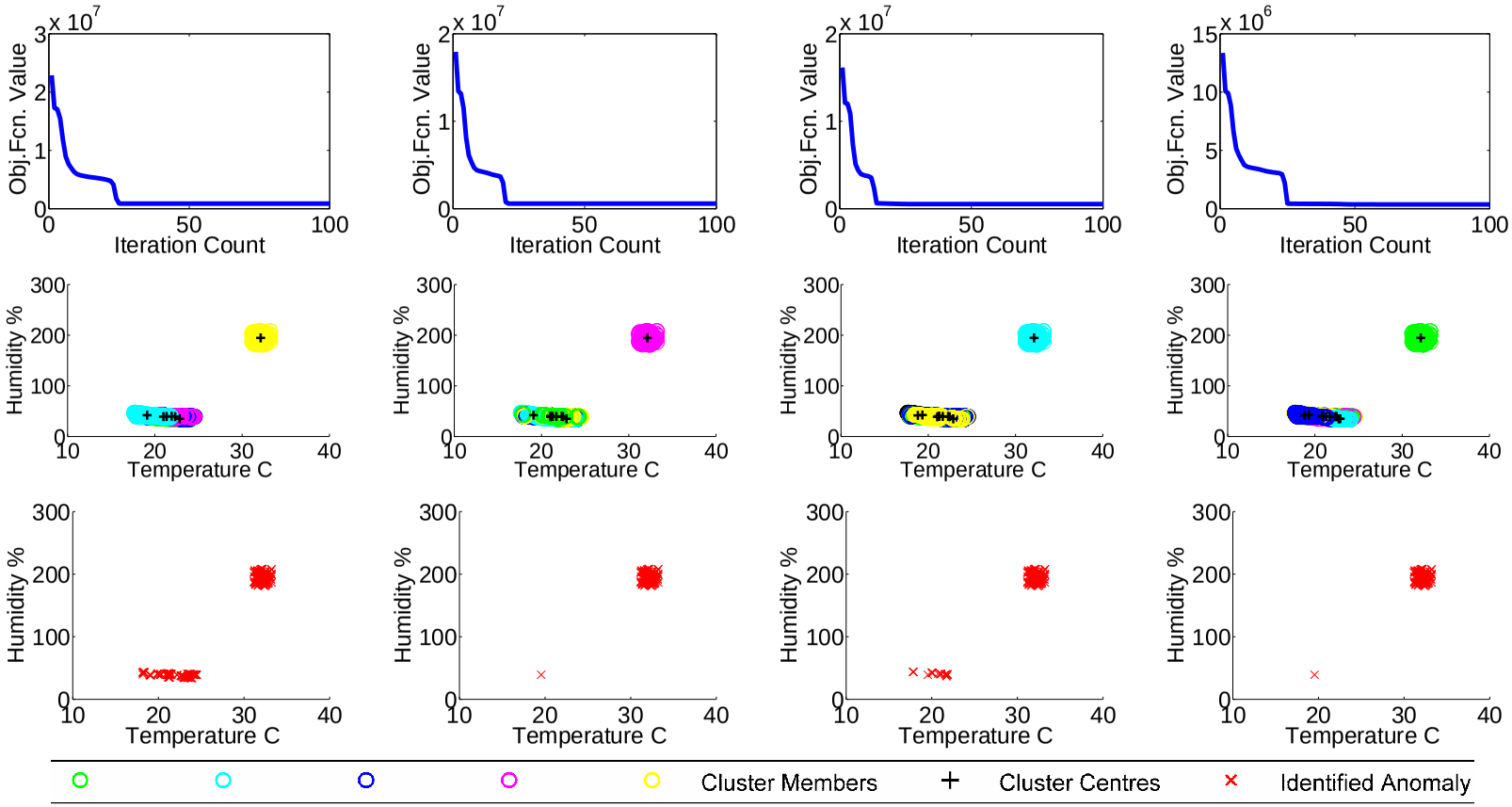

Figure 2.6: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and 12 expected clusters for Intel Anomaly Distribution 2 (*rows 1-3*). Similar to Figure 2.4, for Intel Anomaly Distribution 2, the Obj. Function stabilises with a slightly increased no. of iterations (below 20) with no discernable difference in the cluster distributions as both normal and anomalous data is tightly correlated and packed with 1 cluster representing the anomaly range. There's low impact on isolated anomalies as the no. of clusters increase (from *left* to *right*).

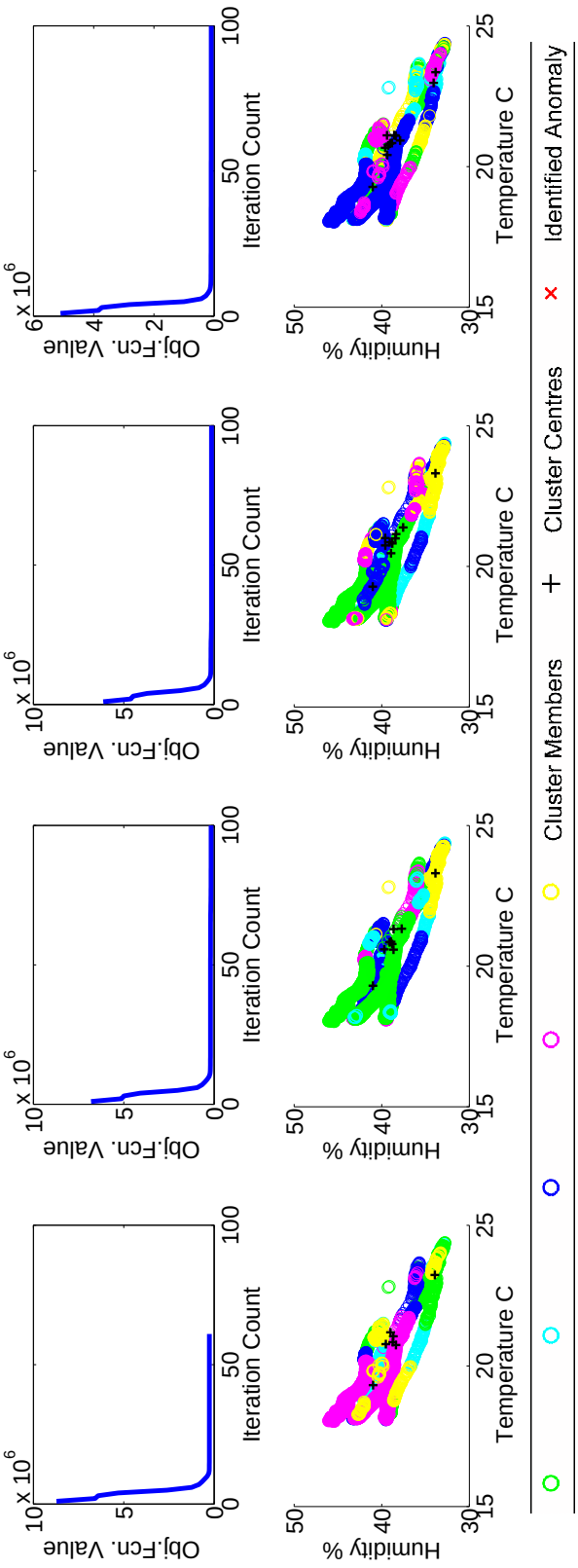

Figure 2.7: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and 12 expected clusters for Intel Normal Data Distribution 4 (*rows 1-2*). For Intel Normal Data Distribution 4, the Obj. Function stabilises with a slightly increased no. of iterations (below 15). The cluster centroids remain tightly together as the data is more focally concentrated (compared to Figure 2.5) with only a slight spread out with the increase in the no. of expected clusters (from *left* to *right*) resulting in more similar cluster memberships for all data points.

0.9953 respectively at nodes S2 and S3.The final results for the Intel data set are calculated at the gateway node S1 based on all the local and intermediate clustering results recieved. These results as shown in Table 2.5, clearly demonstrates that the proposed algorithm has comparable or very high acccuracy in terms of classification accuracy through the sensitivity and specificity metrics. The final average sensitivity and average specificity values come to 0.9509 (95%) and 0.9972 (99%). Therefore, considering the detection of anomalies within a given temporal window of Δt on a hierarchical network topology the proposed framework provides highly accurate classification in clearly identifying both normal data and anomalies.

Table 2.5: Final Classification Accuracy (Node S1): Distributions 1,3 and 2,4 INTEL

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	15	0.0725	198	0.0106	0.9275	0.9894	18395	192
8	9	0.0435	12	6.4540e -004	0.9565	0.9994	18581	198
9	9	0.0435	26	0.0014	0.9565	0.9986	18567	198
10	8	0.0386	34	0.0018	0.9614	0.9982	18559	199
11	8	0.0386	34	0.0018	0.9614	0.9982	18559	199
12	12	0.0580	6	3.2770e -004	0.9420	0.9997	18587	195

2.4.2.2 Classification Accuracy - ISSNIP Data Set

Next, the proposed algorithm is evaluated on the data sets based on the annotated data from ISSNIP. Considering the same assumed hierarchical topology for the sensor network as discussed above the results are given for each of the local, intermediate and final data clustering and analysis stages through Table 2.6 -2.9.

2.4.2.2.1 Local Clustering Results

The results on Table 2.6 for the data distribution with introduced anomalies depicted in Figure 2.3(*top,left*) gives an average sensitivity rate of 0.6035 and an average specificity rate of 0.9998. Therefore, very high accuracy in classifying normal data is achieved with a corresponding very low *false positive* rate. However, the classification of actual anomalies show a marked reduction with a corresponding high *false negative* rate. This can be mainly attributed to the nature of the data distribution itself. As the main hypothesis for a data clustering approach is that the

normal data for a typical spatio-temporal observation domain will be tightly correlated and spatially closer together while the anomalies will be small in number with loose correlation and spatially distanced from the normal data, it drives the proposed framework to identify anomalies based on a threshold of one standard deviation away from the mean on the calculated fuzzy membership scores and inter-cluster distances. However, as it is clear from the data distribution in Figure 2.3 the concerned data has introduced anomalies that gradually increase from the observed normal data sequentially. This is due to the fact that the ISSNIP researchers have artificially manipulated the observed temperature and humidity values through the introduction of a temporally increasing heat source for anomaly generation. Hence, a portion of the annotated anomalous data are spatially very close to the majority of the normal data. This drives the framework to classify them as normal in increasing the *false negative* rate and impacting on the overall sensitivity measured. The *false negatives* ranging from 22-25, directly correspond to these data that are very similar to the majority of normal data although they are annotated as anomalies. Meanwhile the anomalies that are more spatially different in the range of sequentially increasing observations are correctly classified.

Table 2.6: Classification Accuracy (Node S4) Anomalous Data : Distribution 1 ISSNIP

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	22	0.3793	1	2.1589e -04	0.6207	0.9998	4631	36
8	25	0.4310	1	2.1589e -04	0.5690	0.9998	4631	33
9	25	0.4310	1	2.1589e -04	0.5690	0.9998	4631	33
10	22	0.3793	1	2.1589e -04	0.6207	0.9998	4631	36
11	22	0.3793	1	2.1589e -04	0.6207	0.9998	4631	36
12	22	0.3793	1	2.1589e -04	0.6207	0.9998	4631	36

In an industrial sensor network deployment setting, an anomaly feature that sequentially increases as here will be detected by the proposed approach when it passes the derived threshold. As this threshold is set to be only one standard deviation away from the mean for the concerned parameter (i.e membership value or inter-cluster distance) the anomaly will be detected before it can increase up to a level that can be harmful to the system. On the other hand, for this data distribution there is no major imapct to the observed accuracy in the choice of the number of expected clusters from 7-12 with the objective function reaching a stable min-

Table 2.7: Classification Accuracy (Node S6) Anomalous Data : Distribution 2 ISSNIP

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	0	0	5	0.0011	1	0.9989	4585	100
8	3	0.0300	1	2.1786e -04	0.9700	0.9998	4589	97
9	2	0.0200	1	2.1786e -04	0.9800	0.9998	4589	98
10	0	0	3	6.5359e -04	1	0.9993	4587	100
11	11	0.1100	4	8.7146e -04	0.8900	0.9991	4586	89
12	3	0.0300	1	2.1786e -04	0.9700	0.9998	4589	97

imum between 10-15 iterations as shown in Figure 2.8. The local clustering results and the identified anomalies for each number of expected clusters are also given in Figure 2.8.

The results for the second ISSNIP data distribution with anomalies are given in Table 2.7. Here we have very high classification accuracy in both the average sensitivity and average specificity rates of 0.9683 and 0.9995 over the range of expected number of clusters from 7-12. All the objective function values reach a stable minimum between 10-15 as presented in Figure 2.10. The corresponding scatterplots featuring the identified clusters and the anomalies are also shown. This data distribution also has introduced anomalies that sequentially increase from near the spatial domain of the majority normal data (See Figure 2.3(*top,right*)). However, in this case the normal data is more tightly pact through spatial correlations in contrast to the loosely spread anomalies resulting in increased detection accuracy compared to the previous data distribution.

Table 2.8: Classification Accuracy (Nodes S5 and S7) Normal Data : Distribution 3 and Distribution 4 ISSNIP

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	0	NAN	0	0	NAN	1	4690	0
8	0	NAN	0	0	NAN	1	4690	0
9	0	NAN	0	0	NAN	1	4690	0
10	0	NAN	0	0	NAN	1	4690	0
11	0	NAN	0	0	NAN	1	4690	0
12	0	NAN	0	0	NAN	1	4690	0

The results for the two data distributions with no anomalies and representing all normal measurements are as presented in Table 2.8. As both of these data distributions consist only of

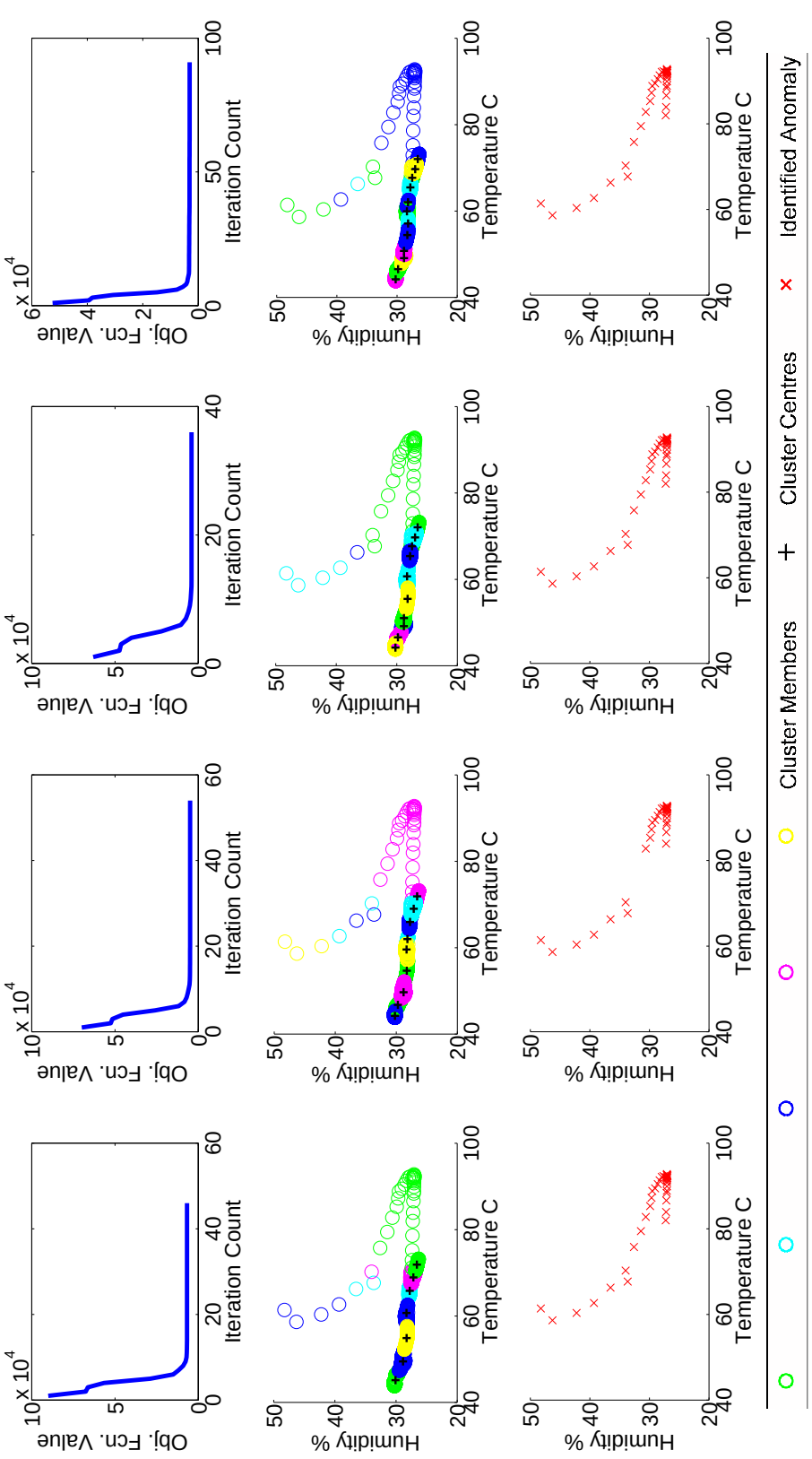

Figure 2.8: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and12 expected clusters for ISSNIP Anomaly Distribution 1 (*rows 1-3*). For ISSNIP Anomaly Distribution 1, the Obj. Function stabilises with a slightly increased no. of iterations (below 15). The cluster distribution spreads sequentially corresponding to the normal data with no clusters in the anomaly range resulting in all anomalies to be identified as individual outliers with high accuracy as the no. of clusters increase (from *left* to *right*).

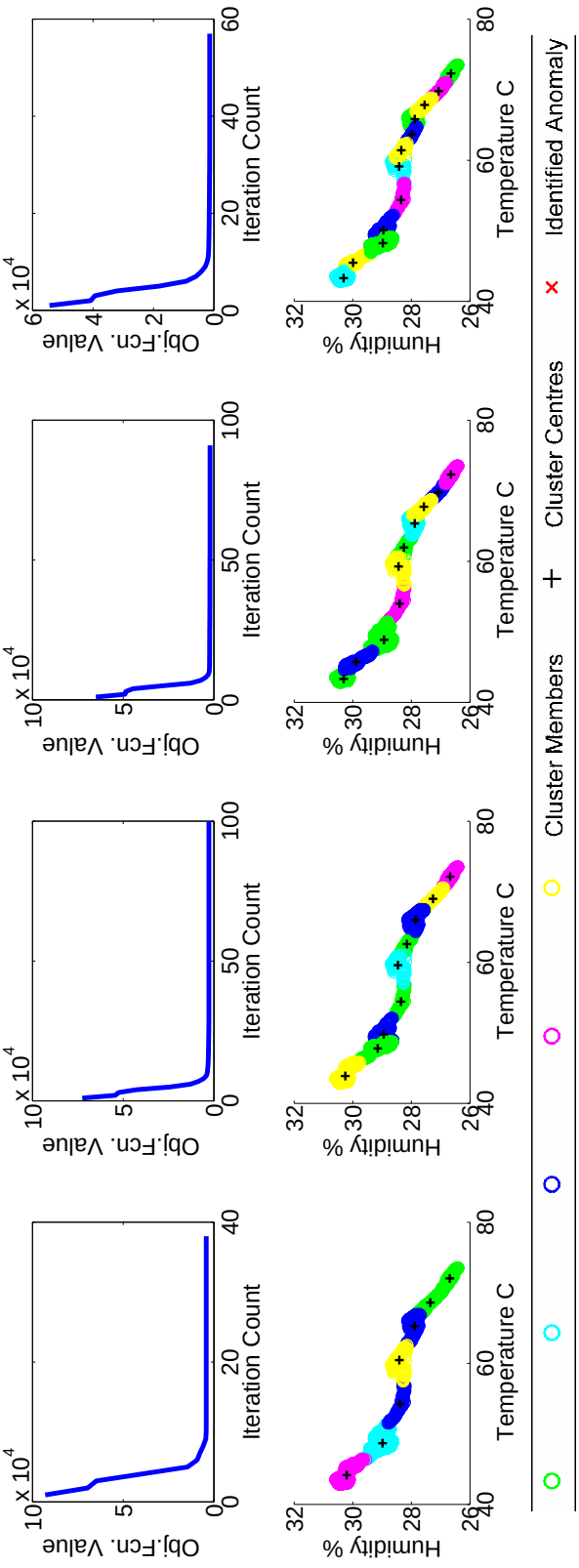

Figure 2.9: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and12 expected clusters for ISSNIP Normal Data Distribution 3 (*rows 1-2*). For ISSNIP Normal Data Distribution 3, The Obj. Function stabilises with a slightly increased no. of iterations (below 15). The cluster centroids spread out sequentially highlighting the sequential nature of the distribution resuting in similar but more tight cluster memberships to each data point as the no. of clusters increase (from *left* to *right*).

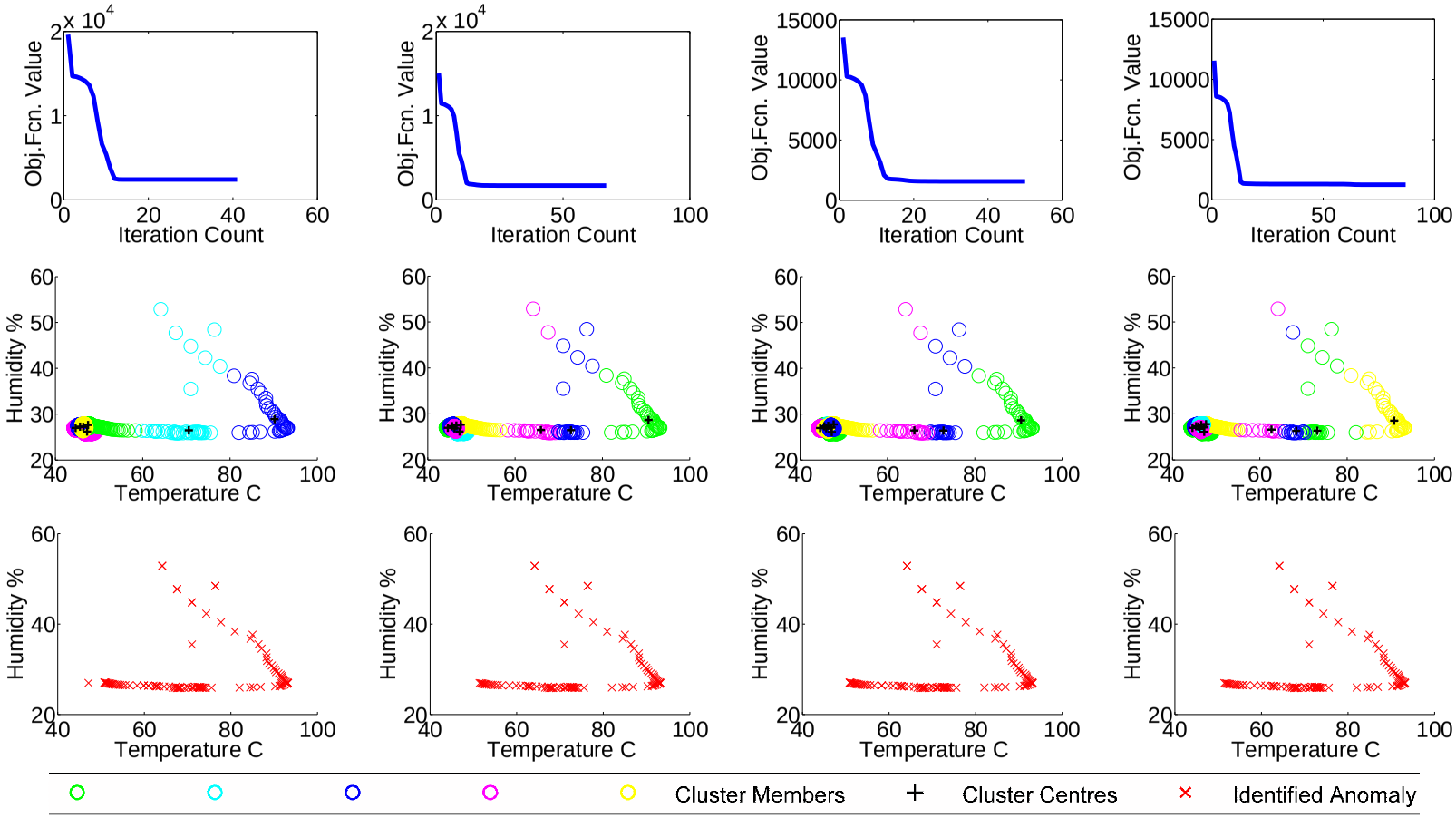

Figure 2.10: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and12 expected clusters for ISSNIP Anomaly Distribution 2 (*rows 1-3*). For ISSNIP Anomaly Distribution 1, the Obj. Function stabilises with a slightly increased no. of iterations (below 15). The cluster distribution is concentrated on the compact normal data and spreads sequentially corresponding to the anomalies (unlike in Figure 2.8) resulting in anomalies to be identified as anomalous clusters as well as individual outliers with high accuracy as the no. of clusters increase (from *left* to *right*).

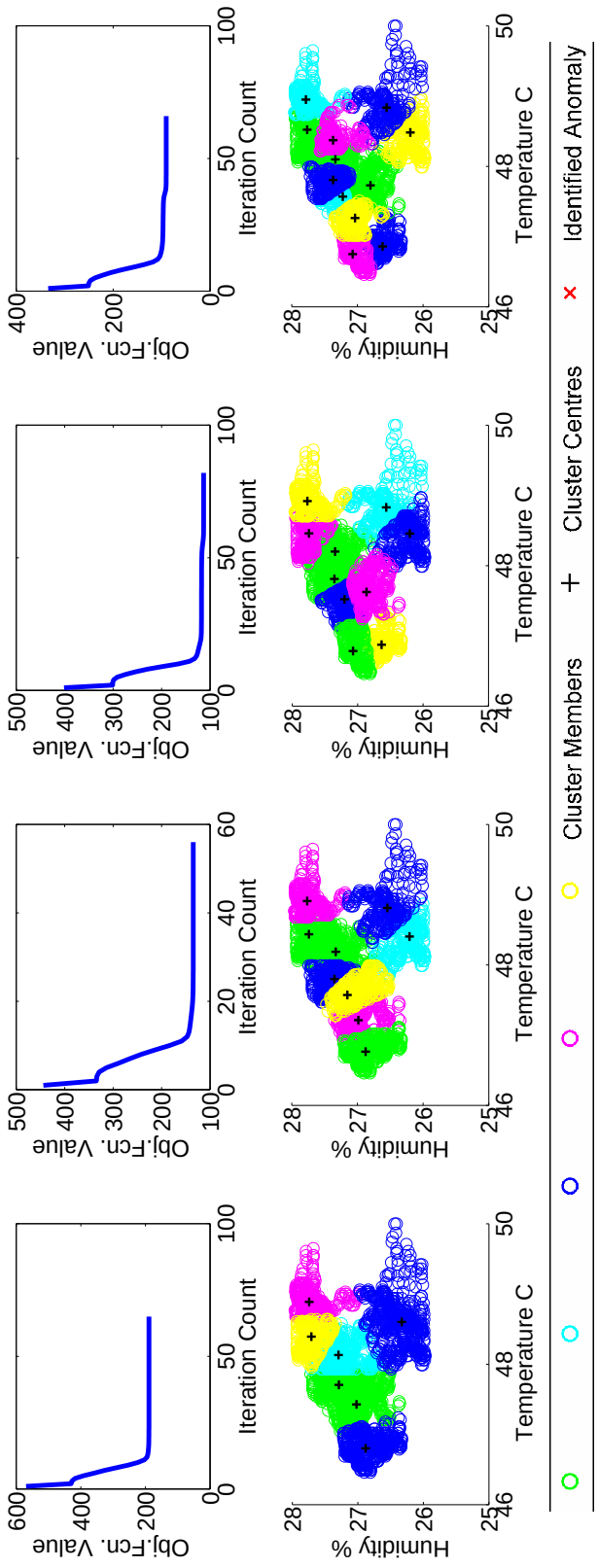

Figure 2.11: Objective Function (sum of squared errors) Variation, Detected Clusters and Isolated Anomalies (from *top* to *bottom*) for 7,9,10 and12 expected clusters for ISSNIP Normal Data Distribution 4 (*rows 1-2*). For ISSNIP Normal Data Distribution 4, the Obj. Function stabilises with a slightly increased no. of iterations (below 15). The cluster centroids spread out uniformly highlighting the uniformly compact nature of the distribution resuting in similar but more tight cluster memberships to each data point as the no. of clusters increase (from *left* to *right*).

tightly correlated spatially compact observations our approach provides an ideal classification accuracy through a constant value of specificity 1 for the range of cluster numbers from 7-12. There are no false positives observed at each phase for both distributions with the maximum number of iterations in reaching a stable minimum in objective function value below15 as shown in Figures 2.9 and 2.11, with the corresponding cluster distributions as depicted.

2.4.2.2.2 Distributed Clustering Results

Considering the secondary results from the two intermediate stages of clustering at nodes S2 and S3 in the hierarchical topology, the average sensitivity achieved was 0.6035 and 0.9683 with an average specificity of 0.9999 and 0.9998 over the range of clusters from 7-12. Results for the final clustering and global anomaly detection at the gateway node for all data distributions are given in Table 2.9. In the context of the data distributions based on the ISSNIP data set the proposed approach provides very high classification accuracy in terms of final average sensitivity and specificity rates of 0.8344 and 0.9998. Here the reduction of sensitivity to 83% corresponds to the sequential nature of the anomaly distribution at node S01 as explained previously. An ideal specificity of 99% is achieved in classifying normal data for all distributions.

Table 2.9: Final Classification Accuracy (Node S1): ISSNIP 1,3 and 2,4

No. Clusters	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
7	22	0.1392	6	3.2255e -004	0.8608	0.9997	18596	136
8	28	0.1772	2	1.0752e -004	0.8228	0.9999	18600	130
9	27	0.1709	2	1.0752e -004	0.8291	0.9999	18600	131
10	22	0.1392	4	2.1503e -004	0.8608	0.9998	18598	136
11	33	0.2089	5	2.6879e -004	0.7911	0.9997	18597	125
12	25	0.1582	2	1.0752e -004	0.8418	0.9999	18600	133

2.4.2.3 Communication Efficiency & Complexity

Considering WSNs deployed in critical infrastructure monitoring it is vital that the network lifetime is sustained and extended in reducing blackouts and process monitoring gaps. Therefore, a distributed approach for data analysis is vital in reducing communication overheads. This is due to the fact that in sensor networks the majority of the consumed energy is spent on

data communication rather than on computation [Pottie and Kaiser, b, Raghunathan et al., 2002]. Here, we investigate the communication overheads incurred at each level of the proposed approach concerning the assumed network model compared to a centralized approach in Tables 2.10 and 2.11. The number of individual data vectors communicated on each wireless link at first and second hierarchical levels on the topology are calculated. Then the overhead reduction percentage is calculated based on the ratio of the communicated data points with respect to a centralized approach where all observations of each node are sent. This is investigated for 6 stages with the expected number of clusters ranging from 7-12.

Table 2.10 provides the analysis for the Intel data distribution. The computed ratios show a least reduction ratio of 0.0322 (96.77%) to a highest ratio of 0.0107 (98.92%) for wireless links on level 1, accumulating up to ratios of 0.0211 (97.90%) to 0.0113 (98.87%) in level 2. Similarly the results in Table 2.11 for the ISSNIP data set gives a least reduction ratio of 0.0120 (98.80%) and a highest ratio 0.0046 (99.54%) in level1, with the final reductions accumulating upto 0.0080 (99.20%) least and 0.0075 (99.25%) highest over the range of cluster numbers from 7-12. Thus, the average communication reduction percentage for the Intel data set is 98.58% on the first level with 98.63% on the second level. The corresponding values for the ISSNIP data set stands at 99.18% and 99.23%. Therefore, this level of communication reduction of more than 98% average at all instances will greatly benefit the concerned sensor network in saving energy and maximizing active life-time.

In the proposed algorithm the time complexity to calculate the membership matrix and achieve a soft partitioning of the data is $O(ndc^2i)$ where n is the number of data points, d the number of dimensions, c the number of clusters and i is the number of iterations. Here, n is the major determinative factor with the highest effecting value for complexity compared to others. However, n is effectively curtailed after one algorithmic step, as from the the second stage onwards only the local outliers are processed with the identified cluster centroids that are communicated (as above results clearly show). Considering the first step, the number of observations can be determined by the operator in limiting n through a specifically chosen time window Δt within the application context.

The message complexity can be analyzed as follows. On each link the basic message com-

plexity will be $O(N)$ where $N = [nC + A_i]$, with C and A being the number of clusters and local anomalies at each sensor node (i). While C varies consistently from 6-12 in this implementation, n represents the link level in the considered hierarchical topology. Therefore, message complexity is concluded to be far lower than a centralized implementation where N consists of all observation vectors at each node accumulating over different network levels. The general *transport capacity* of a WSN consisting of M nodes in a unit area is given by $O(\sqrt{M})$ bit-m/s [Gupta and Kumar, 2000]. With a bit meter defined as the transport of 1 bit over 1 meter, the throughput per node is then given by $O(1/\sqrt{M})$ bit-m/s. Therefore, this becomes the fundamental boundary of communication for a generalized WSN considering node distance and connectivity aspects. Accordingly, the algorithm implementation in the proposed approach is quite feasible for any such condition in a given WSN. At any given time in the presented iterative approach, only the nodes in one level are communicating with their immediate parent nodes which are defined to be only one hop away. No simultaneous transmissions occur between different levels as such other than on that singular level across the network confirming to that of the acceptable throughput limits. Furthermore, for any reasonable scenario of WSNs in an industrial monitoring setting it can be assumed that the average distance between nodes is of $O(1/\sqrt{M})$ offering further sufficient monitoring density. Therefore, it can be clearly stated that the communication requirements are satisfied within the physical power and bandwidth considerations of the network according to transport capacity as derived through distance and node density criteria.

2.4.2.4 Comparative Analysis

Here, the key limitations in other distributed clustering algorithms [Kargupta et al., 2001, Tasoulis and Vrahatis, 2004, Nowak, 2003, Forero et al., 2011] are highlighted against the proposed approach. In [Kargupta et al., 2001] and [Tasoulis and Vrahatis, 2004] two methods are proposed based on *collective principal component analysis* and *k-windows*. [Kargupta et al., 2001] achieves a global clustering of a single centrally located data set distributively at different nodes through local data projection along global Eigen vectors derived using local counterparts and samples of locally projected data. In [Tasoulis and Vrahatis, 2004], this is done based on the

Table 2.10: Communication Overhead Analysis: INTEL Data Set; The number of data points communicated on each wireless link for 2 levels on a hierarchical topology are given with the ratio and reduction percentage corresponding to the variation in the expected number of clusters from 7-12

No. Clusters	Centralized Approach (Level1)	Distributed Approach (Level1)	Reduction Ratio-Percentage (Level1)	Centralized Approach (Level2)	Distributed Approach (Level2)	Reduction Ratio-Percentage (Level2)
7	9400	101	0.0107 - 98.92%	18800	397	0.0211 - 97.90%
7	9400	303	0.0322 - 96.77%	18800	-	-
8	9400	117	0.0124 - 98.75%	18800	218	0.0116 - 98.84%
8	9400	108	0.0115 - 98.85%	18800	-	-
9	9400	132	0.0140 - 98.60%	18800	233	0.0124 - 98.76%
9	9400	110	0.0117 - 98.83%	18800	-	-
10	9400	111	0.0118 - 98.82%	18800	243	0.0129 - 98.71%
10	9400	142	0.0151 - 98.50%	18800	-	-
11	9400	112	0.0119 - 98.80%	18800	244	0.0130 - 98.70%
11	9400	143	0.0152 - 98.50%	18800	-	-
12	9400	112	0.0119 - 98.81%	18800	213	0.0113 - 98.87%
12	9400	113	0.0120 - 98.80%	18800	-	-

positioning, moving and enlargement of windows to capture the most number of data points at different locations and then merging intersecting ones centrally. These are severely hampered by the associated communication costs at each step and being non adaptable for dynamic data that is itself generated distributively. Also a global clustering does not reflect the intermediate corelations that are critical to identify anomalies at different granularity levels in a WSN while a large number of user defined parameters makes the process practically unfeasible.

Table 2.11: Communication Overhead Analysis: ISSNIP Data Set; The number of data points communicated on each wireless link for 2 levels on a hierarchical topology are given with the ratio and reduction percentage corresponding to the variation in the expected number of clusters from 7-12

No. Clusters	Centralized Approach (Level1)	Distributed Approach (Level1)	Reduction Ratio-Percentage (Level1)	Centralized Approach (Level2)	Distributed Approach (Level2)	Reduction Ratio-Percentage (Level2)
7	9380	44	0.0047 - 99.53%	18760	149	0.0079 - 99.21%
7	9380	112	0.0119 - 98.81%	18760	-	-
8	9380	42	0.0045 - 99.55%	18760	140	0.0075 - 99.25%
8	9380	106	0.0113 - 98.87%	18760	-	-
9	9380	43	0.0046 - 99.54%	18760	142	0.0076 - 99.24%
9	9380	108	0.0115 - 98.85%	18760	-	-
10	9380	47	0.0050 - 99.50%	18760	150	0.0080 - 99.20%
10	9380	113	0.0120 - 98.80%	18760	-	-
11	9380	48	0.0051 - 99.50%	18760	141	0.0075 - 99.25%
11	9380	104	0.0111 - 98.90%	18760	-	-
12	9380	49	0.0052 - 99.48%	18760	147	0.0078 - 99.22%
12	9380	110	0.0117 - 98.83%	18760	-	-

In [Nowak, 2003], an incremental EM algorithm is adapted for distributed clustering in WSNs.The data is modelled as a mixture of Gaussians with each component assumed to represent one elementary condition. However, the final outcome treats the whole composite of data sets at individual nodes as a single entity requiring close form expressions for local estimators using sufficient statistics. Furthermore, the mixing probabilities can be unreliable with the contribution of mixture components requiring prior knowledge on distributions and the ideal of

a static homogeneous environment without any dynamicity. A distributed k-means algorithm and finite mixture model are presented for both deterministic and probabilistic partitional clustering in [Forero et al., 2011]. Both methods only achieve a central global clustering of the data domain that is considered distributed to individual nodes. Two major limiting factors are assumptions that the data is independent and identically distributed (i.i.d condition) as well as cluster numbers to be similar at each node. In the distributed k-means version a broadcast step is also required communicating the updated local cluster memberships at each iteration. This acts as a major burden and also allows inappropriate data sharing over a broadcast domain as opposed to a single hop criterion limiting unwarranted exposure. The probabilistic partitional method is presented as a follow-up of estimating mixture density formed by class conditional probability density functions (pdfs). Here, the representativeness of the used tuning scalars as well as the no-consensus constraints on langrangian multipliers remains key obstacles for improved accuracy and robust adaptation. Finding the minimum for the parameter that guarantees stability is also challenging with the presence of different scalar, vector and matrix updates required at different iterative steps. Thus, both remain untenable for a distributed anomaly detection approach in a WSN context.

2.4.2.5 Performance Comparison

Here, the performance of the proposed algorithm is compared with that of the approach in [Rajasegarar et al., 2006]. The fixed-width clustering algorithm is implemented on a hierarchical framework for a three level sensor node topology and tested on the same two data sets from Intel and ISSNIP described in Section 2.4.1. *Sensitivity* and *Specificity* values are calculated for evaluating detection accuracy by varying the specified value for the width parameter w. The range between 0.001 and 0.0125 is chosen in 0.0025 intervals, as values below that will contribute to a large number of clusters with some even being singletons and contributing to higher computational and communicational costs. The parameter K in determining the nearest neighborhood for anomaly detection is kept fixed at 10. It is chosen to be high enough so that it is meaningful in the sense where large numbers of clusters are observed while still being small enough to reduce computational costs. The results are represented in side-by-side

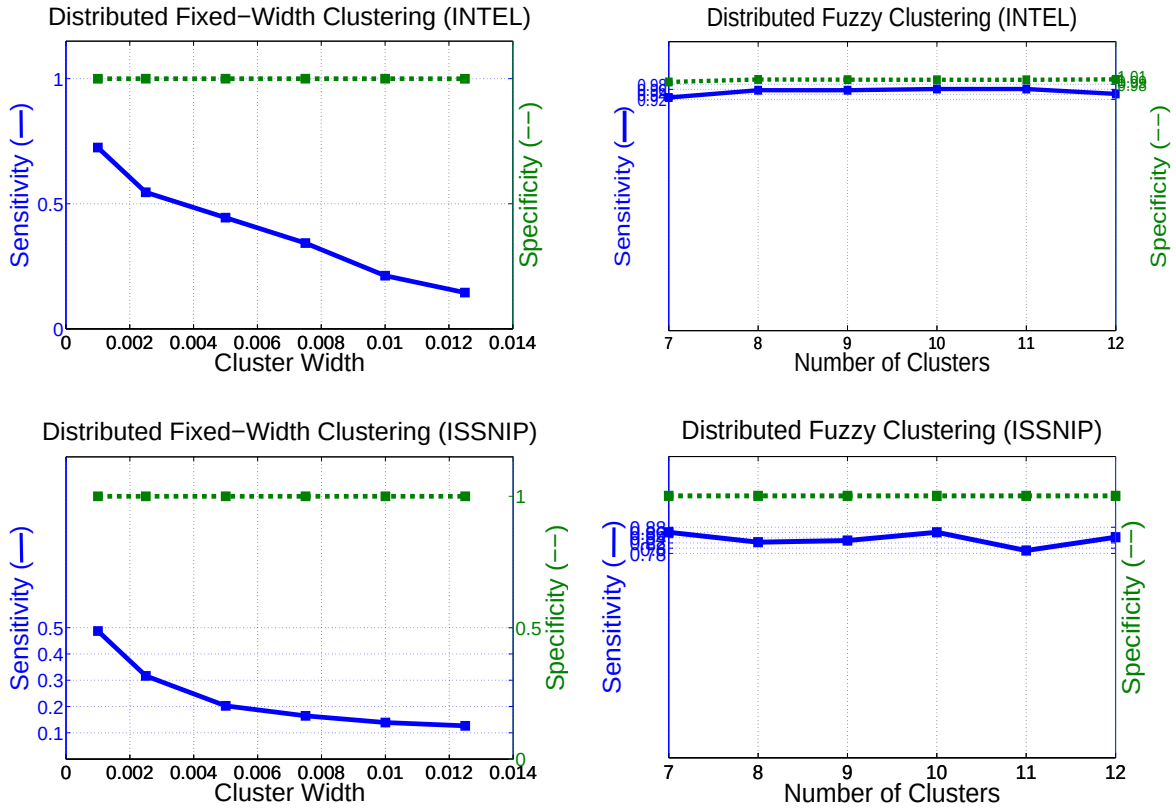

Figure 2.12: Comparison of Sensitivity and Specificity Variations for Anomaly Detection based on Distributed *fixed-width* Clustering (left) and Distributed *fcm* Clustering (right) for Data Sets based on INTEL (top) and ISSNIP (bottom) Distributions

plots (Figure 2.12) for comaprison with the variation in sensitivity and specificity obtained for our proposed method on the same data sets.The graphs show that both approaches maintain a very high degree of accuracy in determining normal data with the specificity being between 99%-100% in all instances. However, while the sensitivity values for the proposed scheme is more or less maintained at an acceptable range between 80% - 86% for ISSNIP data and a high 92% - 96% for Intel data, the results for the fixed-width clustering progressively decreases through out with the increase in cluster width. Furthermore, they are much lower compared to the proposed *fcm* based method as only being between 12% - 48% for ISSNIP data and 14% - 72% for Intel data through-out the investigation range. Therfore, it is clear that the proposed method achieves significant gains in detection accuracy compared to [Rajasegarar et al., 2006] while still being distributed, unsupervised and non-parametric with regard to dynamic and

heterogeneous data.

2.5 Conclusion

In this chapter a distributed anomaly detection framework is presented for industrial wireless sensor networks. There-in, soft partitioning of the data space is performed making use of the *fuzzy c-means* algorithm in an incremental model. An innovative data processing framework was implemented that efficiently and accurately classifies the data in identifying anomalies through maximizing distributed in-network processing in a hierarchical architecture. Evaluation of the proposed approach was performed for different data distributions representing different anomaly profiles based on both real and artificial data sets. The experimental results demonstrate very high detection accuracy with greater reductions in communication overheads on the network. These communication reductions highlight the scalability in extending the model flexibly for large scale distributed sensor networks. The proposed algorithm is comparitively analyzed with that of existing data clustering approaches and it's performance evaluated through sensitivity and specificity analysis. The downside of the current model is the requirement of the number of expected clusters to be specified at the local level. Future research will address this problem in achieving a dynamic allocation of cluster numbers through the specific information gained through each observational domain.

Chapter 3

Dynamic Data Partitioning with an Entropy Criterion for Multi-granular Anomaly Detection (DDP-EC)

This chapter proposes a fully dynamic and unsupervised data partitioning model for contextual anomaly detection in Wireless Sensor Networks (WSNs). This corresponds to the second research question on how to attain the number of data clusters dynamically in a data partitioning based anomaly detection approach for WSNs. A key limitation in the model developed in Chapter 2 is that the number of partitions are to be determined beforehand by the operator. Therefore, the number of clusters remain static unless operator intervention is made with regard to dynamic and evolving data streams. While it is shown in Section 2.4 that a cluster number that is not too small or not too large (as in the range between 6-12) will produce excellent results even over diverse data distributions, it is of interest to have a technique in determining this number dynamically from the data itself. Furthermore, the partitions that are produced in the previous model are limited to hyper-spherical clusters formed through the evaluation of data characteristics at or below second order statistics.

In addressing these issues and improving the formation of clusters towards a dynamic process, an entropy based data partitioning model is introduced. This captures data characteristics that are beyond second order statistics through a point-wise entropy measure that is directly used to determine the number and composition of data partitions in an adaptable and dynamic manner. A relative density measure is also defined to evaluate areas with differential density in identifying outliers and strengthening the entropy based data partitioning process. Furthermore, the proposed model is implemented on a hierarchical network topology for a WSN in the interest of providing multi-granular anomaly detection accompanied by low communication costs through a distributed and in-network data processing approach. The model is therefore specifically designed to fit the environment of large scale pervasive sensor networks with an aim of differentiating sensed data anomalies in a robust and scalable framework.

The rest of the chapter is organized as follows. Section 3.1 gives the contextual environment of industrial WSNs where dynamic data streams are encountered. It discusses the motivation for the proposed model and the specific contributions of this research. The related work and the rationale for integrating an entropy measure with local neighbourhood density for anomaly detection in the given context is presented in Section 3.2. Section 3.3 presents the proposed multi-granular anomaly detection model with subsections focusing on aspects of partitioning entropy based regions and techniques for non-parametric anomaly detection. The framework is evaluated using both real and synthetic data sets corresponding to a variety of data distributions for large integrated WSNs in Section 3.4. The classification accuracy is evaluated over different granular levels from local to global with regard to a hierarchical network topology in this section. Communication costs are also comparatively evaluated with respect to a centralized data processing approach. Section 3.5 concludes the chapter with a discussion of limitations and future work.

3.1 Motivation and Contributions

Wireless Sensor Networks (WSNs) are increasingly gaining widespread traction over a broad spectrum of applications that require an effective and pervasive monitoring mechanism [Buttyan et al., 2010, Bertocco et al., 2008, Luo et al., 2012, Guevara et al., 2012]. They consist mainly

of a large number of low-cost sensor nodes that can collaboratively collect, process and deliver information. Major advantages include scalability, manageability and extensibility within the context of observing large-scale geographically distributed environments/infrastructure, as well as a cost effective and interoperable method to perform multi-granularity monitoring with a variety of sensing options [Akyildiz et al., 2002, Puccinelli and Haenggi, 2005].

Streamlined sensing capabilities and the potential for aggregated intelligence (via parallel processing) in WSNs offer unique opportunities for distributed control as required in many industrial or environmental monitoring applications [Buttyan et al., 2010, Puccinelli and Haenggi, 2005]. Therefore, WSNs are becoming a key enabler for a multitude of process control systems that make decisions through distributed data sensing and communication capabilities [Akyildiz et al., 2002, Bertocco et al., 2008].

Accordingly, there is a vital need for a secure, continuous, reliable and effective sensing mechanism, one that places utmost importance on the integrity of the data. However, the nature of openly commoditised wireless technologies often employed, and the inherent limitations of the sensor nodes themselves in naive implementations would expose the application environment to a variety of threats and vulnerabilities [Djenouri et al., 2005, Shi and Perrig, 2004]. Such threats can: (i) affect the integrity of the network through path and node configurations; (ii) alter routing processes; (iii) introduce illegitimate network operations; (iv) perform illegal modifications or feed falsified data and (v) induce process monitoring blackspots [Djenouri et al., 2005, Shi and Perrig, 2004, Luo et al., 2006b, Phipatanasuphorn and Ramanathan, 2004]. Such threat vectors are easily realised through attacks on WSNs deployed for industrial equipment monitoring purposes [Mo and Sinopoli, 2012, Wang et al., 2010]. The need to deploy sensor networks unattended over a large geographic area in most applications exacerbates the situation further, with physical security of individual nodes not guaranteed [Di Pietro et al., 2009]. This can lead to instances of side-channel attacks [Ravi et al., 2004, Bar-El et al., 2006] and compromised security keys that open windows of intrusion. Another key concern is the unavoidable occurrence of faults in the sensor nodes that can lead to wrong measurements and impact key process parameters [Sharma et al., 2007, Sharma et al.].

Such situations of security attacks and faults often manifest as anomalies in data streams

and pose grave implications to decision-making processes at the heart of process monitoring implementations, and can easily threaten key components of the underlying application. If left undetected these data anomalies will lead to wrong operational and control decisions on industrial infrastructure that impact wider society through power blackouts, congestion and production loss, and can ultimately lead to severe economic losses, human harm and environmental damage. Recent research has only focused on securing WSN communications, often through defining more secure routing protocols [Datta, 2005, Villas et al., 2013] and managing link quality, security keys and reliability [Luo et al., 2006a, Yu and Guan, 2008]. Urgent focus is required for more data-centric security approaches that ensure the integrity of the sensed data. Therefore, methods for accurate and efficient detection of potential anomalies within the sensed data is of severe importance in WSN implementations.

However, for meaningful interpretation of anomalies, it is necessary to identify the common underlying aspects of WSN applications from a data-centric point of view as: (i) involving large amounts of sensed data; (ii) observing dynamic environments where the observation domain consists of unpredictable and dynamic distributions; (iii) heterogeneity of data with differing types in unmatched densities; and (iv) sensing functionality leading to data-driven control and core decision making. These aspects pose significant challenges for contextual anomaly detection that are yet to be addressed satisfactorily in the existing literature and include;

- A heterogeneous environment with dynamic data distributions, requiring an unsupervised, non-parametric and non-probabilistic data classification method.
- Non-homogeneous and multi-density data distributions within any instance of an aggregated observation domain, consisting of both individuals and correlated clusters of normal and anomalous behaviour.
- A resource constrained environment requiring high data-processing efficiency with reduced computational and communication complexities.

3.1.1 Contributions

The aforementioned challenges are addressed here with an innovative entropy based data partitioning framework that dynamically detects anomalies at different levels on a hierarchical topology. Granular evaluation of anomalies is achieved efficiently by using in-network data processing over several analytical stages that correspond to different network levels on an assumed node hierarchy.

More specifically, the data space that is aggregated on each sensor node for a specific time window ΔT, is partitioned onto different cohesive regions in a completely unsupervised and non-parametric process. The different regions are identified based on data correlations compared through a *point-wise entropy* measure. This is further complemented by a *relative density* factor for a defined local neighbourhood at each data point. This effectively curtails the effect of noise through prior segregation of individual outliers. At each analytical stage thresholds are derived adaptively and introduced to identify both anomalous data regions (partitions) and individual outliers. The average entropy value and corresponding density information is communicated along with identified local anomalies to the next hierarchical level. More global correlations are evaluated based on the entropy measure at that network level in identifying representative anomalies at that level of granularity. This is performed at multiple levels until the root of the hierarchy is reached for the considered WSN.

The framework is evaluated using different data distributions that represent a variety of dense and sparse regions for dynamic behaviour. Two major data sets are used based on the publicly available sensor data distributions from Australian Research Council's research network on Intelligent Sensors, Sensor Networks and Information Processing (*ISSNIP*) [Suthaharan et al., 2010] and Intel Research Laboratories [Bodik et al., 2004]. *Sensitivity* and *specificity* analysis is used to calculate the detection accuracy of both normal and anomalous data points while communication complexity is compared to that of a centralized data processing approach. The major technical contributions of the research are summarised below:

- *A dynamic and unsupervised data partitioning method that is responsive to heterogeneous and dynamic data*: A point-wise entropy measure using Euclidean distance similarity is

used to evaluate data correlations in partitioning the data on to different cohesive regions. The negative effects of noise and outliers are mitigated through average relative density comparisons that identifies and segregates such instances prior to the partitioning. The average relative density is defined to be compared over two localised distance neighborhood levels. Entropy is then used directly in determining the number and composition of data partitions directly as opposed to existing entropy based approaches [Gokcay and Principe, 2002, Temel and Aydin, 2007]. Significantly the number of partitions are determined dynamically without prior knowledge in contrast to most existing methods [Rokach, 2010]. Experimental results over a variety of dynamic data distributions demonstrate high classification accuracy of more than 94% in both sensitivity and specificity metrics consistently.

- *Non-parametric and non-probabilistic detection of anomalies that is robust and adaptive over different granular levels*: Anomalies are differentiated through introduced thresholds on observed mean entropy measures and corresponding average relative densities. These are derived from second order statistical information of *mean* and *standard deviation* over different analytical stages in the assumed hierarchy. Unlike in most existing methods [Chandola et al., a, Hodge and Austin, 2004], by focusing only on domain knowledge that is available locally at each network level, any definitions of arbitrary parameters or probabilistic assumptions are avoided. Therefore, a more accurate detection of anomalies is achieved in a robust and scalable manner for dynamic data distributions with differential density. Higher averages in the detected *true positives* and *true negatives* with a corresponding reduction in *false positives* and *false negatives* is observed for different dynamic data domains.

The use of a *distributed data processing framework* enabling a more *efficient* and *scalable* anomaly detection process further complements these innovations. Local data is aggregated over a defined sliding window with data processing performed in-network and in parallel at each node. Only the local anomalies and representative information on defined entropy based data partitions are sent between different network levels. This reduces communication overheads

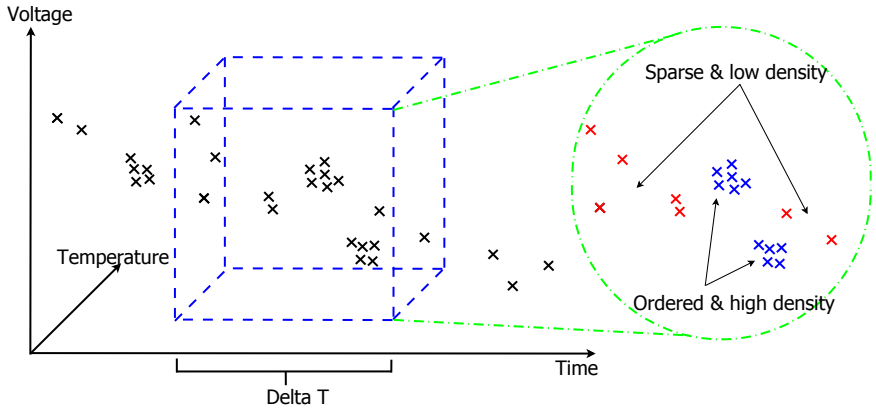

Figure 3.1: Dynamic & Non-homogeneous Data Distributions: The observed data in the context of WSN sensing and monitoring can be dynamic with the distribution changing unpredictably through time. Considering an aggregated distribution instance for a window of time ΔT, the data will also be non-homogeneous with both ordered and non-ordered areas differentiated through observation density.

dramatically when compared to a centralized data processing approach. The use of a data partitioning method also enables an incremental model where the data can be processed as they are sensed while anomalies are evaluated over different granular levels. Communication complexity investigations reveal a cost reduction of more than 85% compared to a centralized approach where analysis is performed globally.

3.2 Related Work and Rationale

Here, we discuss the rationale for adapting a data partitioning approach based on point-wise entropy and average local neighborhood density for anomaly detection. There is extensive literature regarding anomaly detection in both the statistics and database community [Chandola et al., a, Hodge and Austin, 2004, Ramaswamy et al.]. However, these approaches are not directly applicable in the context of a large scale WSN environment, either because they assume prior knowledge on input data distributions, require parametric or probabilistic assumptions or not being tailored to work in a resource constrained environment.

Another factor of key concern is the nature of the data itself. In the considered context the observation domain is assumed to be dynamic and non-homogeneous as highlighted in Figure 3.1. Therefore, regarding any considered anomaly detection approach, the data space is non

static with the distribution changing with time. Furthermore, an aggregated instance of the distribution at any time window ΔT can be non-homogeneous and of differential density.

Existing approaches can be classified onto different classes as (i) Distribution based, (ii) Distance based, (iii) Density based and (iv) Clustering based approaches. Of these, approximating distribution models require significant prior knowledge for data fitting as well as expensive testing/refitting. This is not practical in the sense of dynamically changing data and the low resource capacity environment of WSNs. On the other hand, distance based approaches [Knorr et al., 2000] identify outliers based on the fraction (β) of the data that is further than a defined distance (r) to the majority. However, this can lead to misclassifications and detection problems where the observation domain consists of both sparse and dense distributions within a single time window.

Density based techniques mine outliers depending on the local density of each observation [Breunig et al., 2000, Xie et al., 2012]. The local neighborhood is typically defined by a distance (r) that encompasses the K nearest data points. While this approach handles multi-density distributions with minimum assumptions/parameters, it suffers in the context of anomalies comprising of uniform density micro-clusters similar to normal data. In such instances the parameter K will have to be defined as large as the size of these clusters. Thus, sensitivity to this single parameter brings unreliability and reduced detection rates in the context of observations that comprise of isolated outliers and uniform density micro-clusters of abnormal behaviour. In [Papadimitriou et al., 2003], these issues are addressed using a *local correlation integral* based on statistical measures derived through local neighborhood densities. This identifies outliers non-parametrically and is robust considering dynamic distributions with different densities. However, it cannot detect instances of abnormal behaviour (consisting of micro-clusters) that has same density as normal data while being significantly different in individual values.

Data partitioning/clustering approaches stand out as the most viable in this context when implemented with a non-parametric approach. They offer fully unsupervised classification without any prior knowledge/assumptions and can be adapted over an incremental model that offers different granularity. This is best attained in a hierarchical network topology that

performs the partitioning distributively and evaluates anomalies incrementally over different levels. However, existing data clustering approaches are not optimized for anomaly detection and are mostly centralized processes with high complexity and resource costs. Furthermore, inferring outliers through an integrated formula cannot easily be achieved as the explicit goal is segregating data groups with only implicit attention on the effects of noise/outliers [Xu and Wunsch, 2005, Rokach, 2010].

Existing data clustering algorithms such as [Kumarage et al., 2013] are focused mainly on the two aspects of *distance* or *density* to identify data partitions. Such approaches are limited as they only capture second order characteristics of statistics and tends to work well only for distributions with hyper-spherical or hyper-elliptical clusters. Another concern is how to define the number of expected clusters dynamically without prior knowledge or resorting to static definitions. These concerns are overcome in [Yao et al., 2000] by using point-wise entropy to directly cluster the data towards a fuzzy inference model. However, it uses predefined direct thresholds β and γ in removing the impact of noise and defining the similarity boundary for each cluster. This is purely arbitrary and cannot be adapted for anomaly detection in the context of dynamic multi-density distributions. On the other hand density based clustering methods cannot be used for anomaly detection when abnormal behaviour can consist of regions with similar density to the normal majority. Therefore, a distributed data clustering approach is mandated that directly considers information characteristics above second order statistics (using an appropriate entropy measure) and is robust for multi-density dynamic data observations.

3.3 Distributed Multi-granular Detection Model

In this section we describe the proposed distributed data partitioning framework with the aim for an accurate and efficient detection of observed anomalies over different granular levels in a WSN environment.

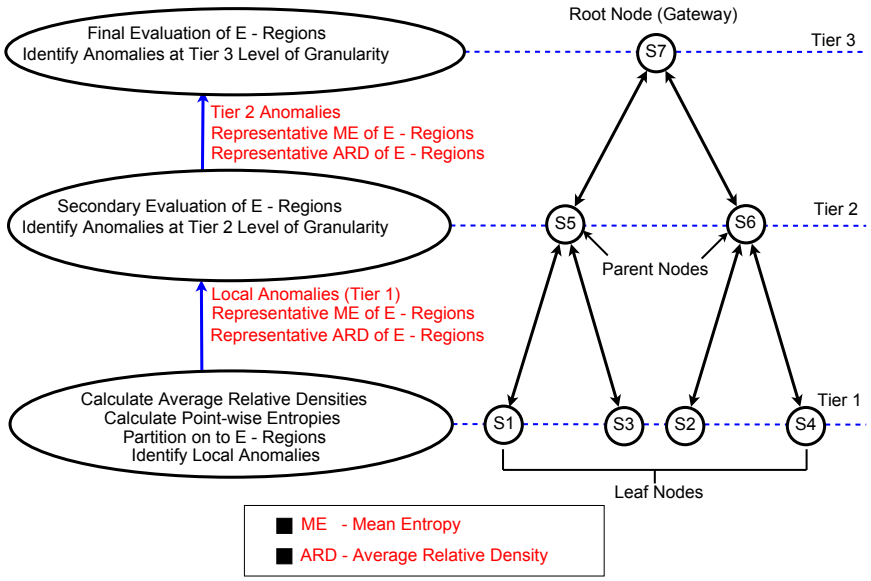

Figure 3.2: Anomaly Detection Architecture: Multi-granularity analysis of anomalies is performed on a hierarchical node topology. The individual processing steps performed at each node are shown respective to network level with the specific data that is communicated among different tiers for a two level sensor node hierarchy. E - Regions are the different data partitions as identified through an entropy criteria

3.3.1 The Sensor Network Model

We consider a network model consisting of a hierarchical topology of sensor nodes. Therefore, the nodes are ordered on to several tiers offering different levels of sensing granularity. These range from individual observation domains at the lowest tier to a domain encompassing the entire network at the highest tier. For each node, there is a *parent* node that receives information and is aware of the status of its *children* (nodes). For example in Figure 3.2, nodes S5 and S6 are parent nodes with S1,S2 and S3,S4 being their respective child nodes. The hierarchical organization of the network can be achieved using any of the existing techniques in the literature [Ganesan et al., a, Sohrabi et al., 2000, Zhang et al., 2008, Zhao et al., 2004, Liu and Haenggi, 2006] and it is assumed that this is done prior to the functionality in the proposed model is implemented.

The model is implemented in-network considering a sliding time window of ΔT at each sensor node. Therefore, each node aggregates it's local observations within the time window and performs the following tasks in a data driven non-parametric approach. First, the *average*

relative densities of each data point are computed in order to identify sparsely populated regions and individual outliers/noise. Then the data distribution is modelled on to different cohesive regions based on the *point-wise entropy* of each point vector considered. Based on these defined regions and corresponding relative densities at each observation, anomalies are identified as both individual data points and cohesive clusters of abnormal behaviour. The resultant local anomalies (at each leaf node) are then communicated together with representative information (mean entropy and related average relative density) on identified regions to the parent nodes at each level. This information is processed at the parent node in evaluating more global correlations and identifying anomalies at that level of granularity. This process is performed recursively until the root of the hierarchy is reached for the concerned (sensor node) topology. Figure 3.2 graphically summarises the architecture of the proposed model. We look at each of the individual tasks in detail in the following.

3.3.2 Estimating Local Density Variations

In the proposed method, the average relative densities for all observations within the considered time window is calculated first. This is carried out in order to identify sparsely populated regions consisting of isolated outliers prior to the actual partitioning of the data. Thus, any effect of these outliers on the partition process is mitigated. We use the following definition in calculating an appropriate measure for such.

Definition 1. *: Average Relative Density*

Consider a set of observations $X = [X_1, X_2, ...X_i...X_n]$*, where each data point* X_i *is a* m*-dimensional observation, with* $X_i = (x_{i1}, x_{i2}, ..., x_{im})$*. Then, the average relative density* (d_{ar}) *of any data point* X_i *is defined as follows based on two distance neighborhoods* r *and* R*.*

$$d_{ar}(X_i, R, r) = \frac{density(X_i, r)}{\{\sum_{Y \in N(X_i,R)} density(Y, r)\} / \left|N(X_i, R)\right|}$$

- *The density of an individual data point is defined as the number (*n*) of observations within a defined distance neighborhood.*

- *Two comparative distance neighborhoods are defined as the* Counting Neighborhood (r) *and* Sampling Neighborhood (R).
- *The counting neighborhood is the neighborhood of radius (r), over which the density (n) is estimated for each data point.*
- *The sampling neighborhood (R) is the neighborhood of radius (R), over which we collect samples in order to estimate the average.*
- *The subset of observations in X falling within the sampling neighborhood (R) of X_i is taken as $N(X_i, R)$ with any such instance represented by Y. Therefore, $N(X_i, R) \equiv \{Y \in X | dist(Y, X_i) \leq R\}$ with $dist(Y, X_i)$ being the simple euclidean distance between Y and X_i.*

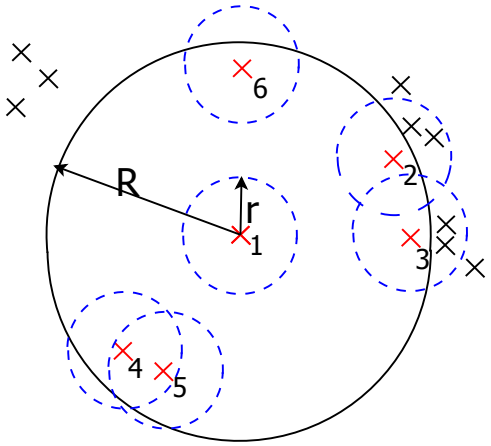

Figure 3.3: Average Relative Density: Average density at a data point is calculated relative to two distance neighborhoods (*Counting Neighborhood (r)* and *Sampling Neighborhood(R)*). The number of observations on each is taken as the local density for each neighborhood. In the above example average relative density is calculated for observation instance 1. Accordingly, six data points come under the sampling neighborhood R each with local densities of 1,3,3,2,2 and 1 (for instances 1 - 6) respectively over counting neighborhoods r. Therefore, the average density at instance 1, is 2(=12/6). Average relative density is then obtained as 0.5(=1/2) through the ratio of local and average densities according to definition above

This definition results in an appropriate measure when comparing local density variations within a larger data distribution as required for contextual outlier identification. An example is given in Figure 3.3, that illustrates the calculation of average relative density over two

distance neighborhoods. In this research we have used R as 0.75 in the normalized distance distributions over the closed interval $[0, 1]$ with a corresponding 0.1 value for r. Thus, r is defined on a range that is small enough to identify minute density variations within the data while the larger value for R provides compensation in a more global view. Therefore, this criteria gives a quantitative measure representative of the degree that a particular observation is isolated compared to it's neighbors. This is then vitally used towards identifying sparsely populated regions and individual outliers based on a defined threshold of one *standard deviation* (σ) away from the *mean* (μ). Thus, for the set of calculated average relative density values, any value that is one standard deviation (σ) away from the observed mean (μ) (regarding the same set of average relative density values) is declared as an outlier. Considering the set of average relative density values as $A = [A_1, A_2, ..A_i, ..A_n]$ with each $A_i = d_{arX_i}$ the threshold is given as follows

$$T = 1/n \sum_{i=1}^{n} A_i + \sqrt{1/n \sum_{i=1}^{n} [A_i - (1/n \sum_{i=1}^{n} A_i)^2]}$$

3.3.3 Estimating Entropy based Data Regions

Once the average relative densities have been computed and sparsely distributed outliers identified, the data is modelled onto different cohesive regions or partitions based on entropy. There are several techniques that make use of entropy in partitioning a particular data distribution [Gokcay and Principe, 2002, Temel and Aydin, 2007]. However, almost all of them use entropy as a measure to validate already identified partitions or optimize an existing number of partitions. Therefore, most of them involve defining an initial partitioning using a separate existing algorithm and later estimating the probability density functions and such (*pdfs*) for each partition in order to measure entropy for validation. This implies that the resultant data partitions are representative of the similarity measure used in the original partitioning/clustering algorithm and are not using an entropy criteria as a property that binds data points to a specific cluster. However, in our approach we divide the data distribution into cohesive partitions/regions ($E-Regions$) using the cumulative point-wise entropy of each data point directly. This has the following desirable properties. (i) Provides a single value criteria

in clustering members together. (ii) Non-parametric calculation when coupled with distance based similarity measure. (iii) Not limited to hyper-spherical or hyper-elliptical data clusters. (iv) Captures data characteristics that are not limited to second order statistics. (v) Adaptive to dynamic data distribution changes and (vi) The number of similarity regions is determined algorithmically without prior knowledge.

Definition 2. *: E - Region*

Let us consider a set of observations $X = [X_1, X_2, ..X_i, .., X_n]$*, where each data point* X_i *is a m-dimensional observation with* $X_i = (x_{i1}, x_{i2}, ..., x_{im})$*. Then a group of E - regions* $E_1, E_2, ..E_i, .., E_k$ *are defined over the spatial distribution of* X *based on entropy similarity and constituting of a subset of all possible subsets of* X *where,*

- *Each defined region* E_i*,*E_j*, is disjoint from one another, i.e.* $E_i \cap E_j = \emptyset, \forall i, j$*.*
- *Each defined region* E_i*, has a maximum similarity difference on entropy between it's constitutive elements that is less than* one standard deviation *(*σ*) of the overall pointwise entropy distribution observed within that region.*
- *We have:* $E_i(max_e) - E_i(min_e) < (1)\sigma$ *holds* $\forall i$ *where,* max_e *and* min_e *are the maximum and minimum entropy values observed between any two elements in* E_i*.*

In deriving the measure of entropy similarity an entropy value for each data point relative to other observations within a concerned data distribution is calculated first. We make use of the following point-wise derivation based on Shannons Entropy definition [Yao et al., 2000, Temel and Aydin, 2007] in determining the entropy value e_{ij} between two data points X_i and X_j using a measure for similarity S_{ij}.

$$e_{ij} = -S_{ij} \log S_{ij} - (1 - S_{ij}) \log(1 - S_{ij})$$

Therefore, the cumulative point-wise entropy for any observation $i \in X$ is given by,

$$e_i = -\sum_{j \in X}^{j \neq i} (S_{ij} \log S_{ij} + (1 - S_{ij}) \log(1 - S_{ij}))$$

where S_{ij} is the similarity measure between data points x_i and x_j normalized to be in the closed interval [0.0 - 1.0]. We take the similarity between two data points as given by,

$$S_{ij} = \exp^{-\alpha D_{ij}}$$

D_{ij} is defined as the proximity measure between the two data points. We use the *euclidean distance* as it is efficient to compute and provides effective proximity.

$$D_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + ...(x_{im} - x_{jm})^2}$$

α is taken as the corresponding value when the similarity measure is substituted by 0.5 similarity (as to be non-partial over a normalized input range) and the mean distance between all data points. Therefore, it is calculated adaptively based on the current data distribution in avoiding arbitrary assumptions as follows;

$$\alpha = -\ln 0.5/\bar{D}$$

Based on these definitions and findings we dynamically partition the data space on to different regions as follows. First the cumulative point-wise entropy for each data point is evaluated by calculating the normalized euclidean distances for the considered data distribution on each sensor node. Then the data point with the least entropy value is selected as the ordered point around which the region is to be defined. However, if the chosen data point lies in the set of points identified as one of sparsely populated individual outliers/noise (as derived through average relative densities in Section 3.3.2), it is removed and the next point with the least entropy is chosen. Then a similarity region is defined based on the entropy by assigning other data points with similarity less or greater than one standard deviation

(σ) of the original selection. Note that the individual outliers identified previously are also considered at this step.These data points are then removed and the next region defined based on the next observed least entropy value. This is repeated until no data points are left and the data distribution is completely partitioned on to disjoint regions with similar entropy while mitigating the effects from local density variations.

3.3.4 Anomaly Detection

Once the considered data distribution at each sensor node per time window ΔT is partitioned in to a set of cohesive similarity regions as detailed in the previous section, anomaly detection is performed as follows. Anomalies are identified based on two major characteristics as (i) individual data points that are anomalous and (ii) E - regions where all constitutive data points on that region are considered anomalous.

Individual Anomalies: These are data points that are located with sparse density on the considered data distribution and are not a member in any of the defined E - regions. Thus, we define an individual anomaly as follows.

Definition 3. *: Individual Anomaly;*

Any data point $X_i \in X$ that is declared an anomaly will satisfy the following criteria,

- *The average relative density (d_{ar}) of X_i is $\geq T$,*
  where $T = 1/n \sum_{i=1}^{n} A_i + \sqrt{1/n \sum_{i=1}^{n} [A_i - (1/n \sum_{i=1}^{n} A_i)^2]}$,
  with A being the set of average relative densities with each $A_i = d_{arX_i}$.
- $X_i \notin E_j, \forall i, j$

Anomalous Regions: These are complete E - regions that are defined according to the proposed data partitioning criteria and identified to be anomalous by the proposed model. These represent dense groupings of data points comprising abnormal behaviour.

Definition 4. *: Anomalous Data Region;*

An E - region identified to be anomalous will have the following properties.

- *A mean entropy value for the region that is higher than one standard deviation away from the average mean entropy value observed at all other defined E - regions for that node.*

- *The relative density of the data point with entropy value that is closest to the mean entropy of the region will have a far lower value compared to the highest relative density observed at that node. Therefore, we use a threshold of* < 0.25 *of* $\max d_{ar}$ *in classifying the region as anomalous.*

As the above definitions make clear one standard deviation away from the mean is used as a parameter in differentiating anomalies. However, as it is a value that is derived from the data itself and applied on both average relative density and mean entropy values independently over different stages it acts as an adaptive parameter as opposed to a more direct one (which can be arbitary) and in keeping with the general assumption that a majority of the data is assumed to be normal.

3.3.5 Distributed Multi Granularity Analysis

Once the data is partitioned and the local anomalies identified with corresponding E - regions at each node, the following information is communicated to respective parent nodes in the hierarchy.

- Detected local anomalies

- The data point with entropy value that is closest to the mean (μ) entropy of each E - region (This data point represents the corresponding E region)

- The locally calculated entropy value (e_i) of each representative data point and its average relative density (d_{ar})

Therefore, each parent node combines the received information in compiling the set of anomalies at that level of network granularity. The entropy values and average relative densities on each E - region is therefore considered based on the union set of all partitions (E - Regions) received at the parent node. The same criteria in identifying anomalous regions as

defined before is applied and the corresponding abnormal regions identified without resorting to additional assumptions. The received individual anomalies are kept as such without any further refinement. The resulting anomalies and the representative information on E regions are then sent to the parent nodes on the next network tier. This is performed recursively until the root of the hierarchy is reached.

If any region that was locally identified as normal is now detected as anomalous, the corresponding member points are requested from the child node. On the other hand, any E - region that was earlier detected as abnormal but is now normal will only be represented by its representative data point as mentioned before, when communicated to the next level. Therefore, at each network level the detected anomalies are refined based on different granularity levels that evaluates correlations across multiple nodes in that network tier based on the expanded domain knowledge. This cross evaluation is vital in the case that a particular sensor node is compromised with all its observations being anomalous. Such data will only be detected as abnormal when compared to a neighboring node as achieved here.

3.4 Results & Evaluation

The proposed model is evaluated through a combination of both real and synthetic data sets. Different data sets are used to highlight the robustness of the model regarding multi-density dynamic distributions. Classification accuracy for both normal and anomalous data points are analysed with the communication overheads over the network hierarchy compared to that of a centralized approach. Experiments are performed using *sensitivity* and *specificity* metrics as the main evaluation criteria. We simulate an assumed three level hierarchical architecture for a wireless sensor network as presented in Figure 3.2. The data sets are then taken as corresponding to the bottom tier of four sensor nodes S1 - S4 which communicates with parent nodes S5 and S6. These in turn are under the purview of root node S7.

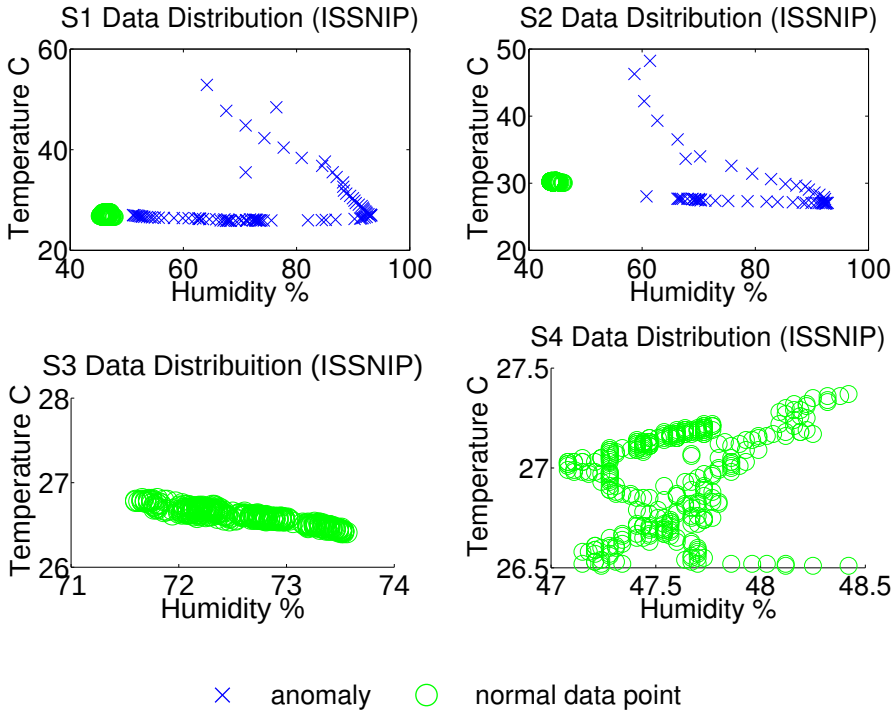

Figure 3.4: Data Distributions based on the ISSNIP Data Set: The two distributions with anomalous data (*top*) represent a set of sequentially distributed anomalies with two tightly correlated normal data sets where one is spatially more closer to the anomalies (*left*) and the other more spatially distributed (*right*). The (*bottom*) two distributions consist of tightly correlated normal data sets where one is sequentially concentrated (*left*) and the other more spatially distributed (*right*).

3.4.1 Data Sets

Two data sets are used consisting of both real sensor mote measurements and artificially created data points that act as introduced anomalies. These are derived as to represent the dynamic multi-density observation environment for a typical WSN application environment. The first data set is based on the annotated data with labelled anomalies available from the Australian Research Council Research Network on Intelligent Sensors, Sensor Networks and Information Processing (*ISSNIP*) [Suthaharan et al., 2010]. Four different data distributions are created corresponding to the sensed data of the first four sensor motes (mote 01 - 04). This data has been collected through the deployment of a multi-hop WSN using TelosB sensor motes measuring temperature and humidity fluctuations over a period of 6 hours in 5 second

intervals. Anomalous observations are included at motes 01 and 03 by artificially manipulating the temperature and humidity values in the observed environment. In our experiments the data from these four sensor motes are taken as corresponding to the data distributions of leaf nodes S1 - S4 according to the assumed three level hierarchy depicted in Figure 3.2. Therefore, nodes S1 and S2 will have anomalous data while S3 and S4 contain only normal measurements. The four data distributions are shown in Figure 3.4.

Table 3.1: Data Format (Intel); Each observation consists of 4 mote attributes and 4 measured parameters

Attrib. 1	Attrib. 2	Attrib. 3	Attrib. 4	Para. 1	Para. 2	Para. 3	Para. 4
date:y-m-d	time:h:m:s	epoch	moteid	temperature	humidity	light	voltage

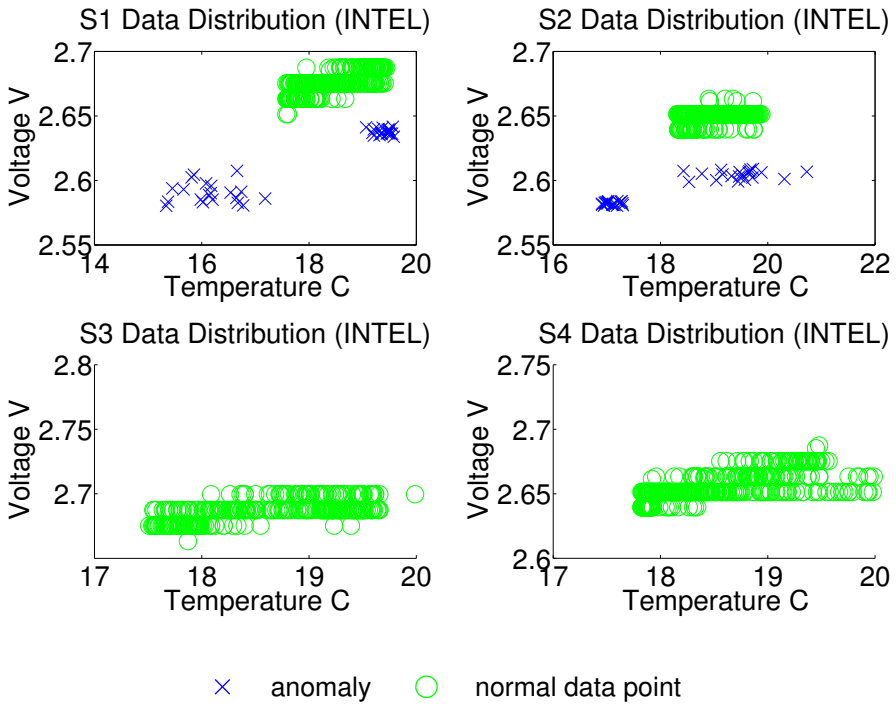

Figure 3.5: Data Distributions based on the Intel Sensor Data Set: Two data distributions representing both spatially dispersed and spatially concentrated clusters of anomalous points are presented (*top*). One features the higher density anomalies closer to the similarly dense normal data (*top, left*) with the other having less dense anomalies close (*top, right*). The (*bottom*) two data sets depict two complete distributions of tightly correlated and spatially focused observations representing normal behaviour.

The second data set is based on the public sensor mote data from Intel Berkeley Research Laboratory and consists of real measurements collected from 54 sensors [Bodik et al., 2004]. Time-stamped topology information is included along with the measured parameters of temperature, humidity, light and voltage using Mica2Dot sensors with weatherboards. The data format takes the form shown in Table 3.1 and isn't annotated. Therefore, in the absence of labelled information on anomalies it is used in our evaluations as follows. The temperature, humidity, light and voltage attributes together are taken as data vectors corresponding to individual observations. Subsequently, four hundred (400) of these observations are considered in sequence per single node in representing the aggregated data over a considered time period ΔT.

First, extreme values in the data are identified through scatter plots as isolated anomalies. The remaining data is labelled as normal for evaluations. Other annotated anomalies are then introduced using two sets of randomly generated data at two of the nodes. Two types of anomaly distributions as randomly dispersed and more focally concentrated are represented via the generated data points. In generating such, random sets of twenty (20) data vectors are drawn from the standard *uniform* and standard *normal* distributions considered over the normal measurements. These are merged with the data distributions derived earlier for mote 01 and 03. The data sets of mote 02 and 04 are kept as originally derived without any anomalies in representing only normal data. The resulting four distributions are then assumed to represent the observation domains of leaf nodes S1 - S4 according to the assumed three level hierarchy as stated before. Therefore, the data distributions of node S1 and node S2 will have the labelled anomalies while nodes S3 and S4 has no anomalous data and represent completely normal behaviour. The data distributions for each of these sensor nodes are as depicted in Figure 3.5.

3.4.2 Evaluation

The proposed framework was implemented for a simulated three level sensor node hierarchy on the matlab environment. Experiments were carried out in investigating classification accuracy and communication complexity in terms of the data communicated using the two created data sets. Therefore, considering each analytical stage on the hierarchy the *False Positive Rate*

(FPR) and the *False Negative Rate* (FNR) were calculated based on the number of False Positives (FP) and False Negatives (FN) observed at each node. A false positive is an instance where a normal measurement is detected as anomalous while a false negative is an instance where an anomalous measurement is detected as normal. The false positive rate is the ratio between the detected number of false positives and the actual normal measurements. The false negative rate is the ratio between the detected number of false negatives and the actual number of anomalies for the considered data distribution.

$$FPR = \frac{FP}{(FP + TN)} \qquad FNR = \frac{FN}{(FN + TP)}$$

Similarly, the corresponding instances of True Positives (TP) and True Negatives (TN) are also determined at each stage leading to the computation of the main *Sensitivity* and *Specificity* values. Sensitivity indicates the probability that a statistical test is positive for a given true positive statistic, while specificity indicates the probability that a statistical test is negative for a given true negative statistic. These are calculated as follows to be the major evaluation criteria for data classification accuracy.

$$sensitivity = \frac{TP}{(TP + FN)} \quad specificity = \frac{TN}{(TN + FP)}$$

The number of data points that are communicated on each wireless link is noted over the three level hierarchy in determining communication savings compared to a centralized approach, where all the observations are communicated to the root node. We begin our discussion by analysing the cumulative point-wise entropy distributions for each data distribution. Limitations for anomaly detection are highlighted when the focus is only on entropy without a contributing density factor. Then, we analyse the classification accuracy of the framework for different data distributions based on calculated metrics on a local, intermediate and final level in the considered three level hierarchy. We conclude the evaluation through a discussion on communication complexity complimented by the experimental results for data reduction over different wireless links. The scatter plots highlighting the identified data partitions in *E - Regions* and their distribution as well as isolated anomalies corresponding to the analysis through mean entropy and average relative density in each data distribution are also presented.

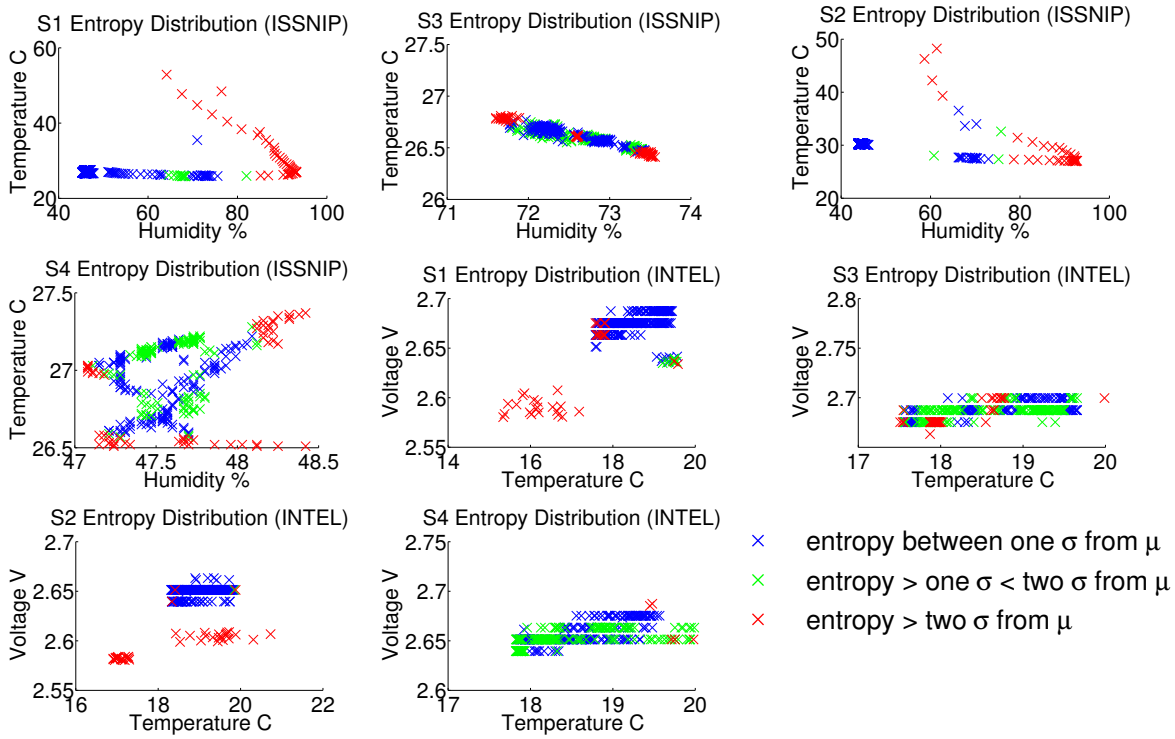

Figure 3.6: Entropy Distributions: Data is plotted based on the cumulative point-wise entropy value of each data point over it's local distribution. Each data point is presented as comparative to the *mean* entropy value on each distribution with the differentiation noted on a *standard deviation* basis. It is highlighted that for multi-density distributions similar entropies are allocated for both normal and anomalous data depending on the orderliness and *local area density* at each data point. Therefore, it is clear that an approach based purely on entropy alone is not sufficient enough to differentiate between anomalous outliers/clusters and the normal data when the considered distributions are non-homogeneous with varying density.

3.4.2.1 Entropy Distributions

Here, we plot the data distributions on both INTEL and ISSNIP data sets based on the calculated values for cumulative point-wise entropy. Figure 3.6, presents these through a three level categorization using the *mean* (μ) and the *standard deviation* (σ) of entropy values on each distribution. We use the intervals to be within one σ of μ, between one and two σ away from μ and more than two σ away from μ. The mean is chosen as the central value to compare as for any anomaly detection criteria the major assumption is that the majority of the data is normal and that the anomalies will be a minority that is significantly different. Thus, we use

μ to represent a middle ground entropy value representing the majority of normal data and compare the deviations. This is done as from information theory, entropy is less for orderly data configurations and high for disorderly configurations.

It can be seen from the corresponding scatter plots that while a majority of normal data resides within one σ from μ and a majority of anomalies in two σ away from μ, some of the normal data is also two σ away with some anomalies being within one σ. This apparent contradiction is due to the fact that the data distributions are representative of dynamic multi-density observation domains. Therefore, the normal data can consist of relatively sparse regions bordering a majority dense region while the majority of anomalies can be sparse with small dense micro-clusters possible in the same data distribution. Thus, when point-wise entropy is calculated it will result in both ordered dense regions (in normal data) as well as small high density areas (in anomalies) having similar values. On the other hand some of the sparsely populated fringe areas of normal behaviour will have similar entropy values with the majority sparse anomalies. Therefore, it is clear that a data partitioning approach that considers only entropy will not be ideal in detecting anomalies where the observation domain is non-homogeneous with multi-density distributions possible. This results in the need for a complimentary density factor that leverages on the entropy deviations for orderly and disorderly configurations in having an effective data partitioning for anomaly detection.

Table 3.2: Classification Accuracy : ISSNIP Data Distributions (Local Phase)

Sensor Node	FN	FNR %	FP	FPR %	Sensitivity	Specificity	TN	TP
S1 (ISSNIP)	3	0.0300	0	0.0000	0.9700	1.0000	300	97
S3 (ISSNIP)	0	NAN	0	0.0000	NAN	1.0000	400	0
S2 (ISSNIP)	0	0.0000	0	0.0000	1.0000	1.0000	342	58
S4 (ISSNIP)	0	NAN	22	0.0550	NAN	0.9450	378	0

Table 3.3: Classification Accuracy : INTEL Data Distributions (Local Phase)

Sensor Node	FN	FNR %	FP	FPR %	Sensitivity	Specificity	TN	TP
S1 (INTEL)	0	0.0000	18	0.0450	1.0000	0.9550	382	40
S3 (INTEL)	0	NAN	12	0.0273	NAN	0.9727	428	0
S2 (INTEL)	1	0.0250	18	0.0450	0.9750	0.9550	382	39
S4 (INTEL)	0	NAN	1	0.0023	NAN	0.9977	439	0

Table 3.4: Classification Accuracy : Distributed Phase (Parent Nodes)

Sensor Node	FN	FNR %	FP	FPR %	Sensitivity	Specificity	TN	TP
S5 (ISSNIP)	3	0.0300	0	0.0000	0.9700	1.0000	700	97
S6 (ISSNIP)	0	0.0000	22	0.0296	1.0000	0.9703	720	58
S5 (INTEL)	0	0.0000	30	0.0357	1.0000	0.9642	810	40
S6 (INTEL)	1	0.0250	19	0.0226	0.9750	0.9773	821	39

Table 3.5: Classification Accuracy : Final Results (Root Node)

Sensor Node	FN	FNR %	FP	FPR %	Sensitivity	Specificity	TN	TP
S7 (ISSNIP)	3	0.0189	22	0.0152	0.9810	0.9847	1420	155
S7 (INTEL)	1	0.0125	49	0.0291	0.9875	0.9708	1631	79

3.4.2.2 Classification Accuracy - Local Phase

Performance on the initial local data partitioning phase is evaluated first. The accuracy of data classification with respect to identifying both normal and anomalous data points is performed through *sensitivity* and *specificity* analysis. Corresponding results on data distributions for ISSNIP and INTEL data sets are as given on Table 3.2 and Table 3.3. The data distributions with abnormal behaviour for the ISSNIP data set represents a set of spatially dispersed (sparse) sequential anomalies compared to the tightly correlated dense majority of normal data. Each has areas of different density in the anomalous data region offering challenges to traditional means of anomaly detection. However, the results in Table 3.2 show that the proposed approach achieves very high accuracy in classifying both normal and anomalous data. While an ideal specificity of 1.0 is achieved in identifying normal data for both distributions at nodes S1 and S2 the respective sensitivity values are 0.97 and 1.0. The very slight reduction for node S1 is due to the fact that a minor portion of the normal data is spatially very close to the anomalies offering little variations in both average relative density and point-wise entropy compared to the data at node S2. Considering the data distributions with no anomalies at node S3 and S4 the proposed approach achieves very high specificity values of an ideal 1.0 and 0.94. Again the slight drop in specificity corresponds to the sparse individual outliers at node S4 that are very close to the majority dense area while the tightly correlated but sequentially spread out distribution in node S3 offers an ideal scenario of normal behaviour.

Considering the INTEL data set, the distributions for normal behaviour are tightly correlated dense regions with only discrete variations over a minute range for the observed voltage parameter. The added anomalous behaviour offer two regions each of multi-density data representing both higher and lower relative densities to the majority normal data. The proposed method again achieves very high accuracy in detecting anomalies with sensitivity values of 1.0 and 0.97 for nodes S1 and S2. Similarly the classification of normal data in nodes S3 and S4 remain high at 0.97 and 0.99 respectively. Therefore, it is clear that the proposed techniques achieve very high detection accuracy with higher *True Negative* and *True Positive* rates with corresponding lower *False Positive* and *False Negative* rates. This is shown to be achieved consistently for different multi-density distributions representative of a dynamic observation environment through the different data sets used. The scatterplots for the data partitioning through identified $E - Regions$ and the isolated anomalies are presented in Figures 3.7 and 3.8 respectively for both ISSNIP and INTEL data sets.

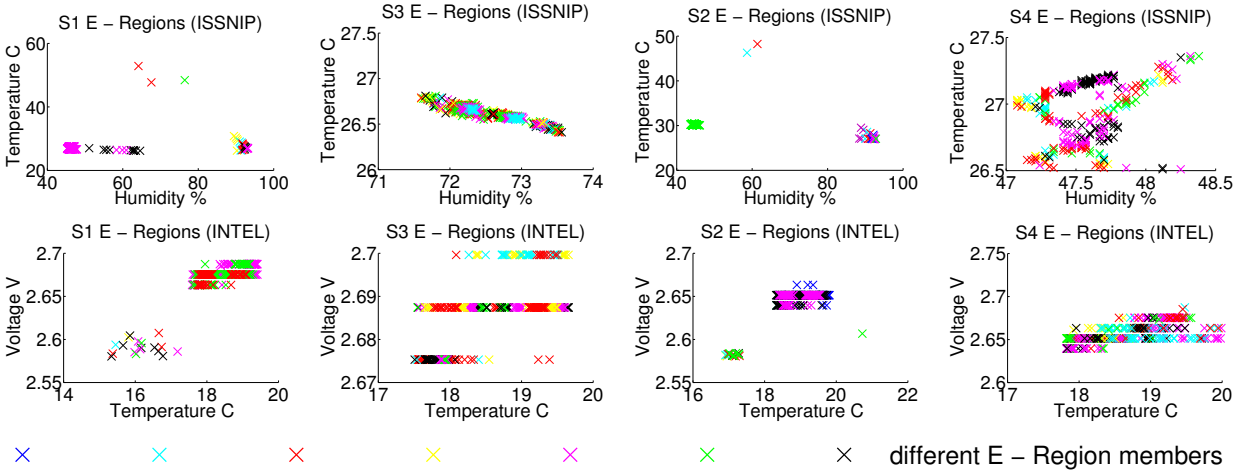

Figure 3.7: Identified E - Regions: Data distributions partitioned on to different cohesive $E - Regions$ using both *cumulative point-wise entropy* as well as *average relative density* for ISSNIP and INTEL data sets. Granular partitions are obtained that reflects *both* orderliness through the entropy measure and local density variations through *average relative density* towards an effective anomaly detection approach. Unlike in Figure 3.6 where only entropy is considered the number of granular partitions are determined adaptively to represent the nature of the distribution as closely as possible.

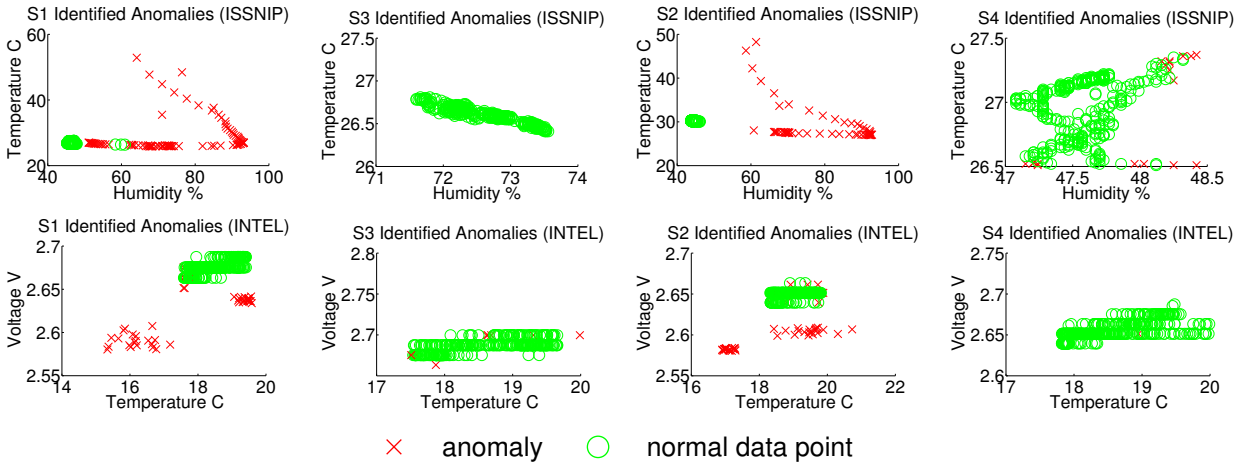

Figure 3.8: Isolated Anomalies: The detected anomalies and the normal data as classified through the proposed approach is presented regarding the local data partitioning phase at each sensor node for ISSNIP and INTEL data sets.

3.4.2.3 Classification Accuracy - Distributed Granular Phase

The intermediate step of anomaly detection offering second tier level of granularity in the considered topology is performed at nodes S5 and S6. The results for both ISSNIP and INTEL data sets are as given in Table 3.4. Anomalies are identified based on the received local anomalies and representative information on $E-Regions$ from the previous local partitioning phase. Therefore, the observation domain will expand to include both normal and anomalous data regions at each node with corresponding density variations. Here, the detection accuracy is maintained consistently as in the previous local phase with sensitivity values of 0.97 and 1.0 with specificity of 0.97 and 1.0 for the ISSNIP distributions at nodes S5 and S6 respectively. Similarly, the results are 1.0 and 0.97 in sensitivity and 0.96 and 0.97 in specificity for the INTEL data set at nodes S5 and S6.

The final values for classification accuracy through sensitivity and specificity analysis is given in Table 3.5. The *False Positive* and *False Negative* rates as well as the *True Negative* and *True Positive* rates are calculated considering the global observation domain at root node S7. Therefore, anomalies are detected over the complete hierarchy consisting of the union set of all four data distributions for each ISSNIP and INTEL data set. We achieve a very high final

sensitivity value of 0.98 for both data sets while the specificity comes up to a similarly high 0.98 and 0.97 for ISSNIP and INTEL respectively. Thus, it is demonstrated that the proposed method achieves high detection accuracy for both normal and abnormal behaviour consistently throughout the hierarchy. No shifts in detection accuracy are observed as the observation domain expands and manages to offer consistently accurate results at multi-granularity levels for different dynamic data distributions.

3.4.2.4 Communication Complexity

Considering the inherent power limitations in wireless sensor networks, it is critical that any proposed data processing framework limits any overheads on energy consumption. In WSNs the majority of consumed energy is spent on data communication rather than computation [Pottie and Kaiser, a, Li et al., 2011, Raghunathan et al., 2002]. Thus, special emphasis should be given on reducing the amount of data communicated on each wireless link for a given sensor network topology. Here, this is investigated considering the communication overheads that are incurred on each wireless link for the proposed hierarchical data processing framework. We compare the amount of data that is communicated to that of a centralized approach where all sensed data is communicated over each considered time window. The communication overhead is calculated based on the ratio of the number of messages communicated in the proposed model to that of a centralized approach in Tables 3.6 and 3.7 for the ISSNIP and INTEL datasets respectively. Therefore, experimental results are given for each of the four wireless links at the first tier and the succeeding two links at the second tier on the considered hierarchy.

Table 3.6: Message Complexity (ISSNIP Dataset)

Sensor Node	# Msgs (Proposed Model)	# of Msgs (Centralized)	Ratio	Overhead Reduction %
S1 Link #1	106	400	0.2650	73.50%
S3 Link #2	66	400	0.1650	83.50%
S2 Link #3	29	400	0.0725	92.75%
S4 Link #4	36	400	0.0900	91.00%
S1/S3 Link #5	135	800	0.1687	83.13%
S2/S4 Link #6	102	800	0.1275	87.25%

Table 3.7: Message Complexity (INTEL Dataset)

Sensor Node	# Msgs (Proposed Model)	# of Msgs (Centralized)	Ratio	Overhead Reduction %
S1 Link #1	76	440	0.1727	82.73%
S3 Link #2	69	440	0.1568	84.32%
S2 Link #3	30	440	0.0681	93.19%
S4 Link #4	27	440	0.0613	93.87%
S1/S3 Link #5	106	880	0.1204	87.96%
S2/S4 Link #6	96	880	0.1090	89.10%

Considering the experimental results on Table 3.6 (ISSNIP data set), the least overhead reduction is 73.5% and the highest 92.75% leading the average reduction in communication overhead to be 85.18% on the first network tier. Considering the second tier level these savings cumulate to an average value of 85.19%. Similarly, the results on Table 3.7 for the INTEL data set shows an average overhead reduction of 88.52% on the first tier and a cumulative 88.53% for the second tier over the considered time window. Therefore, this level of communication reduction of more than 85% in average will lead to considerable savings on energy usage in maximizing active life-time and avoid any disruptions for process monitoring.

3.5 Conclusion

In this chapter an anomaly detection framework is presented that offers multi-granularity analysis for wireless sensor networks. Data is partitioned on to different cohesive regions dynamically based on *cumulative point-wise entropy* and *average relative density* that is computed over two distance-neighborhood levels. This overcomes the challenge of other existing solutions in effectively dealing with the dynamic multi-density nature of the considered observation domain. Consequent to the data partitioning, anomalous behaviour is identified as both individual outliers and groups of correlated observations through defined $E - regions$. Thresholds for differentiating between normal and anomalous are derived adaptively based on second order statistical knowledge available at each analytical stage. This is implemented over an innovative data processing framework that makes use of in-network processing on different network tiers enabling the detection of anomalies at multiple granularity levels on a hierarchical node

topology.

Experimental Results demonstrate higher detection accuracy for both normal and anomalous data with corresponding low false positives and false negatives. The added reductions on communication overheads highlight the scalability in extending the model efficiently over a large scale sensor field. A factor that is not discussed in this chapter is the establishment of the hierarchy in implementing the framework. While the proposed model is based on a static topology that is assumed a priory, this is a limitation considering dynamic topology changes that are possible within the domain of WSNs. Future research will address this issue in dynamically identifying parent nodes where granular analysis is performed over different levels through availability and node density aspects.

Chapter 4

Point-of-View Entropy Evaluations for Real-time Decision Support in Evolving Data Streams (POV-EE)

This chapter focuses on the concerns raised by the third research question in Chapter 1. Specifically, the need to have an anomaly detection procedure that supports decision making in near real-time. The two models that were developed in Chapters 2 and 3 were using a *batch-wise* approach for analysing the data. Therefore, the streaming data that is encountered on each node is aggregated over a small period of time before the analysis is performed on that data set as a group. However, given the nature of most of the industrial applications that utilize wireless sensor networks for pervasive monitoring it is desirable to have an *incremental* approach that is capable of facilitating the determining of anomalies in near real-time. Such an approach is able to determine any anomalous data as they are sensed on each node in the wireless sensor network.

The work in this chapter introduces an anomaly detection model that determines anomalies in an online fashion through an incremental approach. Inspired by the work in chapter 3 an entropy criterion is used to differentiate the anomalies from normal behaviour. Three unique reference points are adapted as part of a Point-of-View (PoV) approach that offer different

lenses to evaluate the relative change in entropy that occurs as a data stream evolve. Therefore, point-wise entropy is calculated relative to these reference points and analysis performed using a *Mean Relative Entropy (MRE)* measure that is defined for each sensed data point. It is shown that this measure is significantly capable of identifying any aberrant form of behaviour in the data when computed respective to the three reference points that are introduced.

In order to identify instances where the data stream may suddenly shift its spatial distribution as part of normal behaviour a secondary analysis is carried out employing a secondary data buffer that stores potential anomalies identified in the primary phase of detection. This dual buffer model with a second analytic phase for potential anomalies works to significantly reduce false alarms that are common with an otherwise unsupervised process. The viability of the proposed model is investigated with respect to accurately detecting different forms of aberrant behaviour as well as evolving patterns of normal behaviour with evident flows of sensor drift.

The rest of the chapter is structured as follows. Section 4.1 gives the background for WSNs in the context of real-time applications of industrial monitoring. The advantages of an incremental approach are presented in contrast to a batch-wise approach in this section before the specific contributions of the proposed model are detailed. Section 4.2 discusses some of the relevant related work in detail. In Section 4.3 the proposed incremental model employing an entropy criterion is presented in detail. This consists of subsections that focus on the defined entropy criterion, the point-of-view approach at computing entropy and the anomaly detection phase. Evaluation of the proposed models regarding accuracy is then performed employing a three pronged approach. This is performed for a variety of data distributions that represent different facets of dynamic and evolving data streams in Section 4.4. The first phase investigates the accuracy in classifying normal behaviour as the data stream evolves with evident spatial drifts. The second phase evaluates the model for detection accuracy with regard to different forms of aberrant behaviour representing anomalies. Finally, a comparative analysis is performed with regard to some of the most qualitatively relevant work in the area. Section 4.5 concludes the chapter.

4.1 Motivation and Contributions

Wireless Sensor Networks (WSNs) consisting of low cost sensor nodes that collaborate to collect and deliver data are increasingly being established as the main form of large scale pervasive monitoring [Akyildiz et al., 2002] [Puccinelli and Haenggi, 2005]. They are being adapted for a variety of industrial scale applications from infrastructure (oil and gas pipelines, power plants, transmission lines) and equipment (turbines, material processing machines) monitoring to meteorological and land use monitoring in agriculture [Buttyan et al., 2010], [Bertocco et al., 2008], [Luo et al., 2012], [Guevara et al., 2012]. Therefore, WSNs are enabling automation with distributed and granular control in such applications through the timely gathering and presentation of information for decision making.

As such, the integrity of the gathered data is of utmost importance. The data should be reliable and accurate in addition to being efficiently sensed and communicated. However, the nature of the sensor networks themselves as well as naive modes of implementation often used introduces different threats and vulnerabilities that directly impact the integrity of the sensed data [Djenouri et al., 2005], [Shi and Perrig, 2004]. In particular, the resource limitations of sensor nodes has resulted in less focus on security as opposed to efficiency making the sensor networks vulnerable to different security attacks. Also in most large scale implementations it is impractical to ensure the physical security of individual nodes and that opens up possibilities of tampering and side channel attacks where the nodes are compromised [Bar-El et al., 2006], [Ravi et al., 2004]. Furthermore, there is always the possibility of node failure through software corruption and hardware malfunction [Rajasegarar et al., 2008]. All of these situations impact the integrity of the sensed data streams through incorrect readings or bogus data [Sun et al., 2007]. Such instances will severely impact the decision making systems at the heart of these applications. Therefore, it is of critical importance to enable a mechanism that effectively detects the anomalous instances and differentiate the normal data on sensor network data streams in mitigating the negative impact they inflict on integrated systems and applications.

However, such an anomaly detection mechanism needs to critically consider and cater to the unique nature of the data that is encountered in WSN applications. Figure 4.1, presents the contextual data environment of WSN applications. These can be noted as (i) continuous

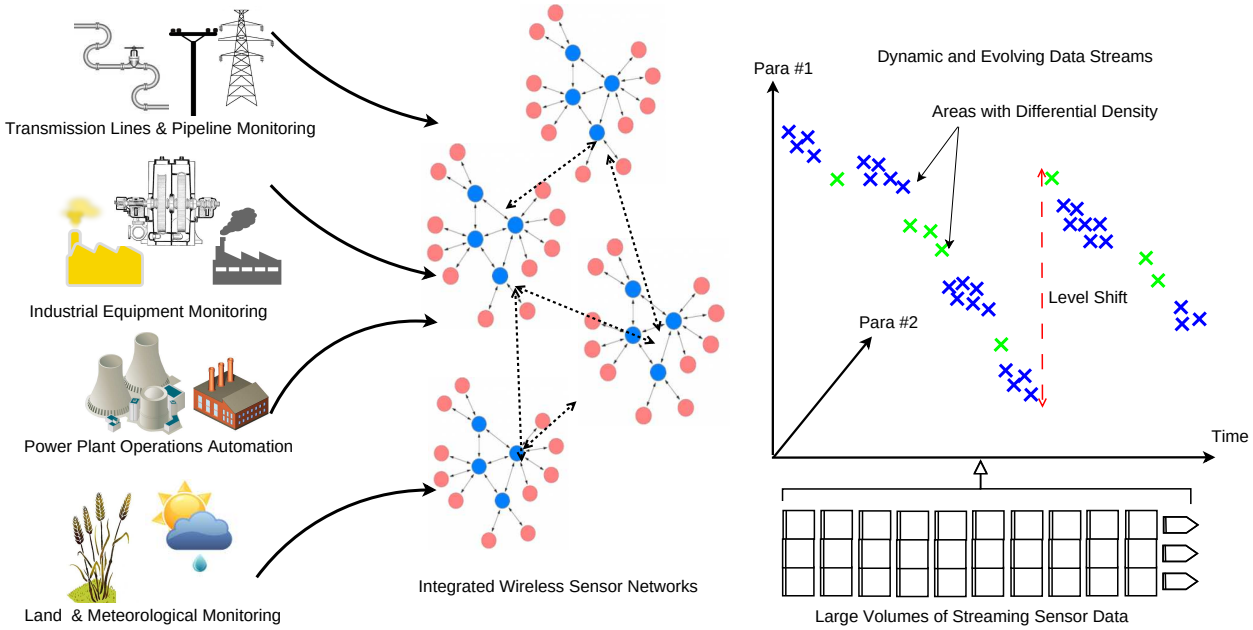

Figure 4.1: Wireless Sensor Networks (WSNs) Application Environment: WSNs are used for distributed sensing and communication regarding a variety of industrial-scale applications. From infrastructure/equipment monitoring to environmental and meteorological monitoring they form a vital information core that enables automation and smart functionality. The different applications with regular time critical sensing paves the way for *very large amounts* of multi dimensional *streaming* data of which ensuring integrity is a core concern in achieving reliable and optimum performance. Such data demographics are *dynamic* and *non-homogeneous* with differential density presenting a challenge for traditional anomaly detection measures. The Sensor networks are often dynamically ordered on to a hierarchical topology offering sensing granularity at different levels with each node independently generating a stream of dynamically evolving data that often enable real time decision support.

streaming of large data volumes, (ii) dynamic fluctuations on data streams with evolving patterns, (iii) data heterogeneity over different streams in unmatched distributions and critically (iv) near real time impact on key decision making through streaming functionality. Considering these facts this research aims to provide a contextual anomaly detection model that successfully addresses the following challenges in contrast to the existing literature [Zhang et al., 2010],[Xie et al., 2011] in this area.

- Dynamic and evolving data streams mandating an unsupervised and non-probabilistic model for differentiating between normal and abnormal.
- Detect anomalies in an online fashion in facilitating real time decision support.

- Efficient in-network procedures with reduced complexity considering resource limitations.

Table 4.1: Deficiencies in Batch-wise approaches in relation to an Incremental approach for anomaly detection over streaming data

Features & Functionality	Batch-wise Approaches [Rajasegarar et al., 2014],[Kumarage et al., 2014],[Kumarage et al., 2013]	Incremental Approaches [Burbeck and Nadjm-Tehrani, 2007],[Wang et al., 2008],[Hill and Minsker, 2010],[Subramaniam et al., 2006]
Managing Streaming Data	Breaks up the data on to stored groups for processing as individual batches Infeasible to store large volumes of data on a single batch	Data processing done in an incremental manner without any selective batch processing Not required to store such amounts of data in an incremental mode of analysis
Dynamicity & Adaptability	Cannot identify evolving data patterns within a batch Cannot determine anomalies as they occur as data is processed as a batch Cannot adapt to sensor drift or level shifts	Capable of identifying data patterns in a fully dynamic manner through incremental analysis Evaluates and detects anomalies as they are sensed in an online fashion Dynamically adapts through incremental analysis over the streaming data
Detection Accuracy	Accuracy depends on the chosen period of time for batch processing of data Accuracy depends on the quality of data within the selected batch of data	Time period has no intrinsic effect on accuracy as detection is performed in a near real-time manner as opposed to predetermined time intervals Accuracy not dependent on selected data groups as data is handled as an incremental series as opposed to batch-wise in determining the normality model

The most common way to detect anomalies in such data streams is to model the normal data and identify any deviations from the resulting normal profile. Such models can be divided on to two categories based on the time when the normal profile is built and the analysis carried out. That is as *incremental models* and *batch-wise models*. A model is termed batch-wise when

the data (sensor measurements in the current context) is stored over a certain pre-determined time period before they are analysed as a group. The normal profile is then built upon this stored data space in order to identify any deviations at the end of each such time period. On the other hand an incremental model doesn't store any amount of grouped data in order to build the normal profile or perform analysis at specific time intervals on the stored groups. In this approach the model is being adaptively updated as the data arrive and any anomalies are determined in a near- real time manner. Therefore, it is advantageous to have an incremental approach rather than a batch-wise approach for anomaly detection within the current context. We compare the main deficiencies of batch-wise models in relation to the clear advantages of an incremental approach for a streaming data context in Table 4.1.

4.1.1 Contributions

The above mentioned challenges are uniquely addressed in this research through the effective use of an entropy criterion to differentiate anomalies. An incremental model is proposed to evaluate relative entropy change and identify anomalies in an online fashion and negate the deficiencies of existing batch models. This is also implemented in-network with data processing performed locally and in parallel at each node in supporting the real-time nature of typical industrial WSN monitoring applications.

Specifically, in the proposed incremental model, a small buffer is maintained at each sensor node in relation to the incoming data streams. The size of the buffer is to be determined according to the memory constraints specific to the considered WSN. This buffer is kept current through continuous updates where each incoming data point is added while the oldest is removed. The relative change in point-wise entropy over the buffered data space is then evaluated with each incoming data point. This is performed through the introduction of three unique Point-of-View (PoV) approaches taken to be as (i) PoV of the mean μ, (ii) PoV of the median η and (iii) PoV of the historic mean μ' (without the influence of latest data point). The relative change in entropy is then evaluated using a defined *mean relative entropy* measure that is used to differentiate anomalies.

A key concern for an incremental approach is to detect level shifts and sensor drifts that

may occur as part of normal behaviour in the data stream. In such instances, the observed data distribution may suddenly shift to a new spatial range before continuing normal behaviour (in that range). For this purpose a secondary buffer is introduced accompanied by a second analytic stage for any potential anomalies. Therefore, any element that is designated as potentially anomalous through relative entropy evaluation in the primary analytic phase will be temporarily stored on the secondary buffer that accumulates the next potential anomalies if detected continuously. The relative change in entropy is then evaluated within the context of the new measurements through the use of the trio of PoV evaluations. This significantly identifies any sudden change in measurement range that can be normal in a particular application domain such as voltage monitoring where the observed voltage may jump between different ranges as part of normal behaviour. Therefore, in the proposed approach such occurrences will not be misclassified as anomalies in reducing the false alarm rate that may otherwise have been common with regard to the unsupervised approach adapted. The model is graphically presented in Figure 4.2.

The proposed models are evaluated using several different data distributions that represent dynamically evolving data streams. These distributions are derived based on two major publicly available data sets from Intel Berkeley Research Laboratories [Bodik et al., 2004] and the Australian Research Council's Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) group [Suthaharan et al., 2010]. Sensitivity and specificity analysis is used to investigate the detection accuracy over different data distributions that represent unique facets of evolving behaviour. The major technical contributions of this research can be summarised as follows.

- *Dynamic evaluation of anomalies in an online fashion for evolving data streams.* Dynamic and evolving patterns of observation behaviour is accommodated for in a two-step model for detecting outliers and abnormal behaviour in near real-time. Sensor drifts and level shifts where sudden spatial changes in the data stream may occur as part of normal behaviour are catered for through the use of a dual buffer model here potential anomalies undergo secondary analysis. Experimental investigations reveal high detection accuracies of more than 98% in average sensitivity for different forms of abnormal behaviour

while still being robust with reduced misclassifications where the data streams evolve with evident sensor drift.

- *PoV approach in evaluating relative change in entropy for dynamic detection of abnormal behaviour.* Abnormalities and outliers present in the data stream is captured through the evaluation of relative change observed in entropy that is computed in a point-wise manner. The point-wise entropy of a data point is therefore, compared to the mean entropy value that is observed over the buffered data space at each node. Three PoV approaches are introduced based on three reference points over which point-wise entropy is computed. Each PoV act as lens providing different insight for more accurate detection of anomalies. The three reference points proposed are the perspective of the mean μ, the median η and the historic mean μ' in relation to the maintained data buffer. Improved accuracy in sensitivity (more than 98% in average) and specificity (more than 96% in average) is achieved with regard to the three PoV's as revealed by the extensive experiments.

These innovations are combined together in a framework that is implemented through efficient in-network procedures with reduced complexity and is uniquely suitable for industrial WSN environments to facilitate anomaly detection in an online fashion.

4.2 Related Work

There exists a wide range of anomaly detection methods [Chandola et al., b], [Chandola et al., 2012], [Patcha and Park, 2007] that have been proposed specific to different application areas. [Zhang et al., 2010] and [Xie et al., 2011] provide a good overview of the techniques specific for streaming sensor network environments. Considering existing batch-wise approaches, [Rajasegarar et al., 2014] and [Kumarage et al., 2014] stands out as they employ data clustering based models that can be adaptive and implemented in a non-supervised manner without prior training. In [Rajasegarar et al., 2014], a fixed-width clustering algorithm is employed to cluster the data on each sensor node with anomaly detection performed using the average inter-cluster distance over a nearest neighborhood. A distributed data processing model is adapted to identify local and global anomalies together with a cluster merging process. However, determining

the width of the clustering algorithm remains a major limitation in addition to the inherent weaknesses of a batch-wise model as discussed in the previous section. [Kumarage et al., 2014] introduces an entropy based data clustering approach that overcomes most of the limitations in [Rajasegarar et al., 2014]. It is more adaptive and responsive to dynamic data fluctuations while being mostly non-parametric in implementation. However. it still inherits the major limitations of a batch-wise approach and the inability to determine anomalies in a near real-time manner. The concept of using point-wise entropy for anomaly detection is adapted towards an incremental model in the current work as inspired from the work in [Kumarage et al., 2014].

Focusing on incremental models of anomaly detection on streaming data, [Wang et al., 2008] provides a multi-pronged detection model based on reducing feature extracted data vectors using PCA. The resulting distance between the vector and its subspace reconstruction is used for classifying and detection. However, it is mainly limited as prior training is required for new data streams with different detection models. Anomaly detection employing a fast incremental clustering model is introduced with a new grid index that improves efficiency in [Burbeck and Nadjm-Tehrani, 2007]. While it has advantages in dynamic adaptability with limited support for evolving data streams through an integrated method for extension and forgetting of out-dated elements, it still employs a semi-supervised model with incremental training data required and does not fully support real-time detection. A scalable auto-regressive data driven model with prediction is introduced in [Hill and Minsker, 2010]. While these approaches mostly parallel the work in change detection on time series data the use of threshold based techniques coupled with prediction intervals can be unreliable over mostly dynamic data streams. Other variants of auto regression in mostly linear time series data such as [Subramaniam et al., 2006] and [Gustafsson, 1996] involves high computational and memory complexity and can be unfit for near real-time decision support in sensor network data streams.

4.3 PoV Approach for Incremental Anlaysis

Considering the limitations in the above work and inspired by employing an entropy criterion to detect any aberrant deviations of data, a fully dynamic and adaptive model for anomaly detection is built for real-time decision support as detailed next.

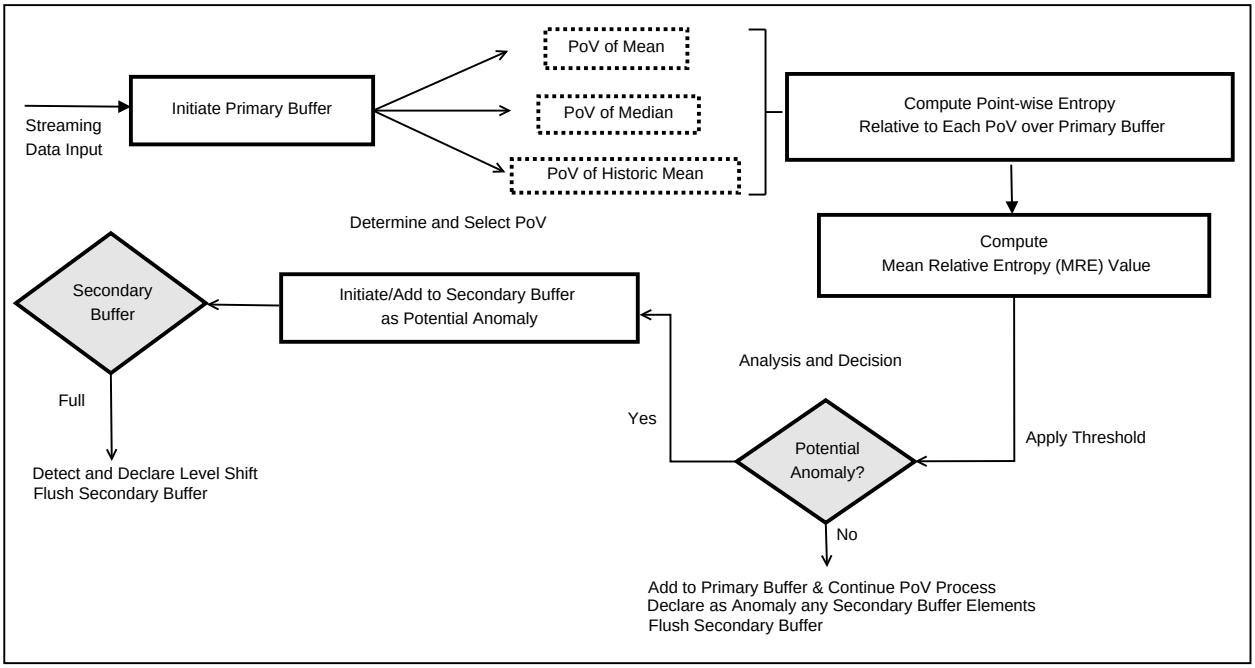

Figure 4.2: Proposed Model for Incremental Anomaly Detection: The main data processing and analytical steps of the proposed dual buffer model for PoV entropy evaluation for streaming WSN data is presented with regard to achieving near real-time anomaly detection

In this section we construct the three PoV analysis models for contextual anomaly detection in wireless sensor networks. The architecture of the proposed model is depicted in summarized form in Figure 4.2. Given that sensor networks naturally produce a data stream the model can be formally introduced as follows. Let $X = \{X_1, X_2, ., X_i, ., X_n\}$ be a finite set of n points where each X_i is a multi-dimensional vector $X_i = [x_{i1}, x_{i2}, x_{i3}, .., x_{id}]$ corresponding to a measurment made by an individual sensor node. As per the streaming model X can only be read in increasing order of the indices i and each node can only have information over a small subset of data it has seen so far as opposed to having random access to all data within the complete set. The subset of data that each node can retain and perform analysis according to memory constraints is taken as the *Primary Buffer Space*: B_{P_j} at each node j. The different nodes in the network are considered to be time synchronized and deployed in either a homogeneous or heterogeneous observation environment where dynamic behaviour is assumed. Therefore, observations will constitute of same or different distributions that evolve over time and are previously unkown and not pre-determined.

4.3.1 The Entropy Criteria

The core functionality in the proposed approach is derived from the use of an entropy criteria in differentiating anomalies from the data stream. This is performed in an online fashion over the buffered data at each node in identifying data anomalies on the fly with the advent of each new observation. The major advantages of having an entropy method to filter out abnormal behaviour from normal behaviour can be listed as follows; (i) Data characteristics that are above second order statistics are represented, (ii) Non-parametric and non-probabilistic computation when combined with a similarity measure using distance, (iii) Evolves with and adaptive to dynamic distribution changes and (iv) robust in capturing a variety of data anomalies from different causes due to the emphasis on randomness or surprise compared to that of the ordered flow expected in normal behaviour. As we are dealing with a data stream consisting of indvidual multi-dimensional observations we use the following point-wise definition that is derived from shannon's entropy.

Definition 5. ***Point-wise Entropy for Multi-dimensional Data***

Considering a set of observations $X = \{X_1, X_2, ., X_i, ., X_n\}$*, with each instance* X_i *being a d-dimensional vector, where* $X_i = [x_{i1}, x_{i2}, x_{i3}, ., x_{id}]$ *the point-wise entropy* E_i *of a particular data point* X_i *in relation to a certain data point* X_j *is taken as,*

$$E_{ij} = -S_{ij} \log S_{ij} - (1 - S_{ij}) \log(1 - S_{ij})$$

,

- *where* S_{ij} *is the adapted similarity measure between the two data vectors* X_i*,* X_j *and is given by,*

  $S_{ij} = \exp^{-\alpha P_{X_{ij}}}, \forall X_i, X_j \in X,$

- *here, $P_{X_{ij}}$ represents the proximity measure between the two data vectors X_i and X_j .*

According to the above definition there are two major factors that need to be determined in order to compute a point-wise entropy value for a certain data point in a stream; (i) Proximity measure $P_{X_{ij}}$ and (ii) parameter α. We use the simple Euclidean Distance $D_{X_{ij}}$ for determining proximity between two data points both for its effectiveness and simplicity in computation. α is then determined as the value resulting from the substitution of the mean distance D_{m_X} as the proximity measure for a similarity value of 0.5 considering a normalized input-range in the interval [0.0-1.0]. This will be non-partial and neutral with regard to the subsequent computations of an entropy value and adaptive to each dynamic domain as D_xm is updated with the evolving data stream. Therefore, we have;

$$D_{X_{ij}} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + ...(x_{id} - x_{jd})^2}$$

$\alpha = -\ln 0.5/\bar{D}$ where $\bar{D} = D_{m_X}$

Based on the above definition for point-wise entropy the following observations are derived.

Observation 1: The value E_{ij} between any two data points will be in the closed interval $[0, 1]$.

$\therefore [0, 1] = \{E_{ij} \in \mathbb{R} | 0 \leq E_{ij} \geq 1\}$

Observation 2: E_{ij} will tend towards zero when the similarity (S_{ij}) is either tending towards one or zero. Therefore, while an average similarity between two data vectors will result in an entropy value that tends towards one the higher the *similarity* or *dissimilarity* between them the entropy value will tend towards 0.

$\therefore$ **when $S_{ij} \longrightarrow$ 1 or 0,**

$E_{ij} \longrightarrow$ **0**

4.3.2 The PoV Approach

The derived point-wise entropy measure is then adapted for analysis over a streaming data model through an innovative Point-of-View (PoV) approach. Considering the Primary Buffer Space B_{P_j} at each node j, where the latest updated set of observations $X_{B_{P_j}}$ is retained, the point-wise entropy is calculated relative to three unique vantage points. These points as selected from the insights from observations 1 and 2 act as lenses offering a clear view on how entropy evolves on a dynamic data stream while each providing an independent perspective relative to its spatial positioning over the distribution of the buffered data space. Accordingly we define three such points as the PoVs of interest. Each PoV is proven to be adequate and capable in detecting anomalies accurately in the next section

Definition 6. ***Point-of-View (PoV)***

Considering the set of observations $X_{B_P} = \{X_1, X_2...X_i...X_n\}$, that is retained on the primary buffer B_{P_j} at a given sensor node j where each instance X_i is a d-dimensional observation ($X_i = [x_{i1}, x_{i2}, x_{i3}...x_{id}]$) the point relative to which the point-wise entropy E_{ij} is calculated over all data points in X_{B_P} is defined as a PoV point of interest. Three such points offering independent perspectives on entropy are defined to be as follows. Note that all operations are point-wise operations.

- *PoV 1: PoV of the Mean - μ*

  This is the point-wise mean *for the complete set of data points X_{B_P} retained in the primary data buffer on each sensor node at a given instance. We have;*

  $$\sum_{i=1}^{n} X_i / \left| X_{B_P} \right|$$

- *PoV 2: PoV of the Historic-Mean - μ'*

  This is the point-wise mean *for the considered set of data points that is in X_{B_P} before the advent of the last element n at a given instance. We have;*

  $$\sum_{i=1}^{n-1} X_i / |X_{B_P}| - 1$$

- *PoV 3: PoV of the Median - η*

  This is the value that seperates the higher half of the data space in the primary data buffer X_{B_P}, from the lower half with respect to numerical value of each observation attribute on each sensor node. We have;

$$X_\eta = [x_{\eta 1}, x_{\eta 2}, ., x_{\eta k}, ., x_{\eta d}]$$

  where $x_{\eta k}$ is the middle value in the ordered set of k^{th} attributes $\forall X_i \in X_{B_P}$

4.3.3 Anomaly Evaluation

Evaluation for anomalies in the data stream is performed in an online fashion with the advent of each new data point. As a new data point gets added to the primary buffer (B_P), each sensor node computes the point-wise entropy $[E_{ij}]$ values over the buffered data relative to the three PoV's of interest. Then for each new incoming data point a *Mean Relative Entropy (MRE)* value is computed according to the following definition.

Definition 7. ***Mean Relative Entropy (MRE)***

Considering a computed set of point-wise entropy values $E = \{E_1, E_2 ... E_i ... E_n\}$ over a buffered data space $X_{B_{P_j}}$, where each instance E_i is calculated relative to a given PoV according to Definition 1 and Definition 2, the Mean Relative Entropy is calculated in relation to any data point in the data space as given by,

$$MRE_i = E_i / [1/n \sum_{i=1}^{n} E_i]$$

- *where E_i is the point-wise entropy between a data vector in buffered data space X and the selected PoV as given in* Definition 1 *and* Definition 2 $\forall X_i \in X$

The above defined MRE value is then used as a measure that signifies the *relative change* in point-wise entropy E_i for the incoming data relative to that of the average entropy encountered over the buffered data from the perspective of a selected PoV. Therefore, if the MRE is less

than a given threshold T the latest data point that contributed to this change is marked as a potential anomaly. The following lemma highlights the basis for the MRE usage in the current context.

Lemma 1: Given an anomalous data point that is significantly aberrant from the normal data, the corresponding MRE value as computed from Definition 3 will significantly deviate towards zero as opposed to a value that tends towards one for the normal data.

Proof: According to *observations 1 and 2*, the E_i value for an aberrant data point will tend towards zero given its dissimilarity to the considered PoV being either the mean, historic mean or the median as defined in *Definition 2*. On the other hand the E_i values for the buffered normal data space will tend towards one. Therefore the mean entropy value over the buffered data space will retain a value that is significantly closer towards one. This results in an MRE value that will increasingly deviate towards zero given that the numerator tends towards zero while the denominator tends towards one as per the definition for MRE in *Definition 3*.

Values of 0.25 and 0.5 for T are investigated in the next section in concluding that a value of 0.5 is significantly successful for correct classification. Therefore, for any computed cumulative point-wise entropy value Ec_i we have;

$$if : E_i/[1/n \sum_{i=1}^{n} E_i] < T \longrightarrow PotentialAnomaly$$

4.3.3.1 Secondary Evaluation

Once a potential anomaly is encountered another secondary buffer space B_S is initiated with that data point being it's first element. Then, the next incoming data point is allocated in place of the last element in the primary buffer and the PoV evaluation on entropy variation performed. If that point is also found out to be potentially anomalous in the context of the primary buffer it is added to the secondary buffer as the next element. This process is continued until the secondary buffer space is filled up when all incoming data are continuously found to

be anomalous with respect to the data distribution retained in the primary buffer X_{B_P}. Once the secondary buffer is filled the PoV operations will be performed on that buffer in evaluating for anomalies within the new context of X_{B_S}. However, if the incoming data points are not anomalous with regard to the data now retained in the secondary buffer while still being anomalous relative to the primary buffer, operator mediation is required. This will signify that the streaming data has suddenly shifted over to a different spatial region (value range) before continuing normal behaviour within that range. This could be part of normal behaviour expected in the considered application. A typical example is voltage monitoring where the observed voltage level may suddenly shift to a different range. Therefore, the proposed model allows for this contingency in identifying such sudden shifts in observed behaviour. This significantly lowers the number of flase positives that would otherwise have inundated the system.

4.4 Evaluation and Results

In this section the proposed PoV entropy evaluation models for real-time anomaly detection in sensor network data streams is evaluated using a three pronged approach. First, the performance of the model is investigated with regard to accuracy in classifying normal behaviour over dynamic and evolving data streams. Adaptability over instances where sensor drift and level shifts are evident is studied. Next, the model is investigated for accuracy in it's primary goal of identifying anomalies that are present in the data stream while the underlying normal data distribution continues to evolve dynamically. Investigations are performed with regard to anomalies that are aberrant in both an orderly and chaotic manner in relation to the normal data distribution. Finally, a detailed comparison is performed in contrasting the strengths of the proposed model as opposed to some of the qualitatively significant related work.

4.4.1 Data Sets

Several data sets are made use of that represent a variety of possible sensed data distributions in the extensive experimentation performed in evaluating the proposed models. These are derived

from the publicly available wireless sensor network data from Intel Research Laboratories at Berkeley [Bodik et al., 2004] and the research network on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) of the Australian Research Council [Suthaharan et al., 2010]. Each individual data set is taken as corresponding to the data stream encountered at a single sensor node and represents different facets of dynamic and evolving behaviour on each node's observation domain. Four separate data distributions (N1 - N4) are derived from the available data corresponding to the multi-hop readings of four sensor nodes (Mote id1 - Mote id4) from ISSNIP. This data is from a multi-hop WSN deployment of TelosB sensor motes that monitor temperature and humidity variations over a period of time in five second intervals. The measurements at nodes one (N1) and three (N3) represent dynamic and evolving patterns (with evident sensor drift and level shifts) as influenced by the artificial manipulation of temperature and humidity in the observed environment. In contrast, the data from nodes two (N2) and four (N4) have less variation while being more concentrated as the fluctuation of the temperature and humidity are as expected in a normal environment. Therefore, we derive four data distributions consisting of four hundred (400) consecutive measurements of temperature and humidity readings each, with N1 and N3 consisting of the portion of data (100 measurements) that evolves continuously in the temporal domain over a dispersed (spread out) spatial range while N2 and N4 constitutes of four hundred (400) measurements of dynamic behaviour that is concentrated (focused) over a smaller spatial range. The four derived data distributions are presented in Figure 4.3.

From the Intel lab sensor data we derive another four data distributions consisting of both a tightly correlated set of measurements focused over a small spatial range as well as a segment that evolves dynamically over an extended spatial range. Again we draw four hundred readings each which are in temporal sequence corresponding to four sensor motes. The data includes real measurements of four parameters (temperature, humidity, voltage and light) together with time-stamped topology information from Mica2Dot sensors with weatherboards. The four data distributions as derived (N5 - N6) are depicted with respect to the temperature and humidity parameters in Figure 4.4. The complete sets of data in the eight derived distributions are labelled as normal for the subsequent experimental evaluations.

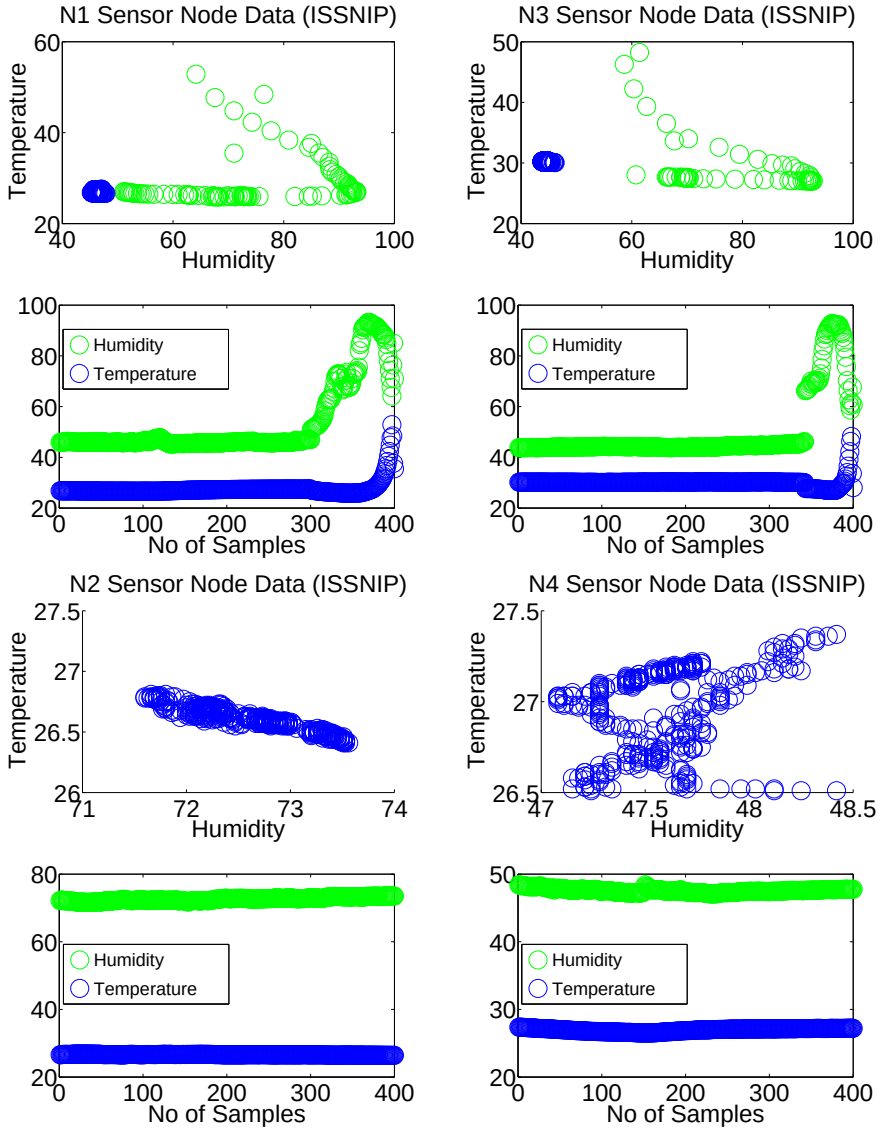

Figure 4.3: Experimental Data Distributions (Normal Data - ISSNIP): The first two data distributions derived from the ISSNIP data (N1 and N3) represents a tightly correlated core set of data that begins a sequential drift in an evolving pattern over a spatially extended range. The next two data distributions derived from ISSNIP (N2 and N4) consists of a more tightly correlated set of data with only limited spatial variation or temporal drift. While both N1 and N3 contain clear patterns of sensor drift N3 also displays a level gap in its readings. Each data distribution is represented from a spatial perspective relative to the measured parameters as well as from a temporal view of each parameter independently. The temporal view clearly shows the sequential and evolving pattern in contrast with the overall spatial drift while the main representation shows the overall distribution of data. The temporal view is presented in the second row underneath the main view for each data set.

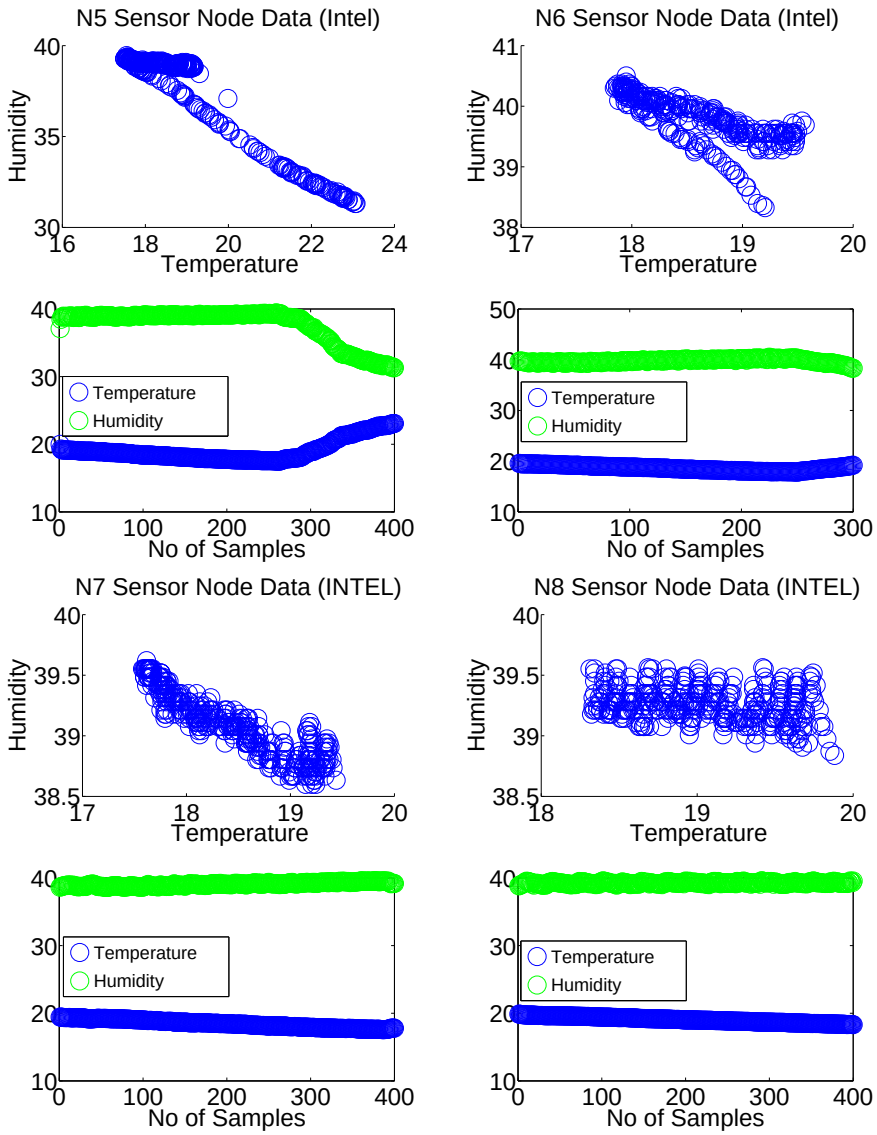

Figure 4.4: Experimental Data Distributions (Normal Data - INTEL): The first two data distributions derived from the (INTEL) data (N5 and N6) represents a tightly correlated set of data that then drifts on a continuous and sequential pattern. This drift is clearly evident in the latter portion of data in distribution N5 as well as to a lesser extent in N6. The next two data distributions derived from (INTEL) data (N7 and N8) feature two sets of data with tight correlation and almost no spatial drift temporally. The temporal view clearly highlighting the evolving patterns and drift is represented below the main view with respect to the measured parameters for each distribution.

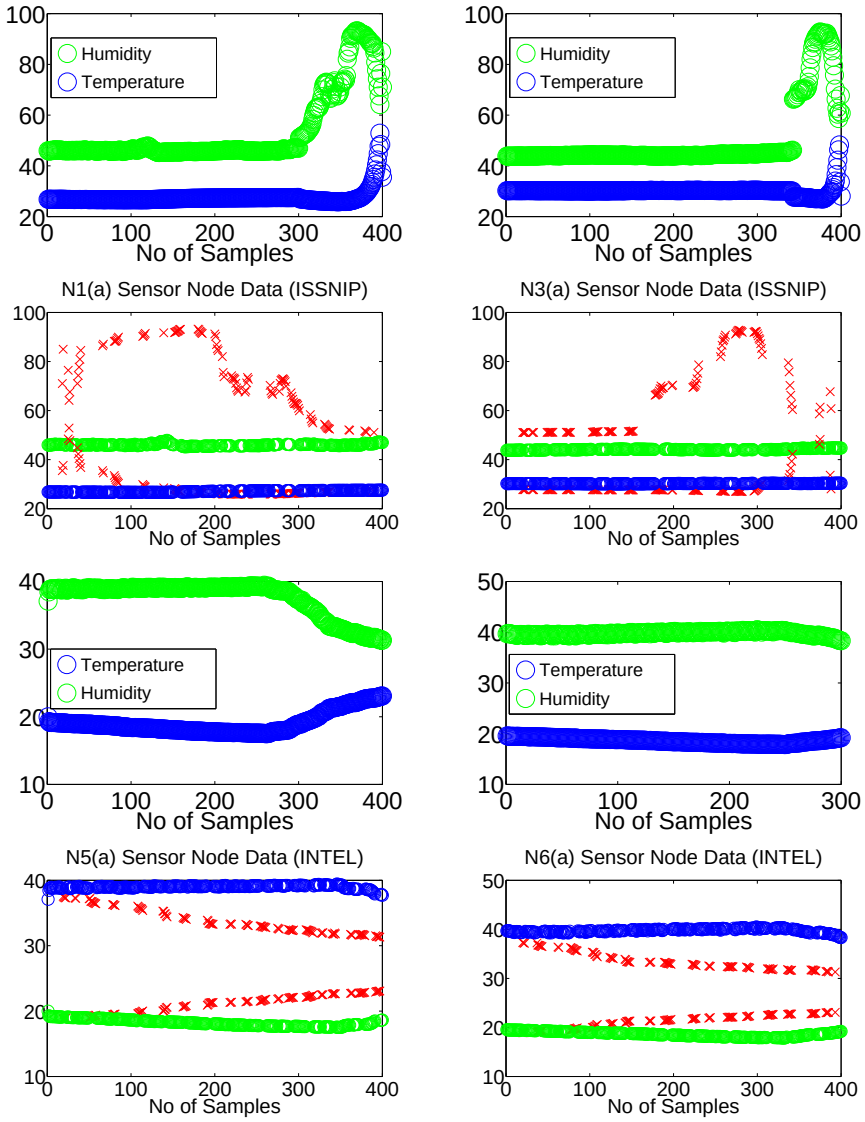

Figure 4.5: Experimental Data Distributions with Anomalies): Four data distributions are derived from the previously composed data sets of N1,N3 (from ISSNIP) and N5,N6 (from INTEL) as containing anomalous data in addition to the majority normal data. The latter portion of evolving data with significant spatial drift are separated and re-integrated with the majority normal data portion in a random manner with respect to both the number and positioning. This adds these latter values as outlying anomalies to the tightly cohesive initial portion of data. Each derived data distribution is shown with respect to the measured parameters of temperature and humidity in a temporal view as two before and after plots. The top figure for each data set gives the initial view with the anomalies-to-be data making up the tail-end of drifting normal data while the bottom figure shows these values randomly inserted over the more cohesive normal data in making up the final distribution for evaluation. In the resulting distributions N5(a) and N6(a) presents an orderly form of aberrant data while N1(a) and N3(a) represent disorderly and chaotic forms of aberrant data as anomalies.

Next, four separate data distributions are derived as to contain anomalous data in addition to the majority of normal data that is encountered. We make use of the previously composed node data of N1 and N3 from ISSNIP as well as N5 and N6 from INTEL. Each of these data sets consist of a tail portion of 100 readings from a total of 400 that spatially deviates significantly in a sequential manner. Therefore, in the purpose of introducing anomalies for evaluation, we separate this portion of data and then re-integrate them with the tightly correlated and spatially cohesive data readings making up the initial portion on each data set. This is performed in a random manner in the sense that a different number of randomly chosen outer readings are inserted at different locations that are also chosen randomly within the initial cohesive normal data portion. These inserted values (100 readings) are then labelled as anomalous with the remainder (300 readings) labelled as normal for use in the evaluation of proposed models. The resulting data distributions produce both orderly and chaotic forms of aberrant data as relative to the majority normal data. The derived data distributions [N1(a),N3(a),N5(a) and N6(a)] are shown in Figure 4.5

4.4.2 Phase #1: Investigations on Evolving Normal Data

In the first phase of the experimental evaluation the accuracy of classifying normal behaviour is investigated in instances where the sensed data distribution shows signs of spatial drift or evolves dynamically. We make use of the eight data sets (N1-N4 from ISSNIP and N5-N8 from INTEL) that were derived as mentioned previously and consisting of all normal data of different distributions with evolving patterns. These data sets are then subjected to the three PoV models on a simulated sensor node using the matlab environment. For each data set we calculate the *false positive rate (FPR)* and the *true negative rate (TNR) or Specificity* using the resulting number of *false positives (FP)* and *true negatives (TN)* from the three different PoV models. A false positive is when a normal reading is falsely classed as anomalous and a true negative is when a normal reading is correctly classed as normal. The ratio between the number of false positives and the actual normal readings gives the false positive rate while the specificity or true negative rate is the ratio between the number of true negatives and the normal readings. These results are computed with regard to two threshold levels of less than

50% and 25% on the *mean relative entropy* value and given in Tables 4.2 -4.4.

Table 4.2: Classification Accuracy: PoV of the Mean - μ

Sensor Node	N1	N2	N3	N4	N5	N6	N7	N8
FP (50%)	23	06	21	21	01	05	11	17
FPR % (50%)	0.0767	0.0150	0.0700	0.0700	0.0025	0.0167	0.0275	0.0425
Specificity % (50%)	0.9233	0.9850	0.9300	0.9300	0.9975	0.9833	0.9725	0.9575
FP (25%)	05	01	06	08	01	02	05	04
FPR % (25%)	0.0167	0.0025	0.0150	0.0200	0.0025	0.0067	0.0125	0.0100
Specificity % (25%)	0. 9833	0. 9975	0.9850	0.9800	0.9975	0. 9933	0. 9875	0.9900

Table 4.3: Classification Accuracy: PoV of the Median - η

Sensor Node	N1	N2	N3	N4	N5	N6	N7	N8
FP (50%)	21	02	42	11	58	06	15	04
FPR % (50%)	0.0700	0.0050	0.1050	0.0250	0.1450	0.0200	0.0375	0.0100
Specificity % (50%)	0.9300	0.9950	0.8950	0.9750	0.8550	0.9800	0.9625	0.9900
FP (25%)	04	01	05	07	00	01	00	02
FPR % (25%)	0.0133	0.0025	0.0167	0.0175	0.0000	0.0025	0.0000	0.0050
Specificity % (25%)	0.9867	0.9975	0.9833	0.9825	1.0000	0.9975	1.0000	0.9950

Table 4.4: Classification Accuracy: PoV of the Historic-Mean - μ'

Sensor Node	N1	N2	N3	N4	N5	N6	N7	N8
FP (50%)	04	06	05	01	03	01	07	04
FPR % (50%)	0.0133	0.0150	0.0167	0.0025	0.0075	0.0025	0.0175	0.0100
Specificity % (50%)	0.9867	0.9850	0.9833	0.9975	0.9925	0.9975	0.9825	0.9900
FP (25%)	00	01	01	00	01	00	00	00
FPR % (25%)	0.0000	0.0025	0.0025	0.0000	0.0025	0.0000	0.0000	0.0000
Specificity % (25%)	1.000	0.9975	0.9975	1.0000	0.9975	1.0000	1.0000	1.0000

The results give an average specificity in classifying normal behaviour as at 95.9% with a corresponding low average false positive rate of 3.7% for the lower threshold of less than 50% in MRE and a high 98.9% and 1.07% for the next threshold level (of less than 25% MRE) over the eight different distributions with respect to the PoV of the mean. Similarly, for the PoV of the median we have an average specificity of 94.7% and 99.2% with average false positive rates of 5.2% and 0.7% over the two threshold levels. The PoV of the historic mean also clearly depicts a higher accuracy level of 98.9% and 99.9% in average specificity with significantly lower

levels of 1.1% and 0.09% in average false positive rate. Therefore, it is demonstrated that the three proposed PoV based entropy evaluation models are clearly successful in classifying normal behaviour with a very low false alarm rate and is adaptive and resilient in instances of dynamic changes in the sensed data stream. The models maintain their high accuracy rate when clear drift patterns are present as in the data distributions of N1 and N3. Furthermore, the use of a secondary buffer in correctly identifying sudden shifts in the observed patter that are non-anomalous is also successful as evident from the results with respect to the data distributions in N3 and to a lesser degree in N1.

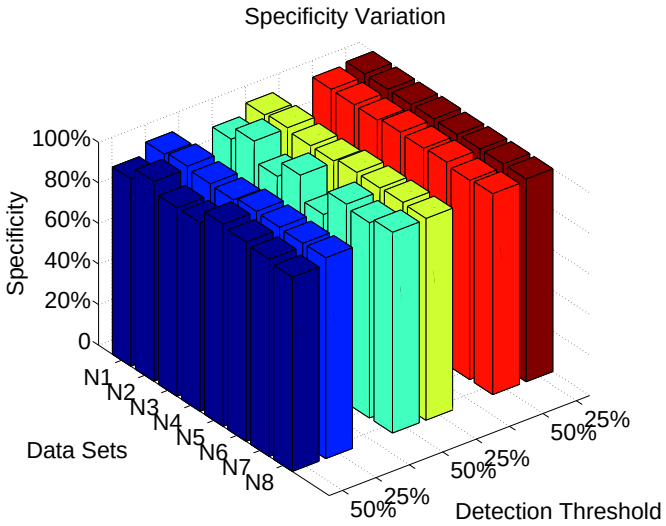

Figure 4.6: Specificity Variation: The variation in specificity over the different data distributions from N1-N8 is depicted relative to two threshold values of <50% and <25% over mean relative entropy (MRE). In all instances the specificity is higher at <25% compared to < 50% while the latter also demostrates higher acccuracy of >80% and more than 90% in a majority of cases.

The variation in specificity relative to the two threshold levels for MRE over the eight different data distributions is comparatively depicted in Figure 4.6.

4.4.3 Phase #2: Investigations on Data with Anomalies

In the second phase of the experimental analysis we use the four data distributions with anomalies [N1(a),N3(a),N5(a) and N6(a)] in a similar manner as in the previous step in calculating the specificity and the false positive rate. However, with the presence of anomalies in addition

to the afore-mentioned metrics we also calculate the *false negative rate (FNR)* and the *true positive rate or Sensitivity* using the resulting number of *false negatives (FN)* and *true positives (TP)* from the three entropy based PoV models. In this context, a false negative is when an anomalous reading is incorrectly classed as normal and a true positive is when an anomaly is correctly classed as anomalous. The ratio between the number of false negatives and the actual anomalous readings gives the false negative rate while the ration between the number of true positives and the anomalous readings give the sensitivity. The results obtained with respect to a threshold value of 50% in MRE value are presented in Tables 4.5 - 4.7 each with respect to the three different PoVs.

Table 4.5: Classification Accuracy: PoV of the Mean - μ

Sensor Node	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
N1(a) -ISSNIP	4	0.0400	01	0.0033	0.9600	0.9967	299	96
N3(a) -ISSNIP	0	0.0000	03	0.0100	1.0000	0.9900	297	100
N5(a) -INTEL	0	0.0000	01	0.0033	1.0000	0.9966	295	100
N6(a) -INTEL	0	0.0000	05	0.0167	1.0000	0.9833	295	100

Table 4.6: Classification Accuracy: PoV of the Median - η)

Sensor Node	FN	FNR	FP	FPR	Sensitivity	Specificity	TN	TP
N1(a) -ISSNIP	0	0.0000	21	0.0700	1.0000	0.9300	279	100
N3(a) -ISSNIP	0	0.0000	05	0.0167	1.0000	0.9833	295	100
N5(a) -INTEL	0	0.0000	28	0.0946	1.0000	0.9054	268	100
N6(a) -INTEL	0	0.0000	02	0.0067	1.0000	0.9933	298	100

Table 4.7: Classification Accuracy: PoV of the Historic-Mean - μ'

Sensor Node	FN	FNR	FP	FPR %	Sensitivity	Specificity	TN	TP
N1(a) -ISSNIP	0	0.0000	06	0.0200	1.0000	0.9800	294	100
N3(a) -ISSNIP	0	0.0000	03	0.0100	1.0000	0.9900	297	100
N5(a) -INTEL	8	0.0769	03	0.0100	0.9231	0.9899	297	92
N6(a) -INTEL	0	0.0000	01	0.0033	1.0000	0.9967	299	100

Considering the results in Table 3.5, PoV of the mean achieves an average sensitivity of 0.99 and an average specificity of 0.99 with corresponding low false positive and false negative rates over the four different data distributions. Similarly the PoV of the median achieves an

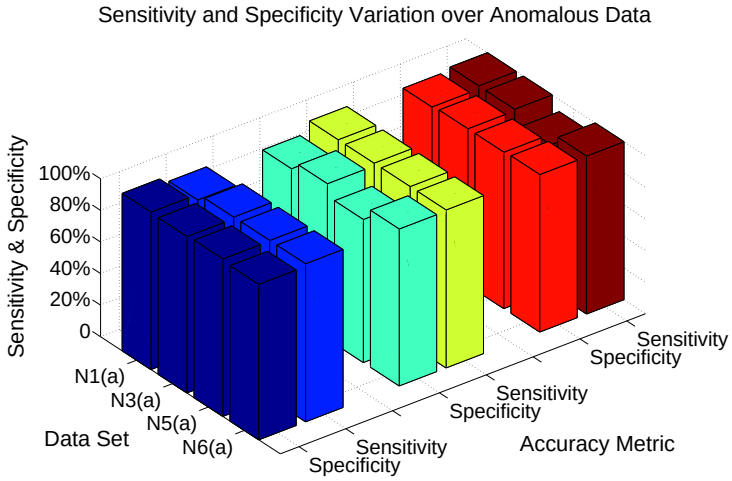

Figure 4.7: Sensitivity and Specificity Variation: The variation in sensitivity and specificity with regard to detecting both normal and abnormal behaviour over the four data distributions with anomalies [N1(a)-N6(a)] is presented above. The sensitivity mainains a near ideal of almost 100% in all instances while being accompanied with a very high corresponding specificity value of more than 90% in a mjority of distributions.

average of 1 and 0.95 in sensitivity and specificity metrics while the PoV of the historic mean outcomes in the values of 0.98 and 0.99 for average sensitivity and specificity. This clearly demonstrates that regardless of the PoV used all three proposed models achieve significantly high success rates in accurately differentiating between anomalous and normal data. The adaptive, resilient and robust features of the proposed methodology is made evident from the fact that each model achieves high performance accuracy for the variety of spatial distributions that the experimental data represent. The high values achieved for the accuracy metrics are consistent with regard to both orderly [N5(a) and N6(a)] and chaotically [(N1(a) and N3(a)] aberrant forms of abnormal behaviour. The relative variation in sensitivity and specificity values over the four different data sets is comparatively depicted in Figure 4.7.

The distributions representing the Mean Relative Entropy (MRE) variation for each of the three PoV models over the four different data sets with anomalies is given in Figure 4.8 and Figure 4.9. Figure 4.8 shows the variation of MRE for the two data distributions [N1(a) and N3(a)] having a more chaotically aberrant form of anomalous behaviour while Figure 4.9 gives the same for the two data distributions [N5(a) and N6(a)] with the more orderly form

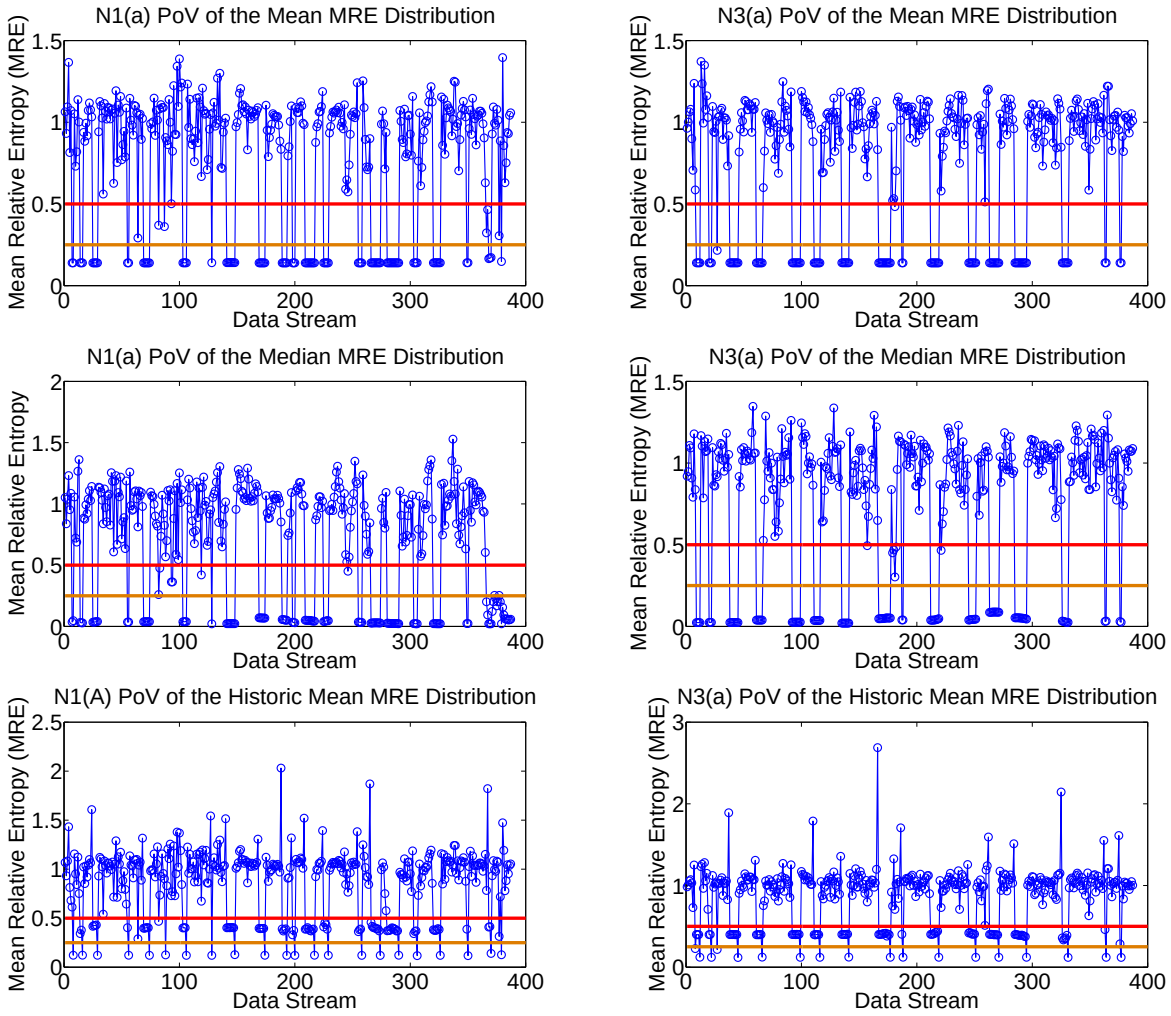

Figure 4.8: Mean Relative Entropy (MRE) Distributions for N1(a) and N3(a) Data: MRE value for normal data remains close to one consistently in the presence of an evolving data stream. The presence of anomalies in a highly chaotic manner [N3(a)] as well as in a less chaotic form while still being clearly aberrant [N1(a)] significantly deviates the MRE towards zero in all three PoV models. This drops to less than 0.25 for PoVs of the *mean* and *median* while it attains a value between 0.25 and 0.5 for most of the anomalous instances in PoV of the *historic mean*. The threshold values of 0.25 and 0.5 are marked in each of the plots.

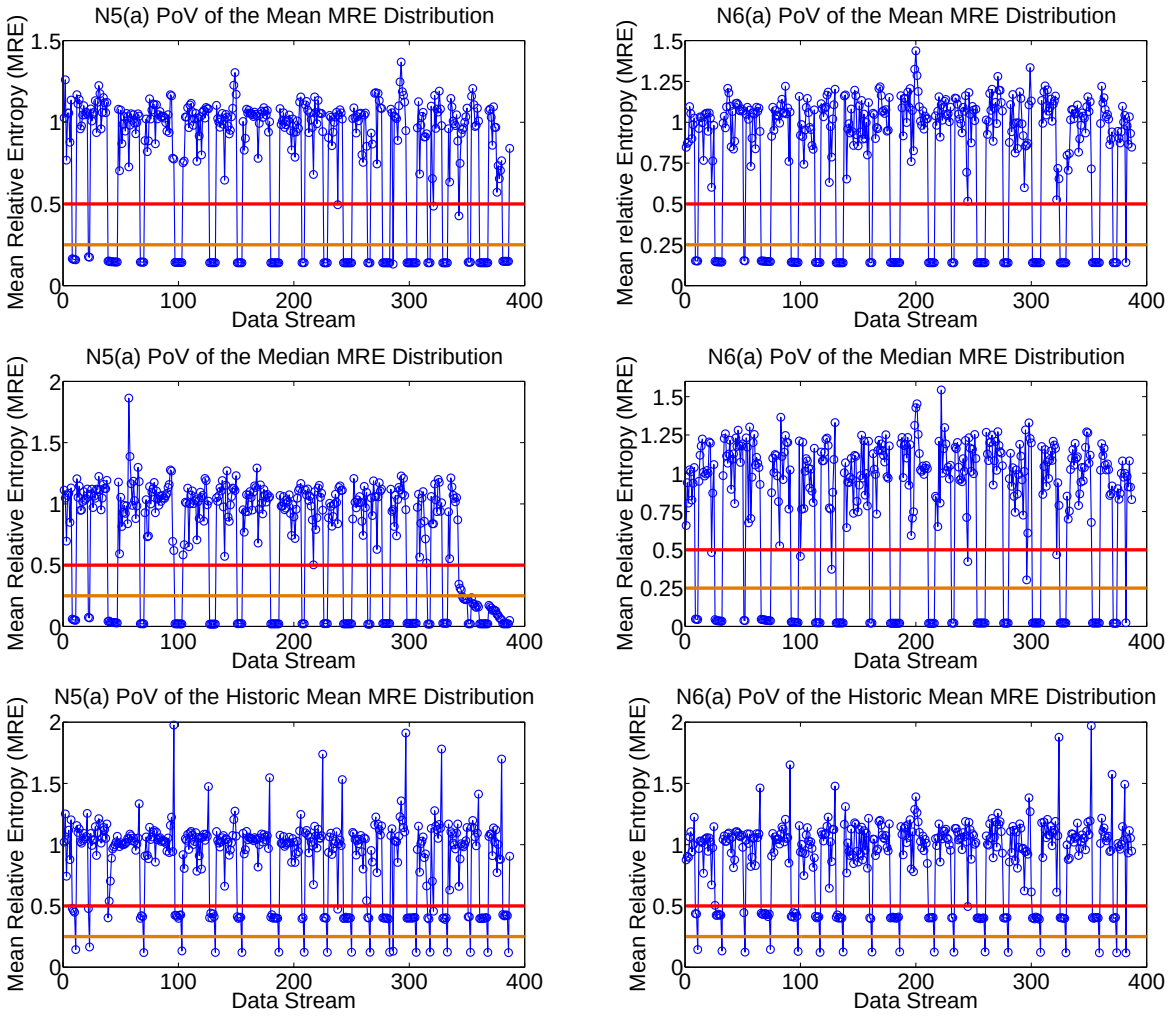

Figure 4.9: Mean Relative Entropy (MRE) Distributions for N5(a) and N6(a) Data: N5(a) and N6(a) presents a sequentially drifting normal data stream interspaced with non-chaotic anomalies. MRE maintains a value range close to one over the normal data while clearly deviating for the anomalies even when they are spatially very close to the normal drifting data for all three PoV models. As for the previous data the PoV of the historic mean attains a value for the anomalies that is between 0.5 and 0.25 while for the other two PoVs the MRE directly drops to that of less than 0.25 n both data sets.The threshold values of 0.25 and 0.5 are marked in each of the plots.

of correlated abnormal data. The two threshold levels of 50% and 25% are also marked on each MRE variation plot. It is clear from the plots that for each of the models the MRE value remains at a range that is significantly close to one. This is maintained even in the event of the normal data evolving in a dynamic manner as well as in the presence of both orderly and chaotically aberrant forms of anomalies. A significant deviation of more than 0.5 typically indicates an anomaly in all models. However, while this deviation goes to the extent of being less than 0.25 in value for the PoV models of the *mean* and the *median*, it hovers above 0.25 and below 0.5 for the PoV of the *historic mean*. Therefore, the use of less than 50% as a threshold for MRE value is successful in all cases while that of less than 25% is unfit for PoV of the historic mean. Given the high performance in the accuracy metrics as discussed previously it can be concluded that a threshold value of 50% is ideal irrespective of the PoV used and that other than in the PoV of the historic mean, using a threshold of 25% achieves an even higher accuracy level. It is also noted that even using a threshold of less than 50% MRE for the PoV of the historic mean, the attained accuracy is significantly higher than the other two PoVs which reaches that level when a threshold value of 25% is used.

4.4.4 Phase #3: Comparative Evaluation

In this section we do a comparative evaluation between the proposed model and some of the most related work in the area of anomaly detction on large volumes of streaming data. Table 4.8 represents a qualitative evaluation of selected work with that of the proposed model in a summarized format. The key metrics that are compared is the prior requirements for building the normal profile, ability to support dynamic data patterns with evolving behaviour, support for real-time analysis and quantitative measurements of accuracy with regard to the *detection rate* and the $false\ alarm\ rate$. A brief description on the major techniques adapted is also included.

Table 4.8: Comparison of Related Work

Scheme	Key-features	Training for Normal Profile	Adaptive for Dynamic Behaviour	Real-time Support	Detection Accuracy
Fuzzy Cluster based Kumarage et al. [Kumarage et al., 2013]	Fuzzy clustering with adaptive statistical thresholds	Fully-Unsupervised	Cluster-dependent	Batch-model	DR-95.08% FAR-0.42%
Entropy based Clustering Kumarage et al. [Kumarage et al., 2014]	Clusters determined using entropy with adaptive statistical thresholds	Fully-Unsupervised	Fully-dynamic	Batch-model	DR-98.42% FAR-2.21%
HSCBS Rajasegarar et al. [Rajasegarar et al., 2014]	Hyper-spherical clusters with nearest Neighbour approach	Fully-Unsupervised	Cluster-dependent	Batch-model	DR-95.15% FAR-1.33%
PCA Wang et al. [Wang et al., 2008]	PCA based multi-pronged detection	Prior-Training	Training Required	√	DR-98.80% FAR 0.4%
ADWICE Burbeck et al. [Burbeck and Nadjm-Tehrani, 2007]	Incremental distance based clustering	Prior-Training	Training Required	√	DR-95.00% FAR-0.03%
ADAM J.Hill et al. [Hill and Minsker, 2010]	Incremental autoregressive model with prediction	Prior-Training	Training Required	√	DR-95.39% FAR-4.60%
Proposed Model (PoV-EE)	PoV Entropy Evaluations	Fully-Unsupervised	Fully-dynamic	√	DR-99.89% FAR- 0.83%

4.5 Conclusion

In this chapter we present an anomaly detection model for real-time decision support in wireless sensor network data streams. An entropy criterion is proposed for differentiating anomalous incoming data over a buffered data space that is retained on each node while considering memory constraints. Three unique Point-of-Views are proposed in computing the entropy with regard to the continuously updated data space of the buffer. Therefore, with the advent of each new sensed data point the point-wise entropy is calculated and its variance evaluated from the perspectives of the (i) mean μ, (ii) median η and (iii) historic mean μ' (without the influence of latest data point). The differential change in entropy with the advent of each new data point is evaluated in relation to the current buffer elements of each sensor node in identifying anomalous points accurately in a unsupervised, non-probabilistic and adaptive manner. A secondary buffer level is introduced in order to correctly identify sudden shifts or spatial drifts in the data that may occur as part of normal behaviour.

Higher values for sensitivity and specificity metrics with lower false positives and false negatives demonstrate the accuracy of the proposed models with regard to a variety of different data distributions with temporaly evolving patterns. The detailed evaluations have also shown the robustness and viability of the proposed models with regards to adaptively handling dynamic and non-homogeneous demographies of sensed data as well as the capability to function in a near real-time manner in supporting decision making even as the data is sensed.

Chapter 5

Conclusion

Wireless Sensor Networks (WSNs) consisting of a large number of low cost, compact sensor nodes coupled with radio transceivers for communication are increasingly becoming the core technology in providing pervasive monitoring for a variety of applications. They are typically deployed in large numbers over a wide area unattended in enabling control and automation through continuous monitoring. Such application environments may either be homogeneous or heterogeneous with dynamic state changes possible. However, given their inherent constraints, WSNs are vulnerable to different faults, malfunction and malicious attacks. Misbehaviour resulting from such occurrences will manifest as observed anomalies in sensor network data streams. If left undetected these data anomalies will lead to wrong operational and control decisions that can impact wider society through severe economic losses, environmental damage and possible human harm. Therefore, it is vital that proper techniques are introduced in accurately detecting these anomalies given the critical role of the sensed data in core decision making. Acknowledging these concerns this research study has focused on introducing novel anomaly detection models in innovative data processing frameworks in a manner that is uniquely suitable for anomaly detection in the current application context of large scale wireless sensor networks.

Existing work on securing and preserving the integrity of WSN data has mostly focused on *prevention* than *detection*. These include designing secure communication protocols, managing link quality and introducing light-weight cryptographic schemes. However, there is always the

possibility that a new kind of attack may succeed with the exploitation of new vulnerabilities as well as the unavoidable occurrence of faults and malfunction mandating an imperative need for detection models. Most of the existing anomaly detection models are not specifically designed to be implemented on WSNs as has been discussed extensively in previous chapters. Of the limited amount of work that are presented uniquely for wireless sensor networks there are many issues limiting their applicability in current application contexts of more dynamic and evolving sensor network environments. The proposed models in this thesis span different areas including the use of a point-wise entropy criterion in differentiating between abnormal and normal data, unsupervised and dynamic data partitioning using distributed fuzzy memberships for data and real time anomaly detection based on relative evaluations of point of view (PoV) entropy in successfully overcoming these problems. Another area that has been explicitly focused on is the formation of statistical techniques for developed detection models that are non-parametric and non-probabilistic while being adaptive with regard to dynamic changes.

The research presented in this thesis has therefore explored the potential of unsupervised data driven methods at anomaly detection that are both scalable with regard to large scale implementation and adaptive in the context of dynamic state changes with evolving data patterns relating to the observation environment. It also focused on the unique features of a WSN environment including the resource constraints in coming up with innovative in-network data processing frameworks that minimize the associated communication complexity of proposed techniques. The complete contributions of this research have been presented in Chapters 2 to 4 of this thesis while a discussion and some remarks on key findings and their implications are presented in Section 5.1. This is followed by a discussion on some limitations and future work in Section 5.2

5.1 Concluding Remarks and Discussion

This thesis has focused on developing novel unsupervised models and associated techniques for distributed anomaly detection in industrial wireless sensor networks. Particular attention was given to developing scalable models that are adaptive and dynamic in the context of evolving data streams while being considerate of the resource limitations in the sensor network environ-

ment. Specially in reducing the communication overhead that is imposed on the network as it is shown that most of the energy use in such wireless sensor networks is on communication rather than computation. The core research contributions in the form of novel anomaly detection models together with in-network data processing frameworks were presented in three chapters. These are Chapter 2 to Chapter 4. The main research contributions there-in have been demonstrated to be successful in addressing the main research questions that were identified in Chapter 1 in line with the motivation provided.

In 2, it was shown that adapting *fuzzy c-means* clustering in an incremental model is capable of achieving unsupervised data partitioning that facilitates distributed anomaly detection in WSNs. The developed model was proven to be scalable and robust when implemented over a hierarchical topology for a considered WSN. The experimental investigations have clearly demonstrated that this approach achieves high accuracy in correctly classifying both normal and abnormal data with less false positives and false negatives. The investigations have also revealed that the number of clusters which is the only value that is operator mediated attains the best results when it is a number between 6-12. A lesser value or higher value would degrade the performance with regard to detection accuracy while a higher value would also inflict more complexity. However, it is demonstrated that within this range (6-12) of statically defined data partitions the model is robust with high detection rates across a variety of data distributions

A point-wise entropy criterion was presented as a viable direct measure to dynamically partition data for the purpose of unsupervised anomaly detection in chapter 3. The number of clusters were defined algorithmically in the proposed framework while adaptive anomaly detection was performed through second order statistical knowledge that is available at different analytical stages. The data processing framework was in-network and distributed with regard to reducing communication costs as compared to a centralized approach. Granular anomaly detection was performed in order to identify anomalies at different levels of granularity in the sensor network. The use of an entropy criterion to directly partition the data overcomes the limitations of hyper-spherical and hyper-elliptical clusters that are spatially local as formed in other data partitioning approaches. It also provides a measure in capturing data characteristics that are not limited to second order statistics. Extensive experiments have demonstrated its

adaptability with regard to dynamic data streams in higher accuracies that are achieved with a significant reduction in communication overheads.

While the presented models were *batch-wise* in the above mentioned chapters, focus was concentrated upon an *incremental* approach that achieve near real-time decision support in chapter 4. Therefore, a Point-of-View (PoV) approach was presented to evaluate the relative change in entropy that occurs as a data stream evolves. Three unique reference points were introduced and was shown to be successful in offering different lenses in evaluating the relative entropy change. The data processing model was local in the sense that the model is implemented at different nodes considering the data stream that is accessible at each node. In order to reduce the false alarm rate through correct detection of sensor drift and spatial level gaps a unique dual buffer model was also introduced in analyzing the data. This buffer space is to be determined based on the memory constraints of each node. It was shown that the model achieves high detection accuracy with low false alarms in a near real-time manner for a variety of data distributions representing evolving behaviour with different facets of abnormality. Some of the main findings of each core chapter are discussed in detail below.

5.1.1 Hierarchical Data Partitioning with Fuzzy Data Modelling for Scalable Anomaly Detection

A soft data partitioning approach based on fuzzy data modelling was presented as the first main contribution of this thesis with regard to achieving distributed anomaly detection in WSNs. The particular goals that this model achieved are;

- A fully unsupervised anomaly classification process without prior training with regard to the unavailability of labelled training data sets.
- Non-parametric and non-probabilistic anomaly detection without prior knowledge of the data distribution in relation to a dynamic observation environment with unpredictable state changes.
- An in-network and distributed data processing framework with high efficiency and reduced communication and computational overheads for the resource constrained WSN

environment.

Specifically, it was shown that fuzzy c-means clustering when adapted incrementally, on several analytical phases over a hierarchical node topology is a viable approach in determining data anomalies in an unsupervised manner that is both scalable and adaptive. Using a fuzzy data clustering process rather than a fixed binary process has allowed the data be evaluated for anomalies at different levels in enabling the model to accurately detect both local and global anomalies from a network perspective. The use of a centroid based data clustering approach has allowed the centroids to be used as representing complete data clusters in significantly reducing the associated communication costs. This has also worked to reduce misclassifications compared to other non-fuzzy classification methods, which are deterministic and restricted to local data correlations. A key problem in a data clustering approach for anomaly detection is determining the correct number of expected clusters to use. In the context of the proposed model the investigations have revealed that the ideal number of clusters should be in the range between 6-12. That is a number that is not too small and not too large with regard to minimizing complexity and resulting in a representative set of data partitions. While this number is statically defined, the experiments have clearly shown that within this range it achieves high classification accuracy in both sensitivity and specificity metrics with low false positives or false negatives. Experimental results showed reduced false positives with the sensitivity range increasing from between (12% - 48%) to (83.44% - 95.1%) compared to a non-fuzzy fixed-width clustering scheme.

The thresholding technique that was used on fuzzy membership scores and inter-cluster distances was also demonstrated to be robust and adaptive in evaluating anomalies over different hierarchical levels. As it only uses second order statistical knowledge (of mean and standard deviation) that is locally available at each analytic stage, the detection process is non-parametric and non-probabilistic. High detection accuracy was consistently maintained in all derived data distributions in the experimentation for both local and global phases of anomaly detection in highlighting the versatility of the model.

The distributed and in-network data processing framework that was introduced has also shown itself capable of significantly reducing the communication costs with regard to a cen-

tralized approach. In the proposed model only the locally identified cluster centroids and corresponding outliers were communicated to the next hierarchical level in reducing communication overheads. This has enabled an efficient incremental model where data can be processed as they are sensed. A communication reduction of more than 98% on average was achieved at all instances of experimentation. This demonstrated the significant benefit to the concerned sensor network in saving energy and maximizing active life-time. Further, the detailed comparison that was performed with regard to a distributed data clustering approach that employ fixed-width clustering has clearly shown that the proposed model far outpaces the latter in terms of classification accuracy and adaptability. Experimental results showed an average sensitivity in detection accuracy of (83.44 – 95.1%) and an average specificity in detection accuracy of (99.73 – 99.98%) compared to (12 – 48%) and (14 – 72%) for an existing data clustering approach employing fixed-width clustering.

5.1.2 Dynamic Data Partitioning with an Entropy Criterion for Multi-granular Anomaly Detection

The next core section of this thesis introduced a fully dynamic anomaly detection framework that offered multi-granularity analysis by employing an entropy criterion to directly partition the data. The key limitation of having a static number of clusters is successfully overcome through a dynamic determination of clusters that is completely data driven. The major goals that were achieved in the proposed model are as follows.

- An unsupervised, non-parametric and non-probabilistic data classification method that is fully dynamic over unpredictable and evolving data distributions.
- A robust detection of both isolated outliers and correlated clusters of abnormal data in non-homogeneous and multi-density data distributions.
- An efficient data processing framework with reduced computational and communication complexities for the resource constrained environment.

Specifically, a fully unsupervised data partitioning model was proposed to partition the sensed data on to cohesive regions dynamically based on *cumulative point-wise entropy* and

average relative density. The point-wise entropy measure was defined using euclidean distance similarity and the density measure was defined to be computed over two distance-neighborhood levels. The effects of noise and outliers were mitigated through prior segregation of such through the use of this average relative density measure. Introduction of this step was shown to be essential in dealing with the dynamic multi-density nature of the observation domain in subsequently facilitating a direct partitioning of the data through entropy. This solves the problem of isolated outliers having low entropy states that are similar to tightly correlated cohesive groups of data. The number and composition of the data partitions in this process were determined in a fully dynamic manner that is scalable over a hierarchical node topology. More significant is that in this approach the number of partitions are determined adaptively without prior knowledge in contrast to most existing methods.

The data partitions that were produced in this model are not limited to hyper-spherical or hyper-elliptical clusters as the entropy measure is capable of capturing data characteristics that are higher than second order statistics. As the entropy represents the level of orderliness in the data these data partitions prove to be much more viable with regard to an anomaly detection process. Granular evaluation of anomalies was attained over the hierarchical topology with minimal communication overhead between network levels. Thresholds in determining anomalies have been derived adaptively at each analytic stage from the second order statistical knowledge that is available. Each stage therefore, focused only on domain knowledge available at that network level in avoiding any arbitrary definitions of parameters or probabilistic assumptions. It is found that this leads to a more robust and accurate detection for dynamic data with differential density compared to existing methods as in [Chandola et al., a, Hodge and Austin, 2004]. Experiments yielded higher averages in the detected *true positives* and *true negatives* with a corresponding reduction in *false positives* and *false negatives* resulting in more than 94% in both sensitivity and specificity metrics consistently for a variety of different data distributions.

Scalability of the approach was highlighted through the high communication cost reduction of more than 85% in average when compared to a centralized data processing approach. The high sensitivity values that were achieved also demonstrated that the proposed model is

successful in accurately classifying both forms of abnormal behaviour (outliers and groups) as evident in the experimental data distributions that were used.

5.1.3 Point-of-View (PoV) Entropy Evaluations for Real-time Decision Support over Evolving Data Streams

The final core chapter of this thesis presented an anomaly detection framework for wireless sensor networks that enabled near real-time decision support. In this framework anomalies are differentiated and detected as they are sensed on each node locally by using an efficient incremental model that employ a unique Point-of-View (PoV) approach at evaluating relative change in entropy. The previous two approaches that were presented focused on *batch-wise* data processing models as opposed to an *incremental* approach that can support real-time detection of anomalies. The particular issues that were addressed are: (i) continuous streaming of large data volumes, (ii) dynamic fluctuations on data streams with evolving patterns, (iii) data heterogeneity over different streams in unmatched distributions and (iv) near real time impact on key decision making through streaming functionality. The following goals were attained with respect to these challenges as mentioned below.

- An unsupervised and non-probabilistic model for differentiating between normal and abnormal for evolving data streams
- Detection of anomalies in an online fashion in facilitating real time decision support
- Efficient in-network procedures with reduced complexity considering resource limitations

The proposed incremental model is implemented in-network with data processing performed in an online fashion. This supports the real time nature of a typical WSN monitoring application in mitigating negative impacts of batch models. In the proposed framework point-wise entropy is evaluated according to three unique Point-of-View (POV) approaches. These reference points were introduced as (i) PoV of the mean μ, (ii) PoV of the median η and (iii) PoV of the historic mean μ' (without the influence of latest data point). Each PoV acts as different lenses providing insight for more accurate detection when used within the context

of an entropy criterion. Then the relative change in entropy was compared through a *Mean Relative Entropy (MRE)* measure that is defined based on the calculated point-wise entropy values for each PoV in determining any anomalies. It was shown that all three introduced reference points are viable for the purpose of anomaly detection with the PoV of the historic mean μ' reporting the highest accuracy level in sensitivity and specificity metrics. Therefore the adaption of a PoV approach for entropy evaluation is shown to be uniquely successful in detecting anomalies in data streams while still accommodating for dynamic and evolving patterns of normal behaviour.

A key problem for incremental models that determine anomalies as they are sensed is how to accommodate instances where the observed data distribution may suddenly shift to a new range (a level shift) before continuing normal behaviour (in that range). This problem is overcome in the proposed approach by adding a second level of analysis regarding any potential anomalies that are initially detected. Thus, any element that is designated as anomalous through evaluation in the primary analytic phase is temporarily stored on a secondary buffer that accumulates the next few incoming data points. Relative entropy change is then again evaluated in the context of this new set of data before final determination of anomalies is made. It was shown that this technique is uniquely capable of identifying any level shifts or sudden sensor drift patterns that would otherwise contribute to a high false alarm rate. Two threshold levels of 25% and 50% were investigated for differentiating anomalies from the computed MRE values over a variety of different behaviour in the data streams. It was found out that while a threshold of 50% can be used regardless of the selected PoV, a threshold of 25% is not suitable with regard to the PoV of the historic mean. The detailed comparisons that were done highlighted the advantages of the proposed approach from a qualitative perspective as well as from a quantitative level. Detection accuracy of more than 98% in average sensitivity was accompanied by a low false alarm rate of less than 0.8% in the proposed approach significantly outperforming others.

5.2 Future Work

This thesis has focused on developing new distributed anomaly detection methods for industrial wireless sensor networks. The proposed models has dealt with the specific problems of how to achieve non-parametric and non-probabilistic anomaly detection in an unsupervised manner, how to make the process dynamic in the context of dynamic data streams and how to achieve near real-time decision support on evolving data streams. Further, distributed data processing frameworks have been introduced in achieving the above efficiently with reduced communication overheads that fits the resource constrained environment of WSNs. While these tasks have been successfully completed with high performance achieved in the proposed models there exists some other possibilities that haven't been the focus within the scope of this research. However, addressing these issues could considerably add value and improve on some of the limitations in the developed models, opening up pathways for future research based on the context of this thesis. These can be discussed as follows.

Determining the expected number of clusters for the anomaly detection approach with fuzzy data modelling

A key limitation in the work proposed in Chapter 2, is that the number of expected clusters need to be determined beforehand by the operator. Therefore, the number of expected clusters is a static attribute within the anomaly detection process. While this issue is addressed through the novel entropy based data partitioning approach presented in Chapter 3, that is a completely different model from the distributed fuzzy partitioning proposed in Chapter 2. Focusing on this issue, methods can be investigated to perform a dynamic allocation of the clusters to this fuzzy partitioning model. In such a model the number of clusters will be determined dynamically based on the information available from the data itself. A potential avenue of research in this direction is to randomly assign a number of clusters and then have a cluster merging or dividing process based on a considered distance measure or other statistical criteria. Another approach might be to compute the similarity of the data before the actual partitioning process. The number of clusters may be determined whether to be higher or lower according to how

similar or cohesive the data is. Exploring such methods to determine the number of clusters in a non-parametric and data driven way will significantly improve the original model based on fuzzy data modelling and make it fully dynamic with regard to evolving data.

Determining the node hierarchy for the anomaly detection framework

A factor that is not focused on the developed anomaly detection models in both Chapter 2 and Chapter 3 is the establishment of the node hierarchy. In implementing the models a static topological hierarchy is assumed a priori. However, in most real WSNs dynamic topology changes are possible. Furthermore, based on location and available resources the amount of processing a certain node can engage in may differ. Methods can be explored to determine different nodes and their role in a hierarchical topology using node connectivity, resource availability and neighborhood density considerations. Based on such considerations, the assumed node hierarchy may designed to be capable of evolving as the network state evolves. A potential direction for such research is to use multi-dimensional Voronoi diagrams for evaluating and determining nodes. This will provide a flexible and adaptive framework to define the assumed topology of a given WSN in a dynamic manner. Such a framework will significantly enhance the capabilities of the developed models in Chapter 2 and Chapter 3 in being more responsive and scalable with regard to network level considerations.

Evolving thresholds for real-time anomaly detection with PoV entropy evaluations

The work produced in Chapter 4 introduces a point-of-view entropy evaluation model to detect anomalies in near real-time. This is primarily done via thresholding a defined *Mean Relative Entropy (MRE)* value which accurately tracks the relative change in entropy as the data stream evolves. However, in the proposed work, these thresholds are fixed at 50% and 25%. Methods could be investigated in having this value determined adaptively based on the MRE distributions that are computed. Different statistical thresholds can also be introduced based on the main analysis on calculated entropy values. This will work to make the afore-mentioned model more adaptive and resilient with regard to fluctuations in observed data distributions.

Another aspect that is not focused on in the proposed model is the determination of the data space. Investigations can be performed specific to different application criteria in selecting the optimum data space to compare relative entropy change for a given data set. Such work will significantly improve the proposed model in terms of detection accuracy as well as make it more efficient with regard to memory allocation and resource utilization.

Bibliography

A. Aggarwal, S. Kunta, and P. Verma. A proposed communications infrastructure for the smart grid. In *Innovative Smart Grid Technologies (ISGT), 2010*, pages 1 –5, jan. 2010. doi: 10.1109/ISGT.2010.5434764. Cited on page 4.

I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393 – 422, 2002. ISSN 1389-1286. doi: http://dx.doi.org/10.1016/S1389-1286(01)00302-4. Cited on pages 66 and 96.

Y. an Huang, W. Fan, W. Lee, and P. Yu. Cross-feature analysis for detecting ad-hoc routing anomalies. In *Distributed Computing Systems, 2003. Proceedings of 23rd International Conference on*, pages 478 – 487, may 2003. doi: 10.1109/ICDCS.2003.1203498. Cited on pages 14, 26, 28, and 33.

H. Bar-El, H. Choukri, D. Naccache, M. Tunstall, and C. Whelan. The sorcerer's apprentice guide to fault attacks. *Proceedings of the IEEE*, 94(2):370 –382, feb. 2006. ISSN 0018-9219. doi: 10.1109/JPROC.2005.862424. Cited on pages 8, 24, 66, and 96.

J. Ben-Othman and B. Yahya. Energy efficient and qos based routing protocol for wireless sensor networks. *Journal of Parallel and Distributed Computing*, 70(8):849 – 857, 2010. ISSN 0743-7315. doi: 10.1016/j.jpdc.2010.02.010. URL http://www.sciencedirect.com/science/article/pii/S0743731510000341. Cited on pages 8 and 24.

M. Bertocco, G. Gamba, A. Sona, and S. Vitturi. Experimental characterization of wireless sensor networks for industrial applications. *IEEE Transactions on Instrumentation and Mea-*

surement, 57(8):1537–1546, 2008. ISSN 0018-9456. doi: 10.1109/TIM.2008.925344. Cited on pages 1, 65, 66, and 96.

J. C. Bezdek, R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(23):191 – 203, 1984. ISSN 0098-3004. doi: 10.1016/0098-3004(84)90020-7. URL http://www.sciencedirect.com/science/article/pii/0098300484900207. Cited on pages 13, 14, 25, 26, and 31.

P. Bodik, W. Hong, C. Guestrin, S. Madden, M. Paskin, and R. Thibaux. Intel lab data (db.csail.mit.edu/labdata/labdata.html), 2004. Cited on pages 13, 25, 37, 68, 84, 100, and 110.

A. Bose. Smart transmission grid applications and their supporting infrastructure. *IEEE Transactions on Smart Grid*, 1(1):11 –19, June 2010. ISSN 1949-3053. doi: 10.1109/TSG.2010.2044899. Cited on page 4.

A. Boukerche, R. W. N. Pazzi, and R. B. Araujo. Fault-tolerant wireless sensor network routing protocols for the supervision of context-aware physical environments. *Journal of Parallel and Distributed Computing*, 66(4):586 – 599, 2006. ISSN 0743-7315. doi: 10.1016/j.jpdc.2005.12.007. URL http://www.sciencedirect.com/science/article/pii/S0743731505002625. Algorithms for Wireless and Ad-Hoc Networks,. Cited on pages 8 and 24.

M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD InternationalL Conference on Management Of Data*, pages 93–104. ACM, 2000. Cited on pages 12 and 71.

K. Burbeck and S. Nadjm-Tehrani. Adaptive real-time anomaly detection with incremental clustering. *Information Security Technical Report*, 12(1):56 – 67, 2007. ISSN 1363-4127. doi: http://dx.doi.org/10.1016/j.istr.2007.02.004. Cited on pages 98, 102, and 122.

L. Buttyan, D. Gessner, A. Hessler, and P. Langendoerfer. Application of wireless sensor networks in critical infrastructure protection: challenges and design options [security and privacy in emerging wireless networks]. *IEEE Wireless Communications*, 17(5):44–49, 2010. ISSN 1536-1284. doi: 10.1109/MWC.2010.5601957. Cited on pages 1, 2, 65, 66, and 96.

V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, a. Cited on pages 16, 69, 70, and 130.

V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, b. Cited on pages 11 and 101.

V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection for discrete sequences: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 24(5):823–839, May 2012. ISSN 1041-4347. doi: 10.1109/TKDE.2010.235. Cited on pages 11 and 101.

V. Chatzigiannakis and S. Papavassiliou. Diagnosing anomalies and identifying faulty nodes in sensor networks. *Sensors Journal, IEEE*, 7(5):637 –645, may 2007. ISSN 1530-437X. doi: 10.1109/JSEN.2007.894147. Cited on pages 11, 14, 24, 26, 28, and 33.

A. P. R. da Silva, M. H. T. Martins, B. P. S. Rocha, A. A. F. Loureiro, L. B. Ruiz, and H. C. Wong. Decentralized intrusion detection in wireless sensor networks. In *Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks*, Q2SWinet '05, pages 16–23, New York, NY, USA, 2005. ACM. ISBN 1-59593-241-0. doi: http://doi.acm.org/10.1145/1089761.1089765. URL http://doi.acm.org/10.1145/1089761.1089765. Cited on pages 11 and 27.

A. Datta. A fault-tolerant protocol for energy-efficient permutation routing in wireless networks. *IEEE Transactions on Computers*, 54(11):1409–1421, 2005. ISSN 0018-9340. doi: 10.1109/TC.2005.172. Cited on pages 8 and 67.

R. Di Pietro, L. Mancini, C. Soriente, A. Spognardi, and G. Tsudik. Data security in unattended wireless sensor networks. *IEEE Transactions on Computers*, 58(11):1500–1511, 2009. ISSN 0018-9340. doi: 10.1109/TC.2009.109. Cited on pages 8 and 66.

D. Djenouri, L. Khelladi, and A. Badache. A survey of security issues in mobile ad hoc and sensor networks. *Communications Surveys Tutorials, IEEE*, 7(4):2 – 28, quarter 2005. ISSN 1553-877X. doi: 10.1109/COMST.2005.1593277. Cited on pages 5, 7, 23, 66, and 96.

W. Du, J. Deng, Y. Han, S. Chen, and P. Varshney. A key management scheme for wireless sensor networks using deployment knowledge. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 1, pages –597, March 2004. doi: 10.1109/INFCOM.2004.1354530. Cited on page 8.

C. Eik Loo, M. Yong Ng, C. Leckie, and M. Palaniswami. Intrusion detection for routing attacks in sensor networks. *International Journal of Distributed Sensor Networks*, 2(4):313–332, 2006. doi: 10.1080/15501320600692044. URL http://www.tandfonline.com/doi/abs/10.1080/15501320600692044. Cited on pages 14, 26, and 28.

M. Erol Kantarci and H. Mouftah. Wireless sensor networks for cost-efficient residential energy management in the smart grid. *IEEE Transactions on Smart Grid*, 2(2):314 –325, June 2011. ISSN 1949-3053. doi: 10.1109/TSG.2011.2114678. Cited on page 4.

M. Erol-Kantarci and H. Mouftah. Wireless sensor networks for smart grid applications. In *Electronics, Communications and Photonics Conference (SIECPC), 2011 Saudi International*, pages 1 –6, april 2011a. doi: 10.1109/SIECPC.2011.5876687. Cited on page 4.

M. Erol-Kantarci and H. T. Mouftah. Wireless multimedia sensor and actor networks for the next generation power grid. *Ad Hoc Networks*, 9(4):542 – 551, 2011b. ISSN 1570-8705. doi: 10.1016/j.adhoc.2010.08.005. ¡ce:title¿Multimedia Ad Hoc and Sensor Networks¡/ce:title¿. Cited on page 4.

L. Eschenauer and V. D. Gligor. A key-management scheme for distributed sensor networks. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, CCS '02, pages 41–47, New York, NY, USA, 2002. ACM. ISBN 1-58113-612-9. doi: 10.1145/586110.586117. URL http://doi.acm.org/10.1145/586110.586117. Cited on page 8.

P. Forero, A. Cano, and G. Giannakis. Distributed clustering using wireless sensor networks. *Selected Topics in Signal Processing, IEEE Journal of*, 5(4):707 –724, aug. 2011. ISSN 1932-4553. doi: 10.1109/JSTSP.2011.2114324. Cited on pages 11, 28, 58, and 61.

S. Gandham, M. Dawande, and R. Prakash. Link scheduling in wireless sensor networks: Distributed edge-coloring revisited. *Journal of Parallel and Distributed Computing*, 68(8):

1122 – 1134, 2008. ISSN 0743-7315. doi: 10.1016/j.jpdc.2007.12.006. URL http://www.sciencedirect.com/science/article/pii/S0743731508000142. Cited on pages 8 and 24.

D. Ganesan, B. Greenstein, D. Estrin, J. Heidemann, and R. Govindan. Multiresolution storage and search in sensor networks. *Trans. Storage*, a. Cited on page 73.

D. Ganesan, B. Greenstein, D. Estrin, J. Heidemann, and R. Govindan. Multiresolution storage and search in sensor networks. *Trans. Storage*, b. Cited on page 29.

J. Gao, Y. Xiao, J. Liu, W. Liang, and C. P. Chen. A survey of communication/networking in smart grids. *Future Gener. Comput. Syst.* Cited on page 4.

E. Gokcay and J. Principe. Information theoretic clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):158 –171, feb 2002. ISSN 0162-8828. doi: 10.1109/34.982897. Cited on pages 16, 69, and 76.

T. S. W. Group. Securing scada and industrial control systems. Technical report, Technical Support Working Group, http://www.tswg.gov/tswg/ip/scada.htm, March (2005). Cited on pages 6, 7, and 23.

J. Guevara, F. Barrero, E. Vargas, J. Becerra, and S. Toral. Environmental wireless sensor network for road traffic applications. *Intelligent Transport Systems, IET*, 6(2):177–186, 2012. ISSN 1751-956X. doi: 10.1049/iet-its.2010.0205. Cited on pages 1, 65, and 96.

V. Gungor and G. Hancke. Industrial wireless sensor networks: Challenges, design principles, and technical approaches. *IEEE Transactions on Industrial Electronics*, 56(10):4258–4265, Oct 2009. ISSN 0278-0046. doi: 10.1109/TIE.2009.2015754. Cited on page 1.

V. Gungor, B. Lu, and G. Hancke. Opportunities and challenges of wireless sensor networks in smart grid. *IEEE Transactions on Industrial Electronics*, 57(10):3557 –3564, oct. 2010. ISSN 0278-0046. doi: 10.1109/TIE.2009.2039455. Cited on page 4.

A. Gupta, A. Mukherjee, B. Xie, and D. P. Agrawal. Decentralized key generation scheme for cellular-based heterogeneous wireless ad hoc networks. *Journal of Parallel and Distributed*

Computing, 67(9):981 – 991, 2007. ISSN 0743-7315. doi: 10.1016/j.jpdc.2007.05.009. URL http://www.sciencedirect.com/science/article/pii/S0743731507000974. Cited on pages 8 and 24.

P. Gupta and P. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388 –404, mar 2000. ISSN 0018-9448. doi: 10.1109/18.825799. Cited on page 58.

F. Gustafsson. The marginalized likelihood ratio test for detecting abrupt changes. *IEEE Transactions on Automatic Control*, 41(1):66–78, Jan 1996. ISSN 0018-9286. doi: 10.1109/9.481608. Cited on page 102.

D. J. Hill and B. S. Minsker. Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. *Environmental Modelling and Software*, 25(9):1014 – 1022, 2010. ISSN 1364-8152. doi: http://dx.doi.org/10.1016/j.envsoft.2009.08.010. Thematic issue on Sensors and the Environment. Cited on pages 98, 102, and 122.

V. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22:85–126, 2004. ISSN 0269-2821. Cited on pages 16, 69, 70, and 130.

H. Kargupta, W. Huang, K. Sivakumar, and E. Johnson. Distributed clustering using collective principal component analysis. *Knowledge and Information Systems*, 3:422–448, 2001. ISSN 0219-1377. doi: 10.1007/PL00011677. URL http://dx.doi.org/10.1007/PL00011677. Cited on pages 11, 27, and 58.

E. M. Knorr, R. T. Ng, and V. Tucakov. Distance-based outliers: algorithms and applications. *The VLDB Journal*, 8:237–253, 2000. ISSN 1066-8888. 10.1007/s007780050006. Cited on pages 11 and 71.

H. Kumarage, I. Khalil, Z. Tari, and A. Zomaya. Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling. *Journal of Parallel and Distributed Computing*, 73(6):790 – 806, 2013. ISSN 0743-7315. doi: 10.1016/j.jpdc.2013.02.004. Cited on pages 72, 98, and 122.

H. Kumarage, I. Khalil, and Z. Tari. Granular evaluation of anomalies in wireless sensor networks using dynamic data partitioning with an entropy criteria. *IEEE Transactions on Computers*, PP(99):1–1, 2014. ISSN 0018-9340. doi: 10.1109/TC.2014.2366755. Cited on pages 98, 101, 102, and 122.

M. Li, Y. Wang, and Y. Wang. Complexity of data collection, aggregation, and selection for wireless sensor networks. *IEEE Transactions on Computers*, 60(3):386–399, 2011. ISSN 0018-9340. doi: 10.1109/TC.2010.50. Cited on page 91.

H. Liang, B. J. Choi, A. Abdrabou, W. Zhuang, and X. Shen. Decentralized economic dispatch in microgrids via heterogeneous wireless networks. *Selected Areas in Communications, IEEE Journal on*, 30(6):1061 –1074, july 2012. ISSN 0733-8716. doi: 10.1109/JSAC.2012.120705. Cited on page 4.

A.-F. Liu, P.-H. Zhang, and Z.-G. Chen. Theoretical analysis of the lifetime and energy hole in cluster based wireless sensor networks. *Journal of Parallel and Distributed Computing*, 71(10):1327 – 1355, 2011. ISSN 0743-7315. doi: 10.1016/j.jpdc.2011.05.003. URL http://www.sciencedirect.com/science/article/pii/S0743731511000992. Cited on pages 11 and 27.

X. Liu and M. Haenggi. Toward quasiregular sensor networks: Topology control algorithms for improved energy efficiency. *IEEE Transactions on Parallel and Distributed Systems*, 17 (9):975–986, 2006. ISSN 1045-9219. doi: 10.1109/TPDS.2006.130. Cited on page 73.

J. Lopez and J. Zhou. Overview of wireless sensor network security. *Wireless sensor network security. IOS Press, incorporated*, pages 1–21, 2008. Cited on page 5.

H. Luo, J. Luo, Y. Liu, and S. Das. Adaptive data fusion for energy efficient routing in wireless sensor networks. *IEEE Transactions on Computers*, 55(10):1286–1299, 2006a. ISSN 0018-9340. doi: 10.1109/TC.2006.157. Cited on pages 8 and 67.

L. Luo, Y. Zhang, and W. Zhu. E-science application of wireless sensor networks in eco-hydrological monitoring in the heihe river basin, china. *Science, Measurement Technology, IET*, 6(6):432–439, 2012. ISSN 1751-8822. doi: 10.1049/iet-smt.2011.0211. Cited on pages 1, 65, and 96.

X. Luo, M. Dong, and Y. Huang. On distributed fault-tolerant detection in wireless sensor networks. *IEEE Transactions on Computers*, 55(1):58–70, 2006b. ISSN 0018-9340. doi: 10.1109/TC.2006.13. Cited on pages 7 and 66.

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, pages 96–103, New York, NY, USA, 2000. ACM. ISBN 1-58113-301-4. doi: 10.1145/345848.345871. URL http://doi.acm.org/10.1145/345848.345871. Cited on page 29.

R. T. Marsh. Critical foundations: Protecting america's infrastructures. Technical report, Presidents Commission on Critical Infrastructure Protection, http://www.fas.org/sgp/library/pccip.pdf, october (1997). Cited on pages 6, 7, and 23.

Y. Mo and B. Sinopoli. Integrity attacks on cyber-physical systems. In *Proceedings of the 1st international conference on High Confidence Networked Systems*, HiCoNS '12, pages 47–54, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1263-9. doi: 10.1145/2185505.2185514. Cited on pages 7 and 66.

Y. Mun and C. Shin. *Secure Routing in Sensor Networks: Security Problem Analysis and Countermeasures*, volume 3480 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. ISBN 978-3-540-25860-5. Cited on pages 8 and 24.

R. Nowak. Distributed em algorithms for density estimation and clustering in sensor networks. *IEEE Transactions on Signal Processing*, 51(8):2245 – 2253, aug. 2003. ISSN 1053-587X. doi: 10.1109/TSP.2003.814623. Cited on pages 11, 28, 58, and 60.

I. Onat and A. Miri. An intrusion detection system for wireless sensor networks. In *Wireless And Mobile Computing, Networking And Communications, 2005. (WiMob'2005), IEEE International Conference on*, volume 3, pages 253 – 259 Vol. 3, aug. 2005. doi: 10.1109/WIMOB.2005.1512911. Cited on pages 11, 14, 24, 26, 28, and 30.

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, pages 315 – 326, march 2003. doi: 10.1109/ICDE.2003.1260802. Cited on pages 12 and 71.

A. Patcha and J.-M. Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, 51(12):3448 – 3470, 2007. ISSN 1389-1286. doi: http://dx.doi.org/10.1016/j.comnet.2007.02.001. Cited on pages 11 and 101.

V. Phipatanasuphorn and P. Ramanathan. Vulnerability of sensor networks to unauthorized traversal and monitoring. *IEEE Transactions on Computers*, 53(3):364–369, 2004. ISSN 0018-9340. doi: 10.1109/TC.2004.1261841. Cited on pages 7 and 66.

G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Commun. ACM*, a. Cited on page 91.

G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Commun. ACM*, b. Cited on pages 27 and 57.

D. Puccinelli and M. Haenggi. Wireless sensor networks: applications and challenges of ubiquitous sensing. *Circuits and Systems Magazine, IEEE*, 5(3):19–31, 2005. ISSN 1531-636X. doi: 10.1109/MCAS.2005.1507522. Cited on pages 2, 66, and 96.

V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava. Energy-aware wireless microsensor networks. *Signal Processing Magazine, IEEE*, 19(2):40 –50, mar 2002. ISSN 1053-5888. doi: 10.1109/79.985679. Cited on pages 9, 27, 57, and 91.

S. M. Rahman and K. El-Khatib. Private key agreement and secure communication for heterogeneous sensor networks. *Journal of Parallel and Distributed Computing*, 70(8):858 – 870, 2010. ISSN 0743-7315. doi: 10.1016/j.jpdc.2010.03.009. URL http://www.sciencedirect.com/science/article/pii/S0743731510000444. Cited on pages 8 and 24.

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*, pages 1 –5, oct. 2006. doi: 10.1109/ICCS.2006.301508. Cited on pages 14, 15, 25, 26, 28, 37, 61, and 62.

S. Rajasegarar, C. Leckie, and M. Palaniswami. Anomaly detection in wireless sensor networks. *IEEE Wireless Communications*, 15(4):34–40, Aug 2008. ISSN 1536-1284. doi: 10.1109/MWC.2008.4599219. Cited on page 96.

S. Rajasegarar, C. Leckie, and M. Palaniswami. Hyperspherical cluster based distributed anomaly detection in wireless sensor networks. *Journal of Parallel and Distributed Computing*, 74(1):1833 – 1847, 2014. ISSN 0743-7315. doi: http://dx.doi.org/10.1016/j.jpdc.2013.09.005. Cited on pages 98, 101, 102, and 122.

S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. *SIGMOD Rec.* Cited on page 70.

S. Ravi, A. Raghunathan, P. Kocher, and S. Hattangady. Security in embedded systems: Design challenges. *ACM Trans. Embed. Comput. Syst.*, 3:461–491, August 2004. ISSN 1539-9087. doi: http://doi.acm.org/10.1145/1015047.1015049. URL http://doi.acm.org/10.1145/1015047.1015049. Cited on pages 8, 24, 66, and 96.

L. Rokach. A survey of clustering algorithms. In O. Maimon and L. Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, pages 269–298. Springer US, 2010. ISBN 978-0-387-09823-4. 10.1007/978-0-387-09823-4-14. Cited on pages 12, 16, 69, and 72.

R. Roman, C. Alcaraz, and J. Lopez. The role of wireless sensor networks in the area of critical information infrastructure protection. *Information Security Technical Report*, 12(1):24 – 31, 2007. ISSN 1363-4127. doi: 10.1016/j.istr.2007.02.003. URL http://www.sciencedirect.com/science/article/pii/S1363412707000052. Cited on page 23.

A. Sharma, L. Golubchik, and R. Govindan. On the prevalence of sensor faults in real-world deployments. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference on*, pages 213–222, June 2007. doi: 10.1109/SAHCN.2007.4292833. Cited on pages 8 and 66.

A. B. Sharma, L. Golubchik, and R. Govindan. Sensor faults: Detection methods and prevalence in real-world datasets. *ACM Trans. Sen. Netw.* Cited on pages 8 and 66.

E. Shi and A. Perrig. Designing secure sensor networks. *IEEE Wireless Communications*, 11 (6):38 – 43, dec. 2004. ISSN 1536-1284. doi: 10.1109/MWC.2004.1368895. Cited on pages 5, 7, 23, 66, and 96.

K. Sohrabi, J. Gao, V. Ailawadhi, and G. Pottie. Protocols for self-organization of a wireless sensor network. *Personal Communications, IEEE*, 7(5):16 –27, oct 2000. ISSN 1070-9916. doi: 10.1109/98.878532. Cited on page 73.

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, pages 187–198. VLDB Endowment, 2006. URL http://dl.acm.org/citation.cfm?id=1182635.1164145. Cited on pages 8, 98, and 102.

B. Sun, L. Osborne, Y. Xiao, and S. Guizani. Intrusion detection techniques in mobile ad hoc and wireless sensor networks. *IEEE Wireless Communications*, 14(5):56–63, October 2007. ISSN 1536-1284. doi: 10.1109/MWC.2007.4396943. Cited on page 96.

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*, pages 269 –274, dec. 2010. doi: 10.1109/ISSNIP.2010.5706782. Cited on pages 13, 25, 39, 68, 82, 100, and 110.

D. K. Tasoulis and M. N. Vrahatis. Unsupervised distributed clustering, 2004. Cited on pages 11, 28, and 58.

T. Temel and N. Aydin. A novel information-theoretic clustering algorithm for robust, unsupervised classification. In *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*, pages 1 –4, feb. 2007. doi: 10.1109/ISSPA.2007.4555489. Cited on pages 16, 69, 76, and 77.

L. Villas, A. Boukerche, H. Ramos, H. de Oliveira, R. de Araujo, and A. Loureiro. Drina: A lightweight and reliable routing approach for in-network aggregation in wireless sensor

networks. *IEEE Transactions on Computers*, 62(4):676–689, 2013. ISSN 0018-9340. doi: 10.1109/TC.2012.31. Cited on pages 8 and 67.

W. Wang, X. Guan, and X. Zhang. Processing of massive audit data streams for real-time anomaly intrusion detection. *Computer Communications*, 31(1):58 – 72, 2008. ISSN 0140-3664. doi: http://dx.doi.org/10.1016/j.comcom.2007.10.010. Cited on pages 98, 102, and 122.

Y. Wang, W. Lin, and T. Zhang. Study on security of wireless sensor networks in smart grid. In *Power System Technology (POWERCON), 2010 International Conference on*, pages 1 –7, oct. 2010. doi: 10.1109/POWERCON.2010.5666729. Cited on pages 7 and 66.

M. Xie, S. Han, B. Tian, and S. Parvin. Anomaly detection in wireless sensor networks: A survey. *Journal of Network and Computer Applications*, 34(4):1302 – 1325, 2011. ISSN 1084-8045. doi: 10.1016/j.jnca.2011.03.004. URL http://www.sciencedirect.com/science/article/pii/S1084804511000580. Advanced Topics in Cloud Computing. Cited on pages 11, 24, 28, 97, and 101.

M. Xie, J. Hu, S. Han, and H. Chen. Scalable hyper-grid k-nn-based online anomaly detection in wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, PP (99):1–1, 2012. ISSN 1045-9219. doi: 10.1109/TPDS.2012.261. Cited on pages 12 and 71.

H. Xu, L. Huang, Y. Zhang, H. Huang, S. Jiang, and G. Liu. Energy-efficient cooperative data aggregation for wireless sensor networks. *Journal of Parallel and Distributed Computing*, 70 (9):953 – 961, 2010. ISSN 0743-7315. doi: 10.1016/j.jpdc.2010.05.009. URL http://www.sciencedirect.com/science/article/pii/S0743731510001103. Cited on pages 11 and 27.

R. Xu and I. Wunsch, D. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645 –678, may 2005. ISSN 1045-9227. doi: 10.1109/TNN.2005.845141. Cited on pages 12 and 72.

Y. Yang, G. Xie, X. Xu, and Y. Jiang. A monitoring system design in transmission lines based on wireless sensor networks. *Energy Procedia*, 12:192 – 199, 2011. ISSN 1876-6102. doi: 10.1016/j.egypro.2011.10.027. Cited on page 4.

J. Yao, M. Dash, S. Tan, and H. Liu. Entropy-based fuzzy clustering and fuzzy modeling. *Fuzzy Sets and Systems*, 113(3):381 – 388, 2000. ISSN 0165-0114. doi: 10.1016/S0165-0114(98)00038-4. Cited on pages 72 and 77.

W. Ye and J. Heidemann. Enabling interoperability and extensibility of future scada systems. Technical report, USC Information Science Institute, november 2006. Cited on page 23.

J. Yick, B. Mukherjee, and D. Ghosal. Wireless sensor network survey. *Computer Networks*, 52(12):2292 – 2330, 2008. ISSN 1389-1286. doi: http://dx.doi.org/10.1016/j.comnet.2008.04.002. URL http://www.sciencedirect.com/science/article/pii/S1389128608001254. Cited on pages 1 and 2.

Z. Yu and Y. Guan. A key management scheme using deployment knowledge for wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 19(10):1411–1425, 2008. ISSN 1045-9219. doi: 10.1109/TPDS.2008.23. Cited on pages 8 and 67.

Y. Zhang, N. Meratnia, and P. Havinga. Outlier detection techniques for wireless sensor networks: A survey. *Communications Surveys Tutorials, IEEE*, 12(2):159 –170, quarter 2010. ISSN 1553-877X. doi: 10.1109/SURV.2010.021510.00088. Cited on pages 11, 24, 28, 30, 97, and 101.

Z. Zhang, M. Ma, and Y. Yang. Energy-efficient multihop polling in clusters of two-layered heterogeneous sensor networks. *IEEE Transactions on Computers*, 57(2):231–245, 2008. ISSN 0018-9340. doi: 10.1109/TC.2007.70774. Cited on page 73.

F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich. Collaborative signal and information processing: an information-directed approach. *Proceedings of the IEEE*, 91(8):1199–1209, Aug 2003. ISSN 0018-9219. doi: 10.1109/JPROC.2003.814921. Cited on page 9.

S. Zhao, I. Seskar, and D. Raychaudhuri. Performance and scalability of self-organizing hierarchical ad hoc wireless networks. In *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, volume 1, pages 132 – 137 Vol.1, march 2004. doi: 10.1109/WCNC.2004.1311531. Cited on page 73.