An Anomaly Detection Model Utilizing Attributes of Low Powered Networks, IEEE 802.15.4e/TSCH and Machine Learning Methods

by

Sajeeva Salgadoe

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Business and Information Technology

Ontario Tech University

Oshawa, Ontario, Canada

December 2019

THESIS EXAMINATION INFORMATION

Submitted by: **Sajeeva Salgadoe**

PhD in **Computer Science**

An Anomaly Detection Model Utilizing Attributes of Low Powered Networks, IEEE 802.15.4e/TSCH and Machine Learning Methods

An oral defense of this thesis took place on Nov 28 2019 in front of the following examining committee:

Examining Committee:

Chair of Examining Committee	Dr. Franco Gaspari
Research Supervisor	Dr. Fletcher Lu
Examining Committee Member	Dr. Miguel Vargas Martin
Examining Committee Member	Dr. Patrick Hung
University Examiner	Dr. Carolyn McGregor
External Examiner	Dr. Ali Miri

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

ABSTRACT

The rapid growth in sensors, low-power integrated circuits, and wireless communication standards has enabled a new generation of applications based on ultra-low powered wireless sensor networks. These are employed in many environments including health-care, industrial automation, smart building and environmental monitoring. According to industry experts, by the year 2020, over 20 billion low powered, sensor devices will be deployed and an innumerable number of data objects will be created.

The objective of this work is to investigate the feasibility and analyze optimal methods of using low powered wireless characteristics, attributes of communication protocols and machine learning techniques to determine traffic anomalies in low powered networks. Traffic anomalies can be used to detect security violations as well as network performance issues. Both live and simulated data have been used with four machine learning methods, to examine the relationship between performance and the various factors and methods.

Several factors including the number of nodes, sample size, noise influence, model aging process and classification algorithm are investigated against performance accuracy using data collected from an operational wireless network, comprising more than one hundred nodes, during a six-month period. An important attribute of this work is that the proposed model is able to implement in any low powered network, regardless of the software and hardware architecture of individual nodes (as long as the network complies with an open standard communication mechanism). Furthermore, the experiment portion of this work includes over 80 independent experiments to evaluate the behaviour of various attributes of low powered networks.

Machine learning models trained using carefully selected input features and other factors including adequate training samples and classification algorithm are able to detect traffic anomalies of low powered wireless networks with over 95% accuracy. Furthermore, in this work, a framework for an aggregated classification model has been evaluated and the experiment results confirm a further improvement of the prediction accuracy and a reduction of both false positive and negative rates in comparison to basic classification models.

Keywords: Security; Ultra Low Powered Networks; IEEE 802.15.4e/TSCH; Anomaly Detection; Machine Learning

AUTHOR'S DECLARATION

I hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize the University of Ontario Institute of Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Ontario Institute of Technology to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

Sajeeva Salgadoe

STATEMENT OF CONTRIBUTIONS

Part of the work described in this thesis has been published as

Sajeeva Salgadoe, Fletcher Lu, An Anomaly Detection Model for Ultra Low Powered Wireless Sensor Networks Utilizing Attributes of IEEE 802.15.4e/TSCH. JCM 14(5): 335-341 (2019)

DEDICATION

Writing this dissertation has been a long journey and I need to rely on support of others to overcome some of the hurdles. I want to thank my thesis supervisor Dr. Fletcher Lu and the committee for all the support. I also would like to thank Jim for allowing me to take time out from the day job to work on the research. I would like to thank my family and friends for all the encouragement given over the years. Anne was a tremendous help and she was always there to help me without any hesitation. Thank you very much for that. A very special thanks to my wife, for taking care of me and my kids during the last few years, her faithful support and belief allowed me to pursue my goals.

I would like to dedicate this work to my late father who decided not to talk to me until I have a Ph.D, and to my wife and the four pillars of my life: Sehansie, Shihan, Dylan and Dihan who gave me all the motivation to work on this research every day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

6LoWPAN	IPv6 over low powered Personal Area Networks
6TiSCH	IPv6 over the TSCH mode of IEEE 802.15.4e
ACK	Acknowledgment (Packet Type)
AES	Advanced Encryption Standard
AMCA	Asynchronous Multi-channel Adaptation
ASH	Auxiliary Security Header
ASN	Absolute Slot Number
BDT	Bagged Decision Tree
BOP	Back Off Period
CCA	Clear Channel Assessment
CFT	Contention Free Time
CIA	Confidentiality, Integrity and Availability
CoAP	Constrained Application Protocol
CPU	Central Processing Unit
CRTC	Canadian Radio-Television and Telecommunications Commission
CSMA/CA	Carrier Sense Multiple Access with Collision Avoid
CSV	Comma-Separated Values
CTS	Clear to Send
DOS	Denial of Service
DSME	Deterministic & Synchronous Multi-Channel Extension
DT	Decision Tree
DTR	Decision Tree Regression
EB	Enhanced Beacons
FFD	Fully Functional Device
FN	False Negatives
FP	False Positives
GTS	Guaranteed Time Slots
HART	High-way Addressable Remote Transducer
HIPPA	Health Insurance Portability and Accountability Act
ICMP	Internet Control Message Protocol

IE	Information Elements
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IoT	Internet of Things
IP	Internet Protocol
IPv4	IP version 4
IPv6	IP version 6
ISM	Industrial, Scientific, and Medical radio band
IV	Initialization Vector
JSON	JavaScript Object Notation
KNN	K-Nearest Neighbours
LD	Link Distance
LLC	Link Layer Control
LLDN	Low Latency Deterministic Network
LoWPAN	Low Powered Wireless Personal Area Network
LoWSN	Low Powered Wireless Sensor Network
LQI	Link Quality Indicator
LRWPAN	Low Rate Wireless Personal Area Network
MAC	Media Access Control
MiTM	Man in The Middle attacks
ML	Machine learning
MTU	Maximum Transmission Unit
ND	Neighbour Discovery
NN	Neural Networks
OSPF	Open Shortest Path First
PAN	Personal Area Network coordinator
PCA	Principle Component Analysis
PHY	Physical layer
PIPEDA	Personal Information Protection and Electronic Documents Act
RF	Random Forest
RFR	Random Forest Regression
RIP	Routing Information Protocol
RPL	Routing Protocol for Low-power and lossy networks
RSSI	Radio Signal Strength Indicator

RTS	Request to Send
SNR	Signal to Noise Ratio
SSID	Service Set Identification
SSL	Secure Socket Layer
STD	Standard Deviation
SVM	Support Vector Machines
SVR	Support Vector Regression
TCP	Transmission Control Protocol
TKIP	Temporal Key Integrity Protocol
TSCH	Time Slotted Channel Hopping
TSMP	Time Synchronized Mesh Protocol
UDP	User Datagram Protocol
WEP	Wired Equivalent Protection

Chapter 1. Introduction

In recent years, a tremendous growth of solutions based on low powered wireless sensor devices has been witnessed [1, 2, 3, 6]. Several factors including technological advancements, cost, simplicity and easy deployment have led to the unfolding of new dimensions, creating richer living experiences and economic benefits [6, 9]. Emerging paradigms such as Internet of Things (IoT) and cloud computing have also significantly, contributed to the growth. The development of wireless communication technologies, including effective Media Access Control (MAC) mechanisms [8, 10], modulation techniques and noise reduction methods have also contributed to the immense growth of wireless sensor networks. Yet, different wireless sensor networks have different requirements operationally, securely and functionally. (In this work, different terminologies namely, Wireless Sensor Networks (WSN), Low Powered Wireless Sensor Networks (LoWSN), Low Rate Wireless Sensor Networks (LRWSN), Low Powered Wireless Personal Area Networks (LoWPAN), Low Rate Wireless Personal Area Networks (LRWPAN) and IoTs are used interchangeably. However, they all refer to wireless networks containing nodes with various constrained resources including processing power, memory, transmission, storage and battery power.

The growth of Internet of Things, including low powered wireless devices is unprecedented and statistics indicate that as of 2016, there were over 6.4 billion such devices on the Internet, up 30 percent from the previous year, with predictions indicating over 20 billion low powered connected devices by 2020 [6, 9]. Due to the diverse nature of low powered wireless data, security controls must be designed to satisfy security, functionality and business requirements of a particular environment. Furthermore, certain types of low powered wireless data, such as personal health data are protected by regulatory laws and unauthorized disclosure or modification of such data may lead to serious consequences [4, 5, 11]. LoWSN deployed in industrial environments generates critical data, which require timely attention and accurate delivery, to prevent unexpected delays and financial repercussions [11]. Intelligent transportation systems and smart home/city applications are heavily dependent on accurate information, retrieved from low powered devices and data unavailability or contamination could be catastrophic [11].

Data protection of wireless sensor networks is based on three fundamental goals: Confidentiality, Integrity and Availability (CIA) [12]. Each of these goals is associated with a set of controls to satisfy the overall security objectives. For instance, controls associated with confidentiality enforce mechanisms to protect data against unauthorized disclosure. Different techniques, such as encryption, access control, authentication, authorization and data classification can be used to prevent unauthorized disclosure [6]. Data Integrity deals with protecting low powered wireless data against unauthorized modification [7]. Techniques, such as access control, authentication, authorization, hash functions and digital signatures can be used to protect the integrity of low powered wireless data [7]. Availability involves data accessibility, in a timely manner and techniques, such as redundancy, backup, hot-standby, intrusion detection and physical security controls. It can be implemented to prevent threats related to the availability of low powered wireless data.

In 2003, IEEE 802.15.4 standard was drafted by the Institute of Electrical and Electronics Engineers (IEEE) to define the MAC and Physical (PHY) layer specification for LRWPAN [8]; it has been widely used in low powered wireless network implementations [1]. However, existing standards such as IEEE 802.15.4 were unable to satisfy the emerging demand for super-low powered wireless requirements. IEEE 802.15.4 has defined a protection mechanism with the use of an Auxiliary Security Header (ASH) [8]. However, the implementation of ASH in a low powered environment would drastically degrade the overall performance. According to Diadone et al. [1], the use of ASH reduces 33.8 percent of the amount of data transmitted in a frame and increases energy consumption by 61.12 percent in 802.15.4 networks. Consequently, in 2012, IEEE defined a MAC amendment for

802.15.4 that was drafted as 802.15.4e to enhance the functionality of 802.15.4-2006 and to better support industrial markets [13, 14].

Despite the fact that threats associated with wireless sensor networks are complex [9], it is important to investigate different venues to secure sensor data. In this thesis, the feasibility of using the characteristics of low powered wireless sensor devices and attributes of IEEE 802.15.4e, with the use of machine learning techniques to detect anomalies in wireless sensor networks is investigated.

The rest of the work is structured as follows: A Background chapter gives a technical overview of the IEEE 802.15.4e amendment and the Time Slotted Channel Hopping (TSCH) media access mode. The Background section also provides a technical overview of machine learning and statistical methods and other supplementary tools utilized in this thesis. A discussion on security related to low powered wireless networks can also be found in the Background section. In the Related Work section, some of the interesting work related to the protection of low powered wireless data is discussed. A Motivation chapter provides a brief outline of the significance of this study. A Research Design and Methods section provide a detailed overview of the study plan including research the questions, data collection and analysis methods, assumptions, threat models and experimental results. A discussion about experimental results, challenges and potential directions for further research can be found in the Summary/Discussion section.

Chapter 2. Background

In the following chapter, background information related to technologies used in this work including communication protocol-specific information, machine learning, statistical methods, a discussion about security threats in network environments as well as known security concerns in wireless network protocols are discussed.

LoWSNs consist of the number of battery-operated wireless sensor devices to measure environmental, physical or physiological properties in discrete time intervals for prolong periods without replacing energy sources (batteries). Several researchers have investigated the energy consumption of low powered devices and according to their findings, regardless of the effectiveness of the hardware architecture of the radio module, the data transmission consumes significantly higher amounts of energy compared to the rest of the activities on a low powered node. Therefore, low powered nodes are specially configured to minimize resource utilization using various techniques including data pre-processing, buffering, summarization and compression. Consequently, data transmission in low powered environments is minimized and highly predictable. Different techniques including conditional rules and machine learning methods can be used to identify a finite number of operational behaviours and activities of a low powered wireless network. Even though low data rates and dense node distribution are characteristics of LoWSNs, conventional open standards such as IEEE 802.11[a, b, g, n] are operated in larger channel-widths and unable to accommodate a larger amount of channels and simultaneous usage of available channels. However, because of the closely connected compact nature of low powered wireless networks, low-powered radio transceivers may trigger negative effects including single-channel interference, multipath fading and hidden node effect. Furthermore, complex modulation techniques and expensive interference avoidance techniques used in conventional wireless protocols require higher computational resources [8, 10].

A number of limitations in adapting conventional network protocols in low powered environments forced the Institute of Electrical and Electronics Engineers (IEEE) to design a standard to operate effectively in low powered environments. Consequently, IEEE 802.15.4e has been introduced to enhance the Media Access Control (MAC) layer functionality to accommodate ultra-low-powered communication and it is considered as the latest generation reliable media access mechanism for low powered wireless networks [10]. The channel agility of a wireless network operating on IEEE 802.15.4e/TSCH mode provides higher reliability in noisy environments. Furthermore, IEEE 802.15.4e only defines the functionality of MAC layer and it allows other open standards such as IPv6 over low powered Personal Area Networks (6LoWPAN), routing protocol for low-power and lossy networks (RPL) and constrained application protocol (CoAP) to be utilized in 802.15.4e environments.

Machine learning methods can be used to train models to approximate potential outcomes based on previous observations. Machine learning methods are used in different applications including descriptive analytics, diagnostics, predictive analytics and prescriptive analytics [60, 61]. In this thesis, machine learning methods are used to predict anomalies in low powered wireless networks operating in IEEE 802.15.4e/TSCH mode. More details about machine learning algorithms and ensemble methods utilized in this work can be found in the Background chapter.

In the following subsection, details about various technologies including IEEE 802.15.4e, 6TiSCH, COAP, machine learning algorithms, optimization techniques, data security concerns in general as well as known security threats in low powered wireless network protocols are discussed.

2.1 IEEE 802.15.4e

2.1.1 Introduction

The original draft of IEEE 802.15.4 has been designed for low powered communication in single-hop environments [15]. Despite the fact that IEEE 802.15.4 is a widely used standard for low powered sensor networks, it has a few fundamental flaws [16], such as (1) high energy consumption, due to the fact that relay nodes must always be active, and (2) utilizing a single channel which may lead to interference and fading [17]. The IEEE formed the 802.15.4e working group to design a low-powered multi-hop standard to satisfy emerging needs for low powered, embedded industrial applications [10]. Several media access control (MAC) behaviour methods are defined by 802.15.4e namely, Deterministic & Synchronous Multi-Channel Extension (DSME), Low Latency Deterministic Network (LLDN), Time Slotted Channel Hopping (TSCH), and Asynchronous Multi-channel Adaptation (AMCA). They are tailored to satisfy various network requirements [10]. The TSCH maintains high reliability and low duty cycles, using time-synchronization and channel hopping. The TSCH mode has emerged from Time Synchronized Mesh Protocol (TSMP) [18] and High-way Addressable Remote Transducer (HART) [19] Technology. In TSCH mode, nodes are synchronized to a slotframe (Fig. 2) structure and to a network coordinator, also known as the personal area network coordinator (PAN coordinator) [10]. The TSCH mode is primarily used in mesh environments, where some low powered nodes are unable to reach a central controller, directly [10]. Furthermore, TSCH mode is specially tailored for environments with low throughput, high latency and small packet size requirement. The following figure depicts a logical topology of a wireless sensor network, operating on IEEE 802.15.4e TSCH mode.

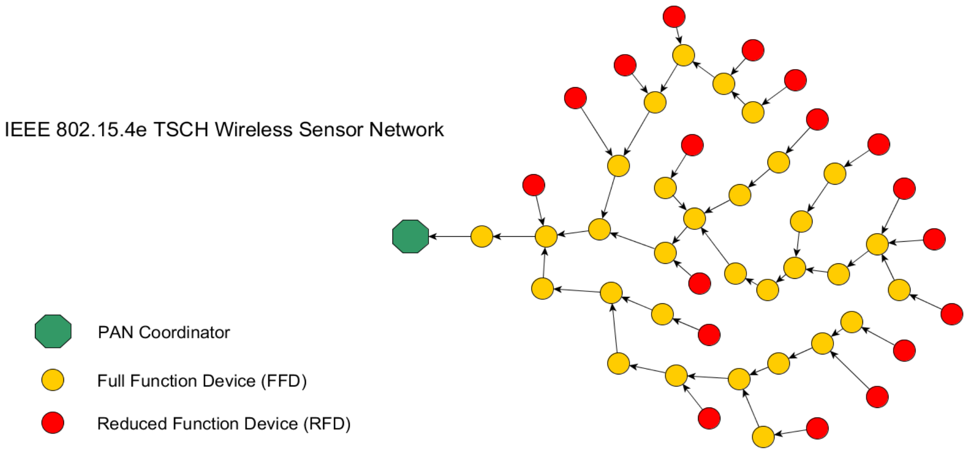

IEEE 802.15.4e TSCH Wireless Sensor Network
PAN Coordinator
Full Function Device (FFD)
Reduced Function Device (RFD)

Fig. 1. IEEE 802.15.4e/TSCH logical topology

A slot-frame is a group of time slots repeated over time and a time-slot is a predetermined period of time used by nodes to exchange data [10]. Each synchronized node follows a schedule, dictating the allowed operation for a particular node, during a timeslot. Several types of timeslot operations are defined by the TSCH mode such as beacon transmit timeslots, beacon listen timeslots, shared timeslots, dedicated transmit/receive timeslots and contention access timeslots [13]. Each timeslot schedule specifies which two nodes are participating in data exchanged, using a specific channel. This should be long enough for a transmitter to send a maximum-length packet and the receiver to successfully acknowledge it [10]. Furthermore, multiple node-pairs can communicate during a single timeslot using different channel offset [10], which increases the network capacity as well as, mitigating the interference and multipath fading, with the use of dynamic channel hopping [17]. Based on the schedule created by the PAN coordinator, an individual node can be put into transmit or receive mode, using a specific channel or switch to sleep mode [10]. The 802.15.4e describes the mechanism to execute the schedule; however, it is the responsibility of the application layer to formulate a schedule, suitable for a particular environment [10]. Shared timeslots are based on a contention access mechanism [10] and they are primarily used for sending management and signalling data such as association/dis-association requests. TSCH is a deterministic protocol where nodes are only awake during timeslots, which have assigned operations for a particular node [13]. The following diagram is an example of a slotframe with ten timeslots.

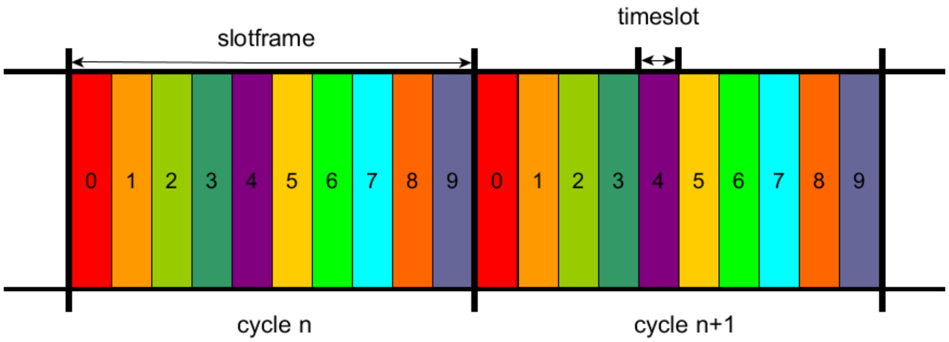


Fig. 2. IEEE 802.15.4e/TSCH slot-frame

Each timeslot can be divided into multiple cells and the amount of cells is dependent on the available channel list [10]. The channel list is formulated using a regulatory requirement and localized factors such as interference [20]. Figure 3 depicts a portion of a schedule that has a slotframe with ten timeslots and five usable channels. Each cell in a timeslot is assigned a node-pair to utilize a unique channel. However, each cell can be shared by multiple node-pairs using a contention access mechanism [10]. In TSCH, the PAN coordinator uses an attribute known as a "link" using two parameters (timeslot number and channel offset) to assign a directed communication between two nodes [10] to a cell. A specific frequency for a particular cell is derived using the following formula.

$freq_{active} = Freq_{list}$ [$(ASN + ch_{Ofset}) \bmod nrOf_{Channels}$] where

$freq_{active}$ = Active Frequency

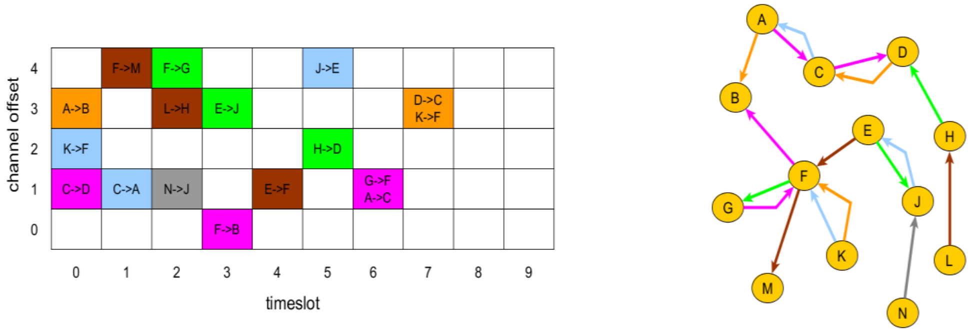
$Freq_{list}$ = Available usable frequencies

ASN = Absolute Slot Number (a unique number used by the TSCH to identify a timeslot and it indicates the total number of slots elapsed since the network was formed [10])

Ch_{offset} = Assigned by the PAN coordinator to a particular "link"

$nrOf_{channels}$ = Number of usable channels in the channel list

Figure 3 describes a sample TSCH schedule and corresponding activities in a wireless network operating in the IEEE 802.15.4e TSCH mode. Each color represents a different frequency used for communication.



channel offset	0	1	2	3	4	5	6	7	8	9
4		F->M	F->G			J->E				
3	A->B		L->H	E->J				D->C K->F		
2	K->F					H->D				
1	C->D	C->A	N->J		E->F		G->F A->C			
0				F->B						
	timeslot									

(a) Frequency Schedule

(b) Dedicated & Shared Links

Fig. 3. IEEE 802.15.4e/TSCH schedule

2.2 Machine Learning

2.2.1 Introduction

Machine learning (ML) is a scientific discipline that addresses the designing of systems to learn automatically and evolve with experience. In recent years, machine learning and related areas have generated a tremendous interest among various groups, including the research community, various industries as well as, government and private institutes.

2.2.2 Methods

Machine learning models can be categorized into several classes based on the learning method and application objectives. In the following, a brief description of different machine learning models is outlined.

2.2.2.1 Supervised/Unsupervised

In supervised learning, a labelled data set is used to train prediction models. Labelled data sets include input data and corresponding output values [63]. The input data is in a form of vector (also known as feature vector) and each element of the vector represents an independent feature. The purpose of the training process is to determine the optimal feature representation [63]. The number of samples required for this training process is dependent on several factors including prediction accuracy objectives, number of input features, number of intermediate layers and the machine learning algorithm. Several machine learning algorithms including Support Vector Machines (SVM), Decision Tree (DT), Linear/Logistic Regression, Naïve Bayes (NB), Random Forest (RF) and Artificial Neural Networks (NN) can be used in supervised learning [63, 64, 65, 66]. The fundamental approach of supervised learning models is to minimize a cost (error) function using an optimization technique such as gradient descent [67]. Machine learning applications, based on supervised learning are commonly used in pattern recognition and in classification including image processing, natural language processing and spam detection [63]. In this work, a number of classification algorithms are utilized to predict traffic anomalies in low powered wireless networks.

In contrast to supervised learning, the primary task of unsupervised learning is to determine common characteristics of unlabelled data [62, 63]. Clustering algorithms such as Mean-shift is commonly used with unsupervised learning models [62]. Unsupervised learning methods are widely used in applications such as feature segmentation, data pre-processing and data mining [62]. Most machine learning techniques utilized in this thesis are based on supervised learning. However, unsupervised clustering algorithms are evaluated to classify wireless data into multiple sub-classes based on input features examined in this thesis.

2.2.2.2 Regression

The regression analysis has been used in statistics and machine learning to identify the relationship between input and corresponding output values. Regression methods including linear, logistic, lasso and polynomial regression are common in regression analysis. Furthermore, regression analysis is based on supervised learning with numerical, continuous and ordinal metrics being common properties of output data used in regression analysis [61]. Various business entities including those in finance, retail, social science and marketing are heavily dependent on regression models in forecasting, predictions, sentiment and trend analysis. In this thesis, regression methods are used with the time-series data to identify relationships between different features.

2.2.2.3 Clustering

Cluster analysis is a process of grouping objects with similar characteristics [61, 63]. Cluster analysis is primarily used with unsupervised machine learning methods and can be seen in a wide range of applications, including data mining and data pre-processing [63]. In this thesis, clustering algorithms are used to classify wireless data into unknown numbers of classes based on similarities in low powered wireless attributes.

2.2.2.4 Classification

Classification is primarily used with supervised learning methods to segregate data into finite numbers of pre-determined classes [63]. As previously noted, a training data set is used by classification algorithms to determine the feature representation and its influence on prediction. Algorithms including SVM, KNN, DT, RF and NN (more details about the classification algorithms utilized in this work can be found in the Algorithms subsection) are commonly used in classification models [66]. Applications such as pattern recognition, biometric identification, image classification, natural language processing and anomaly detection rely on classification algorithms [63].

2.2.3 Algorithms

There are plenty of algorithms available to use with statistical and machine learning models. Algorithm selection is dependent on several objectives including output parameters (regression, clustering and classification), prediction accuracy, and computational complexity. Algorithms such as DT, SVM, RF, NN, Markov methods, linear and logistic regression are common among application designers and researchers [61]. Furthermore, significant progress in the field of machine learning has been observed in recent years, especially related to artificial neural networks. As a consequence, more effective algorithms are invented to construct complex deep learning models including natural language processing, speech recognition, image classification and autonomous driving.

In this work, four classification algorithms (SVM, KNN, ANN and DT), a few regression models including Linear regression, Support Vector Regression, Random Forest Regression, Decision Tree Regression and ensemble methods including Random Forest, Bagged Decision Tree, AdaBoost, Stochastic and Gradient Boosting are used to construct prediction models to identify traffic anomalies in low powered wireless networks operating in IEEE 802.15.4e/TSCH mode.

In the following, machine learning algorithms used in this work are briefly explained.

2.2.3.1 Decision Tree (DT)

A Decision Tree (DT) is a model built in a tree structure and nodes represent a “feature evaluation process” (e.g. participant’s age group) and branches represent the outcome of the evaluation (e.g. age range 20-30) and the leaf node represents the decision, taken after evaluating all attributes (e.g. “political affiliation”) [66]. A decision tree is widely used as a visual analytical tool to describe the influence of a decision and it is a predictive modelling method used in statistics and machine learning [66]. One of the critical tasks in building DT models is to identify sequential feature lists to construct the tree. It is important to select attributes with higher information gain (low entropy) on the top layer of the tree to obtain a faster and accurate decision-making process [66].

2.2.3.2 Neural Networks (NN)

The concept of artificial neural network (ANN) is inspired by the complex biological neural network and artificial nodes known as “neurons (neurodes)” and is interconnected to create a network that resembles a biological neural network [61, 64]. Those nodes contain adaptive weights that can be refined by a learning algorithm. Artificial neural networks introduce a concept called layers [64]. NNs can increase in complexity by adding more layers to the network. In the simplest form of NN, such as feed-forward neural networks, information moves only in one direction [67]. More complex artificial neural networks, such as recurrent neural networks, convolutional networks can be used for complex machine learning tasks. Deep learning algorithms extend the complexity of NN by including additional features such as multiple layers, high dimensional feature representation and complex relationships between nodes [67]. However, processing-time due to complex computations, lack of sufficient training data and over-fitting, caused by excessive layers and nodes, are a common problem with deep learning algorithms [61]. Different generalization techniques such as principal components analysis, feature engineering, cross-validation, pre-processing and regularization can mitigate some of those problems [61].

2.2.3.3 Support Vector Machine (SVM)

Support Vector Machines (SVM) are classification algorithms primarily used in a supervised learning model [65]. Although SVMs are considered as a binary non-probabilistic linear classifier, SVMs are able to perform non-linear classification, by use of higher dimensional feature spaces [65]. In a nutshell, SVMs create a hyper-plane or multiple hyper-planes, in high-dimensional space, to classify data. The optimal classification process is achieved by maximizing the distance between the hyper-plane and the nearest training data points (also known as support vectors) of each adjacent class [65].

2.2.3.4 K-Nearest Neighbour (KNN)

K-Nearest Neighbours is considered as one of the simplest classification algorithms. The classification is based on identifying classes of k nearest neighbours (selection of k value is implementation-specific). The distance can be calculated using different techniques including Euclidean distance. The prediction is determined by identifying the mode (highest frequency) of the k-nearest neighbour classes. One of the main disadvantages of KNN is that no learning function can be derived from a training set. In essence, a whole data set is required to determine the outcome of new input data, which leads to slower prediction processing with larger data sets. However, with smaller data sets, KNN is able to perform significantly faster (training and test process) than some other classification algorithms.

2.2.3.5 Ensemble Methods

Ensemble techniques are based on combining multiple models to improve the prediction accuracy. The ensemble methods are not bound to a particular algorithm and a few different techniques including voting and averaging are used by ensemble methods to determine the possible outcome. Each individual model can be selected using several criteria. For instance, each individual model could be trained with a different algorithm or with a sub-set of input features or with a sub-set of data. Furthermore, different stacking mechanisms such as bagging and boosting are used to pipeline multiple models to improve results. A Random Forest is based on an ensemble technique and it uses multiple DTs to elevate the prediction accuracy.

2.3 Security Threats in Low Powered Wireless Networks

Security requirements for low powered wireless sensor networks are more complex than the traditional wireless data networks. In conventional settings, the majority of nodes act as “clients” and a common set of security controls can be applied. However, in low powered wireless networks, each node may provide a different type of service, which may require different security enforcement. It is crucial to understand the potential threats faced by low powered wireless nodes.

In the following section, wireless protocol vulnerabilities and wireless security threats related to media access and the physical layer of LoWSNs are analyzed.

2.3.1 Vulnerabilities of Wireless Networks

IEEE 802.11 is a set of specifications for MAC and the physical layer of OSI reference model, to implement wireless communication over several frequencies including 900MHz, 2.4 GHz, 3.6 GHz, 5 GHz and 60 GHz [21]. Most wireless devices operate in 900 MHz, 2.4 GHz and 5 GHz frequencies, which is also known as WIFI. Three types of frames, the management, control and data frames, can be seen in most IEEE open standard networks [22]. Management frames establish and maintain the communication with the AP with several types of management frames, including authentication, de-authentication, association request/response, re-association request/response, disassociation, beacon and probe request/response. A few types of management frames maintain a wireless network. For instance, beacons advertise an AP’s presence and capabilities [21]. Unauthenticated stations broadcast probe requests broadcast to obtain information on the available Access Points. Control frames manage the access to a wireless medium and control the communication between wireless nodes. Control frames utilize a few frame types, including Request to Send (RTS), Clear to Send (CTS), Acknowledgement and Power Save, to control the communication. Data frames transmit actual data between nodes. All frame types include a header, a frame body and a Frame Check Sequence (FCS). The frame body carries actual data and in secured wireless networks, it may be encrypted. However, in most implementations, management frames and control frames are not encrypted [22].

2.3.1.1 Protocols

There are several known attacks related to security protocols associated with open wireless standards. A significant portion of low powered wireless sensor networks are operating on IEEE 802.11 and 802.15.4 standards and it is important to understand some of the known attacks associated with those standards. Attacks on 802.11 networks can be classified into several groups, based on the nature of the attacks, namely, key retrieving attacks, availability attacks, keystream retrieving attacks and man in the middle (MiTM) attacks [22]. Certain MiTM variance can be detected by identifying traffic anomalies of corresponding network.

Key Retrieving Attacks

Key retrieving attacks were common in early versions of 802.11 [23], especially, when the networks were protected using WEP methods. Intruders attempt to intercept specific packets carrying key information. Most key retrieving attacks are based on passive mode, therefore protection mechanisms such as firewalls and intrusion detection systems (IDS) may be unable to detect them. The FMS (Fluhrer, Mantin and Shami) attack [27] was one of the first successful attacks on WEP. The attack was based on the weakness of IVs (Initialization Vector) [28, 29]. It takes advantage of the key scheduling algorithm used in RC4. The KoreK attack [30] was based on the fundamentals of FMS; however, it performs seventeen different attacks to retrieve a probable set of keys. KoreK is considered as more effective; however, both those attacks are based on statistical analysis and require the injection of additional packets into the network to improve the efficiency. Arp injection is used by attackers to generate large number of IVs, which can be used by key cracking tools [31]. Dictionary attacks are types of brute force attacks; they can be used as a key retrieval mechanism. This technique is considered an ineffective method in cracking WEP keys; however, dictionary attacks are mainly used to extract weak WPA keys [32].

Availability Attacks

Availability related attacks are also known as DOS attacks and they are common to most versions of the IEEE 802.11 protocol family [22, 25]. Attackers attempt to exhaust network resources to create a denial of service. Since management frames are sent unprotected, most DOS attacks on 802.11 networks are based on a broadcast of forged management frames [22]. However, lack of adequate physical security controls could also lead to DOS attacks. For instance, vandalism, natural disasters and unintentional accidents could disrupt the availability of low powered sensor data. De-authentication attack is one of the most common DOS attacks on 802.11 networks [22]. The attacker monitors the wireless traffic for MAC addresses of client stations, which can be found in unprotected management frames and send forged de-authentication messages to clients on behalf of the AP. However, it is also possible to send a forged de-authentication message on behalf of the client to AP. As a consequence, the client station has to re-authenticate with the AP before resuming the communication. Disassociation attacks [22, 23, 33] are similar to de-authentication attacks and they utilize disassociation messages instead of de-authentication messages. The block acknowledgment flood attack takes advantage of Add Block Acknowledgment (ADDBA) and was introduced in 802.1n protocol [22]. ADDBA allows a client to send a larger block of data without fragmentation. However, an attacker could send an ADDBA request on behalf of a client, which negotiates the block size and the sequence numbers, associated with those blocks. Subsequently, AP accepts only those blocks with the corresponding sequence numbers and legitimate traffic from the victim will be dropped [22]. The authentication request flooding attack is based on flooding the client association table with fake authentication requests and eventually, the AP will not be able to respond to legitimate authentication requests, in a timely manner [22]. Request to Send (RTS) and Clear to Send (CTS) messages are optionally used by 802.11 networks to control access to the wireless medium [21]. A CTS flooding attack is based on continuous flooding of CTS frames to itself or another client and forces the rest of the clients to wait for transmission. A RTS Flooding attack, floods RTS frames with large transmission duration windows, causing other clients to back off from transmitting [22, 33]. Beacon flood attacks are based on advertising the sequence of fake Extended Service Set Identification (ESSIDs) to overflow the list of available networks [22]. Probe requests are used by clients to search for available wireless networks. APs are obliged to respond with a probe response message. The attacker could send a sequence of fake probe requests to overwhelm the AP and cause an attack known as a probe request flooding attack [34]. A probe response flooding attack is also a common DOS attack on 802.11 networks [34]. The attacker replies to probe request messages acting as a valid AP.

Man in the Middle Attacks

Man in the middle attacks (MiTM) are based on impersonation techniques. For instance, Honeypots are created by security admins to attract attackers and redirect their attention from legitimate targets. However, intruders use the same technique to create malicious wireless networks in order to attract users. Using MiTM attacks, adversaries may be able to monitor an entire communication, including application-level data, such as passwords and personal information. However, if the communication is secured using an upper layer control such as Secure Sockets Layer (SSL), an attacker still could launch a replay attack by retransmitting excessive amount of captured wireless packets to create havoc. Evil Twin is also a different variant of Honeypot and advertises an AP with the same network name (SSID) to mislead legitimate clients [22]. Rogue access points are unauthorized APs deployed by intruders or undisciplined users. This may open the door for unauthorized access to the secure network [35].

There are several recommendations and work that have been done to prevent some of the aforementioned attacks [24, 36]. However, most proposed solutions are based on modification of existing standards, which may lead to inconsistency with open standards. Some studies have developed external systems, such as Intrusion Detection System (IDS) to detect attacks related to wireless networks [37]. Ferreri et al. [34] demonstrate how easy it is to launch a DOS attack on 802.11, using Authentication Request Flood (ARF), Association Request Flood (ASRF) and Probe Request Flood (PRF). Wang et al. [25] discuss a DOS attack on 802.11i, using a 4-way handshake protocol [38] and a possible solution, based on 3-way handshake mechanism, using authenticated management frames. Aslam et al. [39] also propose a solution to disassociation DOS attacks using authenticated management messages. However, both those solutions require modification to the firmware of the wireless interface card. Afzal et al. [40] suggest a method to mitigate de-authentication attacks and Evil Twin Attacks, using a signature-based Intrusion Detection System. Research work in [41, 42, 43, 44, 45] also proposes intrusion detection systems, based on different techniques, to prevent DOS attacks on wireless networks. Detection of de-authentication based DOS attacks and a prevention mechanism using intrusion prevention mechanism is discussed in [46]. Previous work improvements [46, 47] implement a machine learning technique to detect DOS based de-authentication attacks. Research work in [48] discusses a specification-based intrusion detection mechanism, which uses both signatures and anomalies to detect attacks on Ad-Hoc networks. Most proposed solutions to mitigate attacks on wireless networks are based on protocol modification, firmware upgrades or via a middleware solution.

2.3.1.2 Physical Layer

Despite the fact that 802.11 is the de facto standard for conventional wireless Local Area Networks, 802.15.4 has been the dominant wireless technology used in low powered wireless sensor networks. 802.15 is a family of standards drafted by the IEEE to accommodate several wireless network requirements. The IEEE 802.15.4 standard was drafted in 2003 and it defines the MAC and Physical layer specifications for LRWPAN [8]. Several amendments were drafted by the IEEE to address concerns related to low powered wireless networks. For instance, IEEE 802.15.4e was drafted to support emerging industrial markets for low rate, wireless devices and to utilize available frequencies effectively, with the use of channel hopping and a sliced-time based media access mechanism [10]. However, most wireless networks based on open standards are operated in ISM and unlicensed bands and interference from other devices including ISM cannot be fully mitigated. Furthermore, physical threats such as vandalism, natural disasters, operational negligence and hardware failures can impact any wireless network including networks operating in 802.15 family protocols. However, a few specific threats related to Physical and MAC layer of IEEE 802.15.4 family networks are discussed in the following segment.

IEEE 802.15.4 (LRWPAN) networks inherit most traditional threats originated in any network. However, it also comprises additional threats, due to characteristics of low powered networks, such as resource-constrained nodes, the broadcast nature of the wireless medium, dynamic topology, physical exposure, lack of physical safe guards and the scale of the network [50]. Furthermore, low powered wireless nodes may be in different states, such as sleep, low power, active, power off and transmit only and this further complicates the definition of network boundaries. In contrast to conventional networks, low powered wireless networks could contain hundreds, if not, thousands of nodes deployed in a single network [51], which lead to more complex network topologies and logical routing structures. Several studies outline threats related to the Physical and MAC layer. Radio jamming is a physical layer attack to cause a denial of service by creating interference on the wireless medium [26]. There are several variants of radio jamming attacks, such as constant jamming, deceptive jamming, random jamming and reactive jamming [52]. An attacker may utilize channel hopping with a single-channel pulse jamming to create interference on an entire range of channels [52]. On the contrary, pulse-band denial targets a particular channel. Intruders can prevent a legitimate node selecting a particular channel, by Jamming a target channel, during the channel energy detection (ED) process. (A channel energy detection process is used by wireless nodes to determine "less utilized" channel, during the auto channel selection process [53]). In activity jamming, the intruder utilizes the Radio Signal Strength Indicator (RSSI) and Clear Channel Assessment (CCA) information to determine network activities [55]. There are a few non-jamming type attacks that can be seen in 802.15.4 networks. For instance, the attacker contaminates the wireless frame by modifying the content, [56] known as message manipulation attack.

2.3.1.3 Media Access Control Layer

Similar to radio jamming attacks on the physical layer, link layer jamming techniques are used in the MAC layer leads to DOS attacks. However, link layer jamming utilizes packets rather than the signals used in radio jamming attacks [57]. Node-specific flooding depletes the power source of a specific target by sending excessive amounts of random data [58]. The IEEE 802.15.4 network uses a carrier sense multiple access with collision avoidance (CSMA/CA) [21] mechanism to access the medium during the contention access period (CAP) [8]. A back-off period (BOP) is used by nodes in CSMA/CA networks to avoid collisions and a random BOP value is selected by the wireless node [8]. The intruder can increase the waiting time for other nodes by selecting a shorter back-off time [58]. A battery life extension (BLE) mode, used by low powered nodes, achieves a smaller initial contention window time. (Initial contention window time defines the range of values for the contention period). However, the intruder could abuse this process by initiating a BLE mode to get a shorter back-off period [58]. The intruder can manipulate this process by reducing the number of back-off periods to one (clear channel assessment reduction attack) or completely omitting the CCA process (clear channel assessment omission attack). Acknowledgment packets are used in 802.15.4 and other networks to control the packet transmission. An adversary can launch an ACK spoofing attack by replying to the sender, using a forged ACK with expected sequence number, on behalf of a legitimate receiver [52, 53]. ACK dropping is another variant of ACK attacks and the intruder uses an intelligent link-layer technique to drop all ACK packets, wasting both the power and the bandwidth of low powered networks [58]. In beacon-enabled networks, the PAN coordinator utilizes Guaranteed Time Slots (GTS) to allocate time for each node, during the Contention Free Time (CFT) period [57]. The intruder can launch a GTS attack by abusing the GTS process [57]. Denial of service against data transmission, during the Contention Free Period (CFP), can be launched by sending de-allocation requests on behalf of legitimate nodes. False data injection attack, during the CFP, can be launched by sending a GTS slot allocation request on behalf of newly associated nodes [59]. Furthermore, the intruder can launch a DOS attack against GTS requests by sending GTS requests using their own identity, instead of using a spoofed ID (false data injection). The intruder can also launch an attack by corrupting GTS slot data [57]. CAP maintenance is used by 802.15.4 networks to balance the CAP and the contention-free access period (GTS) [8]. However, the intruder can force a reduction in the allocated CAP by sending large amounts of GTS requests [59]. 802.15.4 defines a conflict resolution procedure for multiple PAN coordinators in a single network [8]. The Intruder can abuse the conflict resolution procedure by sending multiple PANId conflict notifications, which causes a legitimate PAN coordinator to delay servicing the other nodes [54].

Chapter 3. Motivation

As previously mentioned, the unexpected growth in LoWSN in recent years has led to a massive and heterogeneous collection of low powered wireless sensor devices. Yet, as of today, there is no standardized mechanism to protect over seven billion existing low powered devices. Consequently, according to industry experts, by 2022, 50 percent of security budget will be allocated to remediate security faults in low powered devices. [111].

Most hardware manufacturers are keen to implement open communication standards to provide interoperability between different hardware. However, the selection of software, including operating systems and applications running in low powered devices may be at the discretion of the low powered device manufacturer. The complexity of operating systems and the functionality of applications are exclusively determined by the low powered device manufacturers and they are dependent on various factors, including cost, capabilities, market demand and the regulatory requirements.

Security is a complex domain and the security requirements are dependent on various factors including nature of the low powered wireless data, security objectives, regulatory and law requirements and financial constraints. Furthermore, various regulatory entities such as PIPEDA [104], HIPPA [105], the European Union Directive on the protection of personal data [106] and Electronic Communications Privacy Act [107] outlined the privacy protection laws.

Some of the previous research has demonstrated remarkable results. However, the majority of proposed work requires modification to low powered node's software or amendments to open communication protocols. As previously mentioned, low powered sensor devices are operated in constrained and heterogeneous operating environments and it would be a tremendous challenge to develop a universal solution adaptable to the majority of already deployed low powered devices. Furthermore, most sensor devices are operated in low powered mode, where computational, memory, storage, battery and transmission power is restricted and even implementation of a light-weight solution may affect the performance of sensor devices significantly [108].

The objectives of this study are to explore the feasibility of building a passive traffic anomaly detection mechanism for low powered wireless sensor networks, using characteristics of low powered wireless networks and attributes of 802.15.4e/TSCH, with the use of machine learning techniques. Furthermore, an important goal of this work is to investigate the possibility of designing a universal security solution without compromising the performance of low powered sensor devices and enable them to operate in a heterogeneous environment where, low powered nodes may operate in proprietary hardware and software, interconnected with standardized (open) communication protocols. (Implementation specific details can be found in the Deployment Scenario subsection in the Summary and Discussion section)

Chapter 4. Hypothesis

This thesis is based on an experimental study, using live and simulated data, to determine the relationship between performance factors including prediction accuracy, false positive and negative rates of a traffic anomaly detection model based on characteristics of low powered wireless networks and attributes of communication protocols as input features and various factors including classification algorithm, training-set size, number of nodes, network segmentation, noise influence, retention factor, input-feature segmentation, resource utilization, model aggregation and environmental effects. A number of low powered attributes are evaluated and a detailed description of each of these attributes as well as classification algorithms used in this thesis can be found under the Research Design and Methods chapter.

In the following, a list of research questions, investigated in this work is outlined.

Research Questions:

- What is the behaviour of prediction accuracy and false positive negative rates with regards to various factors including training set size, input parameters, data variance, number of nodes and classification algorithm?
- How consistent is the performance of a prediction model trained with low powered wireless features (retention factor)?
- Which input feature group (physical layer, low powered, IEEE 802.15.4e/TSCH and network layer) is able to provide the most significant results?
- How costly is a particular prediction model on resource utilization (memory, CPU and time)?
- Would an aggregated model based on prediction results from individual models (generated using features from individual classes) be able to improve the prediction results?
- How do environmental changes including severe weather, seasonal effects, Radio Frequency (RF) noise, location specific attributes would impact the performance of a prediction model?

A Comprehensive response to above questions can be found in the Research Method and Design chapter of this work.

In summary, the objective of this research is to investigate whether it is possible to use passively-collected information of low powered networks including the attributes of IEEE 802.15.4e/TSCH, low powered, Physical and Network layer specific to identify the operational behaviour and activities of a LoWSN. If the network activities and the behaviour can be accurately identified using low powered characteristics, the network activity map can be used to detect anomalies and outliers of the corresponding low powered network. This work is further enhanced by investigating the relationship between performance indicators including anomaly detection accuracy, false positive and negative rates and several factors including network size, input features, training set size, noise impact, model retention factor and classification algorithms.

Several machine learning methods are utilized in determining the network activities and the behaviour of LoWSNs and both live and simulated data are used to train classification models and to evaluate performance including prediction accuracy and the false positive/negative rates.

Chapter 5. Related Work

In the following, several interesting studies related to protection of low powered sensor data is discussed. However, most studies related to the security of LoWSNs are about protection against unauthorized disclosure of low powered wireless data. There are also some studies related to the data integrity protection, access control and anomaly detection in low powered environments. However, over 90% of previous work reviewed in this thesis, require modification of existing protocols or modification to software in low powered devices to accomplish the proposed security goals. In the following section, a few interesting works related to the protection of low powered wireless sensor data is discussed.

Das et al. [69] propose a two factor, user authentication mechanism using one-way hash function and XOR operation. Authors of the work insist that the proposed method can prevent password guessing, impersonation and replay attacks. Khan et al. [70, 71] suggest an enhancement to [69] by addressing some of the flaws related to password modification and vulnerabilities related to privileged, insider attacks. However, both the above solutions require modification to the low powered node software. A significant amount of work related to the security of LoWSNs, based on middleware solutions is available. Freitas et al. [72] propose a hybrid encryption mechanism, based on both symmetric and asymmetric keys, with the use of a message authentication mechanism, to secure the sensor data. Daidone et al. [73] suggest a modular middleware solution to guarantee the confidentiality, integrity and the authenticity of low powered sensor data. However, middleware solutions are based on implementing a software code in the application layer of each low powered system. Piro et al. [74] discuss a lightweight mechanism to negotiate link keys in 802.15.4 networks; however, low powered device software has to be modified to accomplish the key, negotiation process. [75 – 80] discuss the use of an efficient low powered security implementation in an application-specific integrated circuit (ASIC), which can be used in low powered devices. However, these types of solutions require a complete re-design of the device's hardware architecture. Hao et al. [81] propose a forecast model of a security situation in low powered wireless networks. Their approach is based on a probabilistic model (Hidden Markov model) to forecast the security posture of a given situation. Skarmeta et al. [6] discuss a decentralized mechanism, to protect data privacy, in low powered wireless networks. Their solution is based on the use of a lightweight token, to access network resources and optimized implementation of the elliptic curve algorithm is required in each node. Marin et al. [82] also, propose a solution based on Elliptic Curve (ECC) for low powered networks. Zhang et al. [83] discuss a peer-to-peer security validation mechanism to protect low powered wireless nodes. Henze et al. [11] suggest a mechanism to reinforce the security controls when data is leaving the perimeter of a low powered wireless network. Liu et al. [84] suggest an immunology based approach to secure low powered wireless networks. According to their model, security controls should be adapted to the changing threat model. However, the above solution [84] is based on node-specific threats and individual nodes should be configured to define base-line values for the normal operation.

Related Work - Anomaly Detection

Anomalies are defined as an observation that appears to be inconsistent with the remainder of the data set [85] and the process is also referred to as outlier detection or deviation detection. An anomaly may be caused not only by intrusion, but also by other phenomena such as hardware failure, battery drainage, software defects and configuration malfunction [86]. Several types of attacks could cause anomalies in low powered wireless networks namely, communication related attacks, nodes being compromised, denial of service attacks, impersonation attacks, protocol specific attacks and random failures [85, 87]. Primarily, there are two types of anomaly detection mechanisms; prior-knowledge based and prior-knowledge free [85]. Anomaly detection based on prior-knowledge, produces a normal profile based on known knowledge [85]. However, prior-knowledge free systems rely on a training procedure to formulate the normal profile [85]. Several training mechanisms, such as machine learning, data mining and graph-based detection are available to create the normal profile. The normal profile constitutes the healthy state of a system. It can be defined, using prior-knowledge or training. A point anomaly can be defined as individual data that diverges from the norm. Contextual anomalies are abnormalities of data in a particular context [88]. Collective anomalies are a group of linked data instances that differ from the normal profile [89].

There are several interesting works, related to anomaly detection, in low powered wireless sensor networks. However, most solutions require installation of additional software in individual nodes. Some interesting and comparable approaches are discussed below.

A probabilistic model to identify anomalies, using kernel density estimators, in a distributed environment is proposed by Palpanas et al. [90]. Tiwari et al. [91] discuss a combination of a probabilistic model and a rule-based scheme to detect certain types of anomalies such as black-hole and selective forwarding attacks. Rajasegarar et al. [92] introduce an anomaly detection mechanism in a distributed environment, with the use of a k-mean clustering algorithm. A multi-agent based mutual detection scheme, where the network forms multiple clusters using neural networks and a k-means clustering algorithm is discussed by Wang et al. [93]. Agah et al. [94] suggest a model based on game theory to find anomalies in LoWSNs. A decentralized rule-based approach is suggested by Silva et al. [95] with rules created based on the application and security requirement of the low powered nodes. A multi-hop acknowledgment mechanism is proposed by Yu et al. [96] to actively detect the packet forwarding path, from the source node to the base station, using the ACK packets. A one-way hash chain, pre-shared encryption key and message authentication code (MAC) are used with this proposed scheme. Ho et al. [97] developed a mechanism to protect against replica, node attacks. Low powered nodes are grouped and labelled (Group Deployment Point (GDP)) based on the location, during the deployment of the LoWSN. When a packet is received by a low powered node, it validates the location of the sending node, using a GDP for anomalies. Onat et al. [98] propose a scheme to detect anomalies using statistical measures of the packet arrival process. Each low powered node maintains a normal traffic profile for each neighbouring node and also, keeps two buffers for receiving data and intrusion data. These two buffers are analysed using descriptive analytic tools (Mean, Standard Deviation) to determine anomalies. A malicious node detection scheme, based on auto-regression (AR) model is discussed by Curiac et al. in [99]. This model uses time-series data to determine malicious activities. A hop count mechanism is used by Dallas et al. [100] to detect sinkhole attacks. Each node is responsible for examining the hop count of the packets passing through and each node and compares the hop count with the information provided by distance vector routing protocol and the base-station. The base-station is responsible for forming the network and maintaining the routing information by periodic broadcasts. Authors of [101] propose a method to detect anomalies using a clustering technique. Xiao et al. [102] discuss a machine learning based anomaly detection scheme with the use of Bayesian classification algorithm. In their work, throughput, packet loss rate, and the packet average delay of LoWSN are selected as the feature vector for the machine learning algorithm. Although the packet loss rate and throughput are not utilized in our work to detect anomalies, some similar objectives can be observed in this work. Authors in this work [102] were able to detect traffic anomalies with over 90% accuracy and to reduce the false-positive rates below six percent. A simulated environment (NS2) with 100 low powered nodes is used in the above work [102] to generate experiment data and details about anomalous data creation process is not discussed. Hence, we are unable to conduct a direct comparison with Xiao’s work. An anomaly detection mechanism to detect outliers, using the distance between current measurement and its neighbours, is proposed by Abid et al. [103]. Authors utilize k-nearest neighbours (KNN) for classifying low powered nodes.

Chapter 6. Research Methods & Design

Low powered wireless sensor networks create more deterministic behaviour patterns compared to conventional wireless networks. A few examples of deterministic behaviour of low powered network activities are listed below.

- A low powered wearable sensor, attached to a patient, may be configured to transmit a value of physiological aspect measured in every predetermined time interval or when a threshold value is reached
- In an industrial setting, a low powered sensor would measure the internal temperature of complex machinery and be configured to send an alert when a threshold value is reached
- The sensor battery level decreases after each transmission and computation, until minimum operable value is reached
- In a smart-home sensor network, the motion sensor, closed to the main entrance, may be activated immediately, after the door is opened
- In process automation, the sensor dispatches a notification when a certain process is completed
- In a smart environment, the actuator may be notified to activate the cooling system when the control system receives an alert from the temperature sensor
- A sensor device may send periodic updates to a specific service port, located in a unique destination
- Sensor notification messages may contain payload values that can be grouped into a finite number of classes

These behaviour patterns can be used to identify a finite number of contexts for a low powered wireless network. Subsequently, context data can determine acceptable baseline values for normal operation and detect outlier/anomalies of the corresponding low powered wireless network. Different approaches including rule based decision mechanism and machine learning prediction model can be used to identify the acceptable behaviour (baseline values) of a particular LoWSN. However, in this thesis, several machine learning techniques including classification, regression, clustering and ensemble methods are evaluated to construct a most effective model to detect traffic anomalies in LoWSNs.

Both live and simulated experimental data are used in this work and details, with respect to, the research questions and methods are described in the following section:

6.1 Research Model

This study is based on a quantitative research model and the following attributes are tested against research questions; values of the attributes could be categorical or continuous.

- Time slot number
- Timestamp
- Battery power level
- Source address
- Destination address
- Service identifier
- Sequence number
- Packet payload size
- Signal strength (RSSI)
- Link quality indicator (LQI)
- Link distance (LD)
- Signal-noise ratio (SNR)

Hereafter, the above list of attributes is denoted by Attributes $^{[\text{low power}]}$

Attributes$^{[\text{low power}]}$ = { Time slot number, Timestamp, Battery power level, Source address, Destination address, Service identifier, Sequence number, Packet payload size, Signal strength (RSSI), Link quality indicator (LQI), Link distance (LD), Signal-noise ratio (SNR) }

An important characteristic of Attributes$^{[\text{low power}]}$ is that, the data associated with those attributes can be collected passively, without implementing any additional software in individual nodes or introducing any extra overhead into the packet payload.

The following is a brief description of each input feature used in this work.

Time slot number

Wireless networks operating in IEEE 802.15.4e/TSCH mode rely on a synchronization mechanism initiated by PAN coordinator [8]. The PAN coordinator is responsible for the formation and maintaining the network operation in TSCH mode. The PAN coordinator utilizes a unique value known as the Absolute Slot Number (ASN) to assign a time slot and a channel number for a particular communication. ASN is a 5-byte number initiated during the formation of the network and it maintains the uniqueness for a longer period of time. ASN is distributed by PAN coordinator using Enhanced Beacons (EB) and it can be easily accessible by extracting payload portion of Information Elements (IE) in the EB header [8].

Battery power level

Battery level indicates the amount of battery power available for a particular low powered node. In low powered environments, devices are expected to operate in a prolonged period, without replacing the power source. However, the battery level of a wireless node decrease, in relation to the computational and data exchange operations, carried out by the individual low powered wireless node. In most environments, certain instructions and operational procedures may be available for the battery replacement process and consequently, a strong negative relationship should be found between battery level and the time.

Timestamp

The timestamp includes actual time for various operations, such as packet-creation time, transmission time, received time and acknowledged time. Timestamps of a single communication (such as packet creation time, transmission time, received time and acknowledged time) produce unique characteristics of a particular data path such as transmission delays, congestion, retransmission, and send-receive relationship. Furthermore, the timestamp can be used as an independent variable in time-series and regression analysis.

Source Address

The source address is used by the source node to uniquely, identify itself in data exchange. The low powered wireless sensor networks could utilize different mechanisms including MAC address, IPv4, IPv6 and labels to uniquely identify low powered nodes. The source address is a mandatory attribute for a legitimate communication and is primarily used to identify the origin of a communication [109]. However, in this work, the IPv6 addressing scheme is used in related experiments.

Destination Address

Similar to the source address, the destination address is used by the destination node to identify itself in a wireless network. During a unicast communication, the destination address is a mandatory attribute [109]. However, certain packets such as EB utilize broadcast address to deliver network schedule to all nodes within the broadcast domain [109].

Service Identifier

The service identifier is used by a destination node to deliver messages to a proper destination (application). For instance, the destination node may provide multiple services, such as a service that collects alerts, summarizes data, logs data and applies specific sensor data. For instance, the Transport layer of IPv4 protocol stack, utilize a 16-bit number (also known as port number) to uniquely identify different services (applications) [110].

Sequence number

Sequence numbers are used in different layers to manage the communication between nodes in each layer. Sequence numbers are increased sequentially (in most low powered implementations) and they are also used to associate messages to corresponding acknowledgments). In this work, sequence numbers used in the MAC layer and network layer are examined.

Packet Payload

The packet payload carries data between nodes; it could contain control information, management information or actual low powered sensor data [109, 110]. Data generated by individual nodes are pre-processed due to constrained resources in local nodes and consequently, a repeatable set of messages belonging to individual nodes should be identified. Although, payload data encryption is a common practise in most network environments, in low powered networks, implementation of complex security mechanisms is uncommon [1].

Signal Strength (RSSI)

Received Signal Strength Indication (RSSI) of neighbouring nodes is dynamically determined by individual nodes using control messages [21]. These messages are used by management modules including routing process to build a logical topology of a wireless network. Different factors such as obstructions and mobility of individual nodes may impact the RSSI values of corresponding nodes and potential RSSI fluctuation patterns can be identified by observing signal strengths of neighbouring nodes for longer periods.

Link quality indicator (LQI)

Link quality of neighbouring nodes is determined by each node using several factors including RSSI, distance, interference (RF), packet loss and transmission rates. However, LQI computation is hardware-specific and different LQI reading can be observed by nodes running different wireless interface adapters.

Link distance (LDI)

Link distance between nodes is determined using control messages and they approximate the distance between nodes. LDI has a different significance in different protocols. For instance, in conventional WIFI networks, the average distance between nodes could be significantly larger than that of in LoWSNs. However, fluctuation of LDI in a particular wireless network defines characteristics of individual low powered nodes.

Signal-to-noise ratio (SNR)

Signal to noise ratio is a combination of multiple attributes including RSSI and the noise impact. RSSI is primarily determined by the transmission and receive sensitivity of radio transceivers and obstructions such as walls and trees. However, noise can be caused by several factors including interference, multipath, obstructions and faulty nodes. Noise can also be time-specific, location-specific or event-specific. However, by observing SNR readings for a longer period, a SNR value map for individual nodes can be obtained.

6.2 Design

6.2.1 Assumptions

The following experiments are based on a number of assumptions. However, the experimental work is divided into four separate groups based on distinctive characteristics of input features and each experimental group is based on an independent set of features. Furthermore, the data used in this work is collected from several environments and consequently, a different set of assumptions has to be made for each experimental group.

Classification models based on machine learning are heavily relied on number of samples used in the training and validation process. However, it is equally important to have balanced number of samples to represent each label class. Unfortunately, it is a known challenge to collect adequate amount of anomalous data in operational environments. Therefore, most anomaly detection solutions are relied on a manually engineered mechanism to generated anomalous data. Anomaly data used in this thesis are created using three different techniques (more details about anomaly data creation techniques can be found in the Experimental Settings section). Those techniques are utilized in such a way that they are able to simulate various attacks and other traffic anomalies in a reasonable manner. For instance, spoofing attack is a form of attack where a node is trying to identify itself as a different node. Replay attack is a situation where previous transmission (packet) is retransmitted with or without modifying the payload with a malicious intention. Defective node or higher noise in the wireless medium could also lead to higher packet loss, transmission delays or retransmission of packets which could also create traffic anomalies. These attacks can be detected by various methods including analysing the data path, timestamps, timeslot value (ASN), source identity (MAC, IP) and node specific attributes including battery value, signal strength (RSSI), link distance, RF noise (SNR). In this thesis, more than 12 different parameters including the above-mentioned attributes are carefully examined to simulate wide range of potential threats.

6.2.2 General Settings

6.2.2.1 Hardware & Software Settings

Hardware	Intel 5i-2400, 4-core, 3.1 GHz, 8 GB ram
Operating System	Ubuntu 16.04 LTS
Applications	OpenWSN 1.8.0, Python 3.5.1, Scikit-learn 0.18.1, Pandas 0.18.1, Wireshark 2.2.4

Tab. 1. Hardware and Software Settings

6.2.2.2 Simulator Settings

Adaptation	IETF 6LoWPAN
IP/routing	RPL
MAC	IEEE802.15.4e

Physical	IEEE802.15.4-2006
Application	COAP

Tab. 2. Simulation settings

OpenVisualizer component provided by OpenWSN is used to generate low powered nodes operating in TSCH mode in a simulated environment. Network topology is manually created and the link quality and packet drop rate (PDR) are manually adjusted to simulate a realistic network environment. An unofficial draft of 6TiSCH, implemented by OpenWSN is used to provide IPv6 support for IEEE 802.15.4e/TSCH network. Furthermore, OpenWSN also provides an unsophisticated implementation of a TSCH scheduling mechanism which provides a simple time allocation mechanism for each node. In the OpenWSN simulation environment, PAN coordinator and the root node for RPL based routing process is manually selected.

RPL generates a routing structure based on a rank-based mechanism. Once a suitable root node is selected, RPL initiates the route formation process by generating Destination-Oriented Directed Acyclic Graph (DODAG) for each node. Packet capturing process (Wireshark) is manually started, when the routing convergence process is completed. Third-party dissector has been used by the Wireshark to identify wireless packets operating in TSCH mode. Captured data is stored in the packet capture format (*.pcap). In this work, captured data is converted to JSON format for further processing.

Captured IEEE 802.15.4e/TSCH data is pre-processed and formatted to use with the machine learning methods. A small portion of captured data is labelled as ‘error/malformed’ by capturing software due to unknown reason. However, malformed data is not removed from the training and testing data sets to maintain consistent results. More details about data extraction, individual attributes, numerical conversion and data normalization can be found in each experimental-group results section.

This thesis is based on an experimental study and the objective is to provide a comprehensive comparison of various performance indicators of traffic anomaly detection models against several critical factors. More than 10 input features, four classification algorithms, three performance indicators including prediction accuracy, false positive and negative rates are compared against over 12 factors including training set size, classification algorithm, noise influence, retention factor, number of nodes, label ratio, unseen data, threshold values, model aggregation, network segmentation, resource utilization and seasonal effects. Though, different techniques such as two-dimension (2D) & three-dimension (3D) diagrams, tables, multi-graphs and graph-summarization were utilized to minimize the number of diagrams and pages, we still had to use 124 diagrams to interpret the important findings of this work. The details are provided in order to allow for experimental replication.

6.2.2.3 Machine Learning Settings

In this thesis, several machine learning methods are used to learn the normal behaviour of low powered networks. The four following machine learning algorithms are used to build classification models.

- Support Vector Machines (SVM)
- K-Nearest Neighbours (KNN)
- Neural Networks (NN)
- Decision Tree (DT)

Several factors have contributed to the selection of those four algorithms including the diversity in optimization and computational costs. SVM is based on identifying decision boundary using support vectors and optimization is based on maximizing the distance between support vectors. In this work, SVM utilizes a linear kernel and it attempts to separate the data linearly with the cost of accuracy. However, KNN doesn’t utilize any optimization method and algorithm is based on identifying K nearest neighbours using Euclidean distance. Furthermore, KNN does not rely on prediction model specific parameters like other three selected algorithms and that could lead to different behaviour in resource utilization. NN provides a complex and higher accuracy with the cost of resources. DT utilizes simple statistical methods such as GINI Index or cross entropy to build a decision tree. Decision Tree is considered as a fast non-linear model with not so optimal accuracy with larger feature vector. Throughout this work, these four classification methods are also referred to as default classifiers. Besides the four default classifiers, other techniques, such as ensemble methods, are evaluated to improve the prediction accuracy and other performance indicators. Furthermore, linear regression is used in regression analysis with time-series data.

Default algorithm-specific parameters provided by the Sci-kit Learn libraries has been used by the classifiers during the training process. Some of the important parameters used by each classification algorithm are listed in the following table.

SVM	Kernel:linear, degree:3, regularization:1, tolerance:1e-3, prob:false
KNN	n_neighbours:5, algorithm:auto, leaf_size:30, power:2
NN	Hidden_layer_size:100,activation:relu, solver:adam,lr:0.001,n_layers:2

DT	Criterion:gini, splitter:best, min_split:2, min_sample_leafs:1

Tab. 3. Classification Algorithm Settings

In machine learning, various methods and techniques are utilized to obtain the most effective models. Principle Component Analysis (PCA) is used in machine learning to determine the most effective input feature vector. Cross validation is used to discover the most effective model by comparing different models and fine-tuning the hyper-parameters associated with individual classifiers. Receiver Operation Characteristics (ROC) and Area Under the Curve (AUC) are used in machine learning to understand the behaviour of true positives against false positives and primarily used to discover the most effective model by regulating various positive rates parameters. However, the objective of this study is to compare the performance (prediction accuracy, false positive/negative rates) against various factors and finding an optimal prediction model would hinder the primary objectives of this thesis. Therefore, utilizing model optimization tools as well as manual adjustment of hyper-parameters of individual classifiers is intentionally avoided.

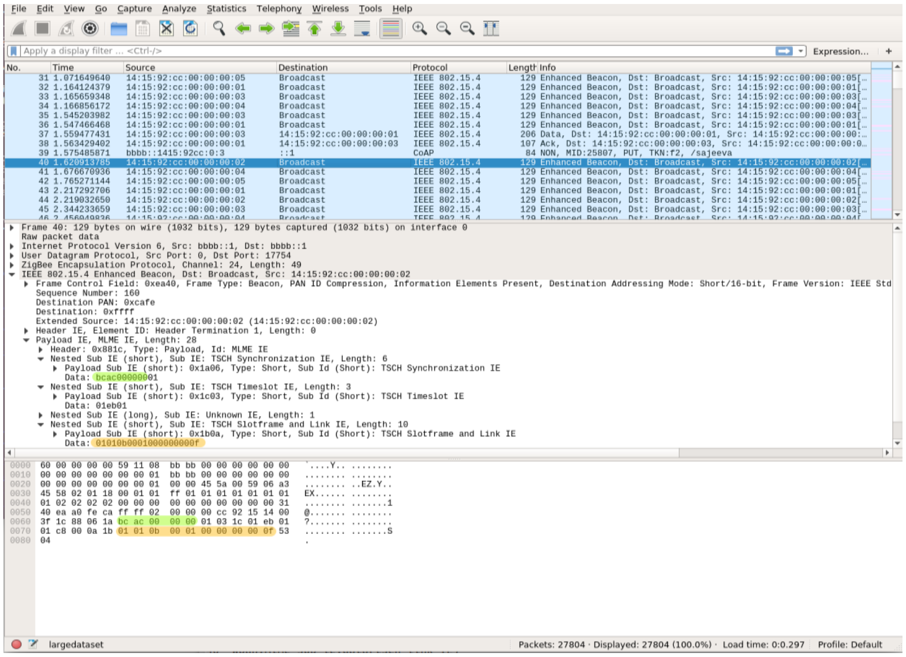
6.3 Experimental Settings

6.3.1 IEEE 802.15.4e/TSCH (Protocol Specific) Characteristics

6.3.1.1 Background

Absolute Slot Number (ASN) is used by wireless networks operating in IEEE 802.15.4e TSCH mode to uniquely identify timeslot used by a particular packet. (Detailed information about IEEE 802.15.4e and supported modes can be found in the IEEE 802.15.4e subsection in the Background section). Absolute Slot Number can be found in Information Elements (IE) located in the payload section of the

802.15.4e Enhanced Beacon (EB). Information Elements are used by EB to relay network management information, including the communication schedule to wireless nodes. IE consist of keys and corresponding values and ASN can be found in an IE belonging to the payload section of EB. The following diagram highlights ASN (green) and schedule data (yellow) extracted from an EB.



```
No.      Time         Source                    Destination               Protocol      Length Info
     31 1.071649640  14:15:92:cc:00:00:00:05   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:05[…
     32 1.164124379  14:15:92:cc:00:00:00:01   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:01[…
     33 1.165659348  14:15:92:cc:00:00:00:03   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:03[…
     34 1.166856172  14:15:92:cc:00:00:00:04   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:04[…
     35 1.545203982  14:15:92:cc:00:00:00:03   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:03[…
     36 1.547466468  14:15:92:cc:00:00:00:01   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:01[…
     37 1.559477431  14:15:92:cc:00:00:00:03   14:15:92:cc:00:00:00:01   IEEE 802.15.4   206 Data, Dst: 14:15:92:cc:00:00:00:01, Src: 14:15:92:cc:00:00:00:…
     38 1.563429402  14:15:92:cc:00:00:00:01   14:15:92:cc:00:00:00:03   IEEE 802.15.4   107 Ack, Dst: 14:15:92:cc:00:00:00:03, Src: 14:15:92:cc:00:00:00:0…
     39 1.575485871  bbbb::1415:92cc:0:3       ::1                       CoAP             84 NON, MID:25807, PUT, TKN:f2, /sajeeva
     40 1.620913785  14:15:92:cc:00:00:00:02   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:02[…
     41 1.676670936  14:15:92:cc:00:00:00:04   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:04[…
     42 1.765271144  14:15:92:cc:00:00:00:05   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:05[…
     43 2.217292706  14:15:92:cc:00:00:00:01   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:01[…
     44 2.219032650  14:15:92:cc:00:00:00:02   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:02[…
     45 2.344233659  14:15:92:cc:00:00:00:03   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:03[…
     46 2.456049836  14:15:92:cc:00:00:00:04   Broadcast                 IEEE 802.15.4   129 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:04[
```

```
▸ Frame 40: 129 bytes on wire (1032 bits), 129 bytes captured (1032 bits) on interface 0
  Raw packet data
▸ Internet Protocol Version 6, Src: bbbb::1, Dst: bbbb::1
▸ User Datagram Protocol, Src Port: 0, Dst Port: 17754
▸ ZigBee Encapsulation Protocol, Channel: 24, Length: 49
▾ IEEE 802.15.4 Enhanced Beacon, Dst: Broadcast, Src: 14:15:92:cc:00:00:00:02
  ▸ Frame Control Field: 0xea40, Frame Type: Beacon, PAN ID Compression, Information Elements Present, Destination Addressing Mode: Short/16-bit, Frame Version: IEEE Std
    Sequence Number: 160
    Destination PAN: 0xcafe
    Destination: 0xffff
    Extended Source: 14:15:92:cc:00:00:00:02 (14:15:92:cc:00:00:00:02)
  ▸ Header IE, Element ID: Header Termination 1, Length: 0
  ▾ Payload IE, MLME IE, Length: 28
    ▸ Header: 0x881c, Type: Payload, Id: MLME IE
    ▾ Nested Sub IE (short), Sub IE: TSCH Synchronization IE, Length: 6
      ▸ Payload Sub IE (short): 0x1a06, Type: Short, Sub Id (Short): TSCH Synchronization IE
        Data: bcac0000000001
    ▾ Nested Sub IE (short), Sub IE: TSCH Timeslot IE, Length: 3
      ▸ Payload Sub IE (short): 0x1c03, Type: Short, Sub Id (Short): TSCH Timeslot IE
        Data: 01eb01
    ▸ Nested Sub IE (long), Sub IE: Unknown IE, Length: 1
    ▾ Nested Sub IE (short), Sub IE: TSCH Slotframe and Link IE, Length: 10
      ▸ Payload Sub IE (short): 0x1b0a, Type: Short, Sub Id (Short): TSCH Slotframe and Link IE
        Data: 01010b0001000000000f
```

```
0000  60 00 00 00 00 59 11 08  bb bb 00 00 00 00 00 00   `....Y.. ........
0010  00 00 00 00 00 00 00 01  bb bb 00 00 00 00 00 00   ........ ........
0020  00 00 00 00 00 00 00 01  00 00 45 5a 00 59 06 a3   ........ ..EZ.Y..
0030  45 58 02 01 18 00 01 01  ff 01 01 01 01 01 01 01   EX...... ........
0040  01 02 02 02 02 00 00 00  00 00 00 00 00 00 00 31   ........ .......1
0050  40 ea a0 fe ca ff ff 02  00 00 00 cc 92 15 14 00   @....... ........
0060  3f 1c 88 06 1a bc ac 00  00 00 01 03 1c 01 eb 01   ?....... ........
0070  01 c8 00 0a 1b 01 01 0b  00 01 00 00 00 00 0f 53   ........ .......S
0080  04                                                 .
```

```
largedataset          Packets: 27804 · Displayed: 27804 (100.0%) · Load time: 0:0.297     Profile: Default
```

Fig. 4. ASN located in the payload section of IE

6.3.1.2 Approach

The objective of this portion of experiments is to determine whether ASN can be used to map traffic activities of low powered wireless networks operating in the IEEE 802.15.4e/TSCH mode. Subsequently, traffic behaviour patterns can be used to design a prediction model to identify anomalies in the corresponding network. ASN is a unique number used by IEEE 802.15.4e networks, operating in TSCH mode, to determine a time-slot and a channel id for a particular communication between two nodes. Even though each active node is able to determine active ASN by analysing the last received Enhanced Beacon, according to IEEE 802.15.4e/TSCH, nodes are not obliged to retransmit ASN in unicast data exchange.

IEEE 802.15.4e is specially tailored for wireless networks consisting of low powered nodes. In this thesis, data collected from a simulated wireless network, operating in TSCH mode, is used . However, it is important to emphasize that, individual simulated nodes are deployed in isolated virtual environment (sandboxed) and simulated nodes are operated in fully functional operating systems which can be implemented in certain hardware architectures. Therefore, the Media Access (MAC) and the Network layer parameters are not significantly affected by the simulation nature of this network. However, data associated with physical layer attributes could be significantly influenced by various environmental factors and data collected from an operational network is used in this thesis to perform experiments related to physical layer attributes.

6.3.1.3 Related Attack Vector

Adversaries take advantage of vulnerabilities of a particular system to compromise the system and corresponding data. However, systems operating in unrestricted resources are able to implement different techniques, including encryption, source-destination authentication and access control to protect data. Yet, devices operating in constrained resources are unable to utilize computationally expensive security controls to protect the data. Consequently, most low powered wireless networks are operating in minimum or non-existing security posture to preserve expensive resources, including battery power. Therefore, low powered wireless networks are vulnerable to most attacks witnessed in conventional networks. In the following, some possible attacks can be mitigated using a potential prediction model based on ASN as an input feature is discussed.

Different techniques, such as sequence numbers are, used by wireless nodes to synchronize the communication between them. Due to the low-security nature of wireless networks operating in constrained resources, intruders are able to intercept wireless communication between nodes, quite easily [12, 23, 50]. Furthermore, the broadcast nature of wireless communication further elevates the exposure. Adversaries are able to influence the synchronization between two nodes by altering sequence numbers. Furthermore, an intruder may be able to by-pass the authentication process and attempt unauthorized access to a wireless node. In conventional networks, different techniques such as sequence number randomization and header encryption/authentication are used to prevent those attacks.

Replay attacks have several intentions, including gaining access to another node or creating a denial of service attack, by harming synchronization, between two nodes. Furthermore, IEEE 802.15.4e doesn’t provide a specific mechanism to protect against replay attacks and it may rely on an upper layer implementation. The following diagram demonstrates a possible replay attack scenario on a network operating on a TSCH mode.

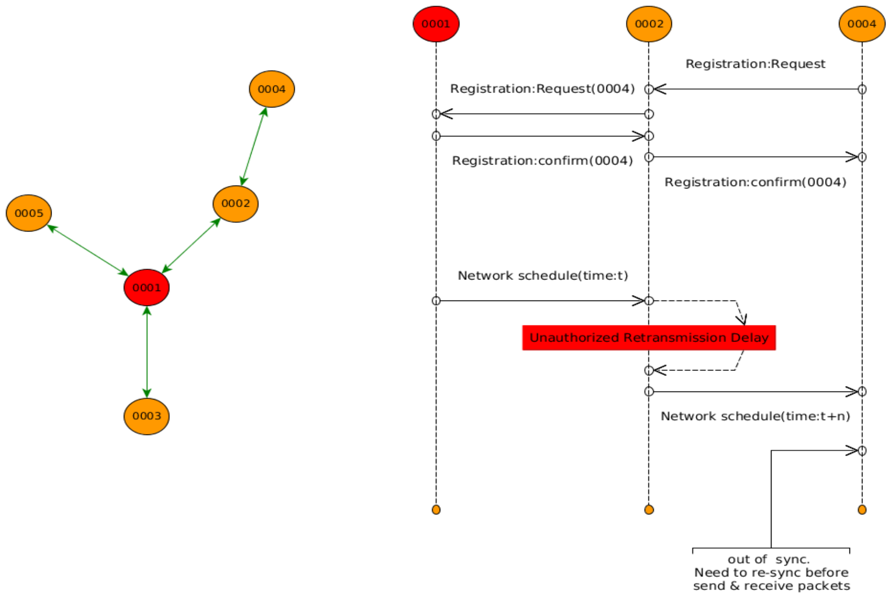

0004
0002
0005
0001
0003
0001
0002
0004
Registration:Request
Registration:Request(0004)
Registration:confirm(0004)
Registration:confirm(0004)
Network schedule(time:t)
Unauthorized Retransmission Delay
Network schedule(time:t+n)
out of sync.
Need to re-sync before
send & receive packets

Fig. 5. A potential DoS attack based on de-synchronization

Wireless networks, operating in IEEE 802.15.4e/TSCH, utilize EBs to synchronize the network and to relay the communication schedule to network nodes. Furthermore, 802.15.4e/TSCH networks are dependent on intermediate nodes to relay EB and other packets including data, control and management to remote nodes. In a synchronized environment, remote nodes should be able to receive EBs and data packets within a specific time slot, assigned by the PAN coordinator. Extended delays in receiving EB that contains scheduler information or synchronization information, may lead to the instability of a wireless network. Several factors could contribute to this phenomenon, including defective intermediate nodes, interference, congestion and malicious intention, such as hostile attack or deliberate sabotage, by a trusted intermediate node [12, 50, 53]. As a consequence, remote nodes may require re-synchronization or re-registration before participating in the data exchange and this may lead to a denial of service (DOS) attack.

The Absolute Slot Number (ASN) is transmitted in the payload portion of Information Elements (IE), in Enhanced Beacons (EB) and over 50 percent of experimental data collected from 802.15.4e/TSCH network are EBs. Certain threats such as transmission delays, network congestion and replay attacks are time sensitive and if ASN has a strong relationship with time, ASN can be used, as a parameter in time-series analysis. Furthermore, ASN changing rate (delta) can be used with classification models to predict anomalies caused by various attacks, including replay, spoofing and denial of service attacks.

6.3.1.4 Assumptions

The simulation (OpenWSN) is used to collect experimental data operating in IEEE 802.15.4e/TSCH mode. Furthermore, this particular simulator is specially designed to operate in TSCH (Time Slot Channel Hopping) mode using supplementary features, including CoAP (Constrained Application Protocol), RPL (Routing Protocol for Low-Power Lossy Networks) and minimal 6TiSCH (draft-ietf-6tisch-minimal-21). However, since the experimental network is operating in a simulated environment, some critical impact factors, such as interference, environmental effects, hardware failures, physical security threats, location-specific influences and human errors are not properly accounted. The following experiments are based on several assumptions and they are listed below.

- Critical elements of IEEE 802.15.4e/TSCH are accurately implemented by the simulator (OpenWSN)
- Operational parameters such as transmission-delay, distance, network formation process and routing convergence time are reasonably accurate.
- Functional restrictions of the simulator such as number of nodes, number of hops, transmission time and receive

sensitivity computations don’t impact the experiment’s outcome.

- Radiofrequency (RF) effects such as interference and multipath fading don’t influence data transmission, including transmission delay and retransmission rates, significantly.
- Simulated low powered wireless nodes are able to produce a comparable reading of networks operating in heterogeneous hardware.
- Experiments conducted using a smaller number of nodes is able to produce generalized results.
- Minimal scheduling mechanism (draft-ietf-6tisch-minimal-21), used in an experimental wireless network, adheres to the IEEE 802.15.4e /TSCH standard requirement.
- Expedited data collection techniques, used in experiments, won’t influence the accuracy of findings.
- Control, management and data packet rates of operational wireless networks are comparable to packet ratios observed in the simulated environment.
- A similar routing convergence process including convergence time and routing path calculation process should be observed in functional wireless networks
- Most low powered wireless networks are operating in fairly stable working conditions, including longer battery life, hardware and software stability.
- Most wireless nodes operating in IEEE 802.15.4e/TSCH mode have a higher permanence or drifting in a predictable trajectory.
- Security and physical restrictions do not prevent capturing tools from collecting wireless messages between nodes and the PAN coordinator.

- The PAN coordinator is operating with unconstrained resources, when direct data capturing method (from PAN coordinator), is used.
- The PAN coordinator is operating on known standards or provides the functionality to access the wireless packet stream.

6.3.1.5 Settings

6.3.1.5.1 Topology & Configuration

The wireless topology and routing structure of the experimental network is depicted in the following diagram.

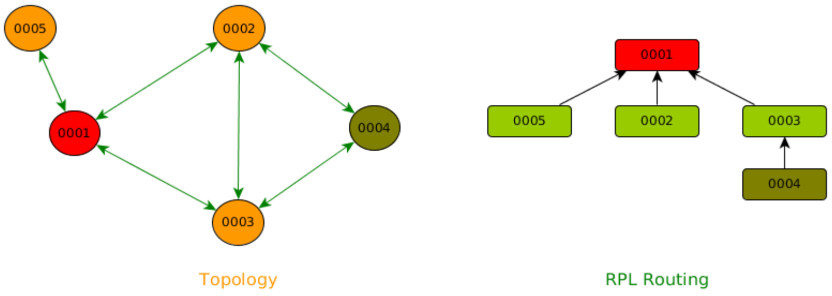


Fig. 6. IEEE 802.15.4e/TSCH wireless and RPL topology for simulated network

The node, with a label 0001, is assigned as the PAN coordinator and the topology indicates a partial mesh network. However, the RPL selects a direct path to the PAN coordinator from all nodes, but node 0004. The following table summarizes the network and environmental configuration used in these experiments.

Number of nodes	5
IP configuration	Node0001: 14:15:92:cc:00:00:00:01 Node0002: 14:15:92:cc:00:00:00:02 Node0003: 14:15:92:cc:00:00:00:03 Node0004: 14:15:92:cc:00:00:00:04

	Node0005: 14:15:92:cc:00:00:00:05
DAGroot	Node0001
Control node	Node0004
Application	Node0004 is configured to send a probe in every 5 seconds. Probe is a randomly selected value from a set of samples. Node0004 also transmits active ASN.

Tab. 4. IEEE 802.15.4e/TSCH network configuration

6.3.1.5.2 Data Collection

One of the main responsibilities of the Enhanced Beacon (EB) is to deliver the communication schedule to each node. The PAN coordinator also includes the active time-slot number (ASN) in the payload section of EB and intermediate nodes are responsible for relaying EBs to remote nodes. However, a certain portion of wireless packets, including control data, such as RPL, neighbour discovery (ND) and acknowledgments, as well as, packets carrying node-specific unicast traffic, such as sensor readings, are not required by the IEEE 802.15.4e standard to transport ASN in the payload. The EB ratio of a particular network is dependent on several factors, including the configuration of sensor nodes, PAN settings, number of nodes, network configuration settings and stability of the network. Yet, a data set with a higher EB packet ratio is critical for following experiments.

The following diagram summarizes ASN vs non-ASN packet ratio for different sample sets.

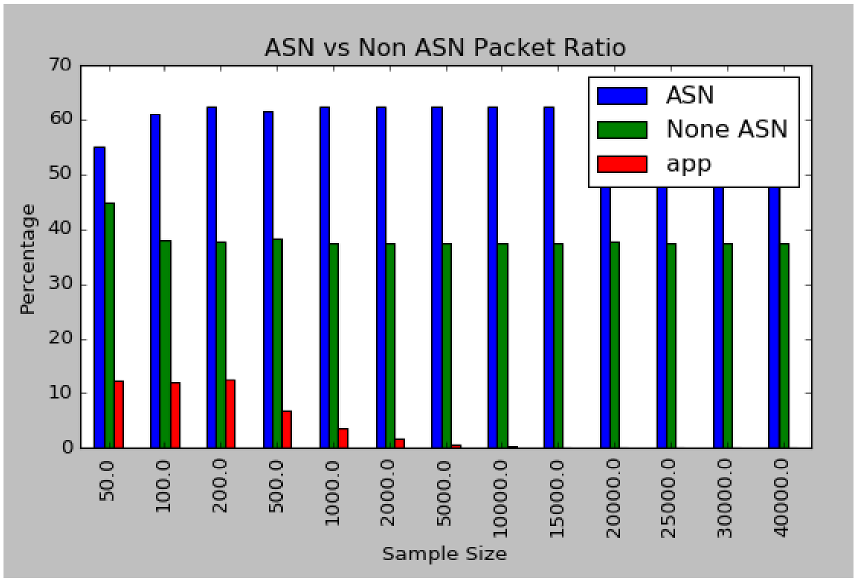


Fig. 7. ASN vs non-ASN packet ratio

As per the above diagram, the ASN & non-ASN packet ratio fluctuates during the formation of an IEEE 802.15.4e/TSCH based network. A higher number of RPL and Neighbour Discovery (ND) packets can be observed during the initial stage. ND is used by 6TiSH to assign IPv6 addresses to nodes. The RPL is responsible for defining the best available path to reach each node in the network. Initial traffic anomaly behaviour could be caused by excessive packet rates of RPL and ND. However, the network seems to produce consistent packet ratios after the first few hundred packets. The following experiments are based on several assumptions, including the stability of the network. However, instability of wireless networks may lead to regular topological changes and this may hinder the accuracy of findings. However, recurring factors, such as seasonal effects and scheduled hardware/software maintenance, can be identified by prediction models, trained with larger datasets.

6.3.1.6 Input Parameter set

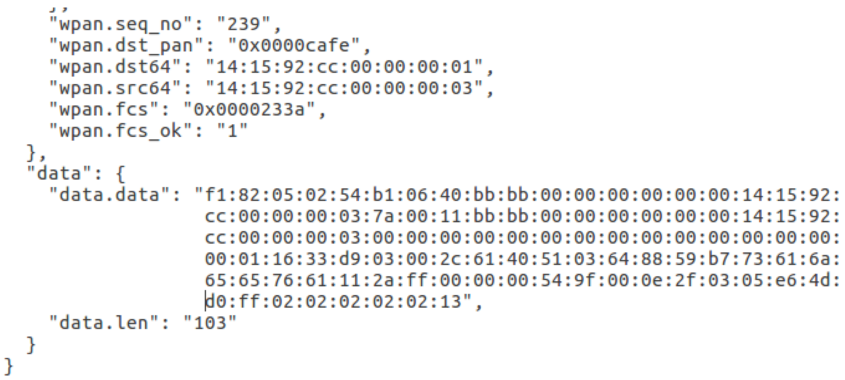
Packet Name	Packet Type	Description

frame.time_epoch	Float64	Absolute time used by network (formatted to float64)
frame.time_relative	Float64	Relative time (initiated to 0 by capturing tool)
frame.number	Int64	Frame Id used by 802.15.4
wpan.mlme_sub_ie.data	5 bytes (Hex)	ASN encapsulated in payload potion of IE represented by 5-byte Hex
data.data	Hex (up to 127 bytes)	Application-specific data (including ASN embedded in packet payload)

Tab. 5. IEEE 802.15.4e/TSCH attributes examined

6.3.1.7 Data Extraction

Time related data (frame.time_epoch, frame.time_relative) and frame identification (frame.number) can be directly extracted from the header section of captured wireless data. However, time-slot identification (ASN) (wpan.mlme_sub_ie.data) and application-specific data (data.data) are stored in the payload section of Information Elements (IE) in IEEE 802.15.4e/TSCH packets, using variable-length hexadecimal. Some special methods are developed to extract embedded attributes from IEs. The following diagram depicts a data portion of a unicast packet extracted from 802.15.4e/TSCH network.



```
      "wpan.seq_no": "239",
      "wpan.dst_pan": "0x0000cafe",
      "wpan.dst64": "14:15:92:cc:00:00:00:01",
      "wpan.src64": "14:15:92:cc:00:00:00:03",
      "wpan.fcs": "0x0000233a",
      "wpan.fcs_ok": "1"
    },
    "data": {
      "data.data": "f1:82:05:02:54:b1:06:40:bb:bb:00:00:00:00:00:00:14:15:92:
                    cc:00:00:00:03:7a:00:11:bb:bb:00:00:00:00:00:00:14:15:92:
                    cc:00:00:00:03:00:00:00:00:00:00:00:00:00:00:00:00:00:00:
                    00:01:16:33:d9:03:00:2c:61:40:51:03:64:88:59:b7:73:61:6a:
                    65:65:76:61:11:2a:ff:00:00:00:54:9f:00:0e:2f:03:05:e6:4d:
                    d0:ff:02:02:02:02:02:13",
      "data.len": "103"
    }
  }
```

Fig. 8. IEEE 802.15.4e/TSCH Unicast packet payload with embedded ASN

6.3.1.8 Sample Sets

The objective of the first experiment, in the following experiment series is to examine the relationship between prediction accuracy and training sample size. As previously mentioned, higher amounts of control data, including ND and RPL packets are observed during the formation of 802.15.4e/TSCH networks. Therefore, to prevent outliers, the data capturing process ignores packets generated during the first 30 seconds of network formation. However, malformed data including packet errors and capturing errors are used in training models to obtain more realistic results.

Ten different sample sets, ranging from 50 packets to 40000 packets, are tested in the following experiment. While 80 percent of the samples from each sample set are used in the training process, the remaining samples are used to evaluate the effectiveness of the model. The following table describes the different sample sizes and the percentage summary of data, used in experiments.

Sample size	EB(%)	RPL(%)	App(%)	Ack(%)	Others(%)
50	55.1	12.24	12.24	20.41	0.0
100	61.62	11.0	12.12	15.15	0.0
200	62.31	12.56	12.56	12.56	0.0
500	61.72	12.63	7.01	7.01	11.62
1k	62.46	12.31	3.5	3.5	18.22
2k	62.48	12.36	1.75	1.75	21.66
5k	62.37	12.36	0.7	0.7	23.86
10k	62.41	12.36	0.35	0.35	24.53
15k	62.37	12.39	0.23	0.23	24.77
20k	62.37	12.4	0.18	0.18	24.89
25k	62.38	12.38	0.14	0.14	24.95
30k	62.4	12.37	0.12	0.12	25.0
40k	62.45	12.36	0.09	0.09	25.02

Tab. 6. Packet ratio of IEEE 802.15.4e/TSCH based simulated network

6.3.1.9 Labelling

Labelled data is used by supervised machine learning methods to train prediction models. In the experiments, binary classification models are used and training data is labelled, using 0 (negative) and 1 (positive). The positive portion (labelled as 1) of data is selected randomly from each sample set. The negative portion of samples is generated using two different techniques. Those are:

1. Sequence Permutation

   With sequence permutation technique, data belonging to one or more input features are shuffled in a random order to generate anomaly data. However, with this technique, there is a possibility of assigning a legitimate packet as anomalous due to the randomness of the permutation process and extra attention is given in this work to prevent such behaviour.

2. Noise Injection

   Noise injection is based on inducing a random noise with a predetermined variance to input data. The labelling process based on noise injection is described in the following. The probability distribution function of a normal distribution can be derived using several methods, including Gaussian distribution function. In the following experiments, the Gaussian distribution function is used to calculate random noise values used in the negative data set generation process. The randomness of a noise sample is based on an adjustable, standard deviation value.

Gaussian distribution function

$$f(x) = [e^{-(x-\mu)^2/(2\sigma^2)}]/\sqrt{2\pi\sigma^2}$$

Where $\boldsymbol{\mu}$ is the mean value, $\boldsymbol{\sigma}$ is standard deviation.

3. Node Generated Noise

   With this technique, random data with a different timestamp and payload value are generated in individual nodes to produce anomalous data. However, using this (node generated noise) method, only certain types of traffic behaviours including anomalies due to network congestion, defective node, replay attack, payload modification could be detected.

Experimental results related to different labelling techniques and classification methods can be found in the Experiment Result chapter.

With all experiments in this work, prediction models are trained and validated using datasets with an equal number of different labels , if not otherwise specified. Positive and negative data sets are mixed in a random order to prevent biased learning processes. 80 percent of mixed samples (positive and negative) are trained with different machine learning methods and the remaining 20 percent of mixed data samples are used to evaluate the accuracy of prediction models. However, a separate sample set extracted from the same distribution is used to validate the model using false positive and negative rates.

6.3.1.10 Regularization

Regularization is used in machine learning to cope with bias and to create prediction models, with a higher generalization power. In essence, the regularization process augments certain input parameters that cause an improvement in the outcome. A similar technique is used in this work to prevent ASN replay attacks, by regularizing, time-window size. The following diagram is used to further discuss the regularization technique, used in this work.

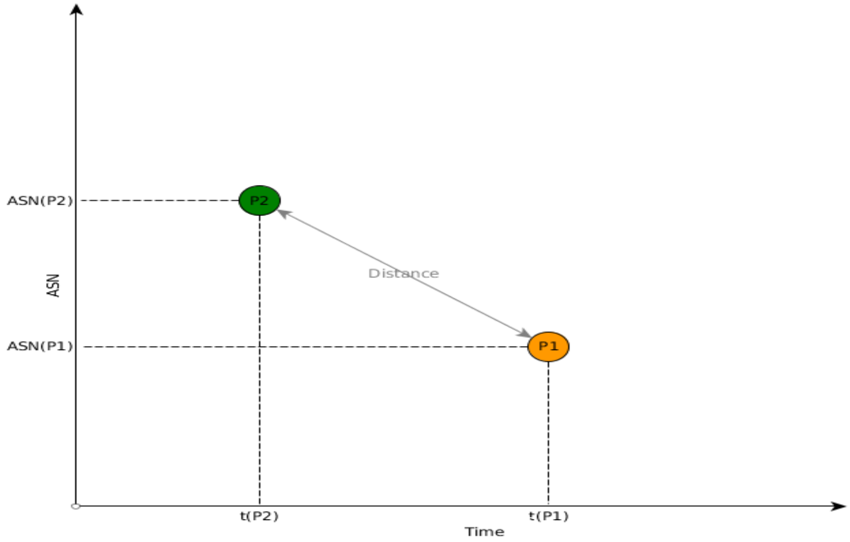


Fig. 9. Distance between two vectors

The Euclidean distance between two vectors, is calculated using the following formula.

Vector A = $\{a_1, a_2, a_3, \ldots, a_n\}$
Vector B = $\{b_1, b_2, b_3, \ldots, b_n\}$

The Euclidean distance between A and B (denoted as Euclidean (A, B))

$$\text{Euclidean}(A,B) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

In the above example, the Euclidean distance between point P1 and P2 is calculated as:

$$\text{Euclidean}(P_1,P_2) = \sqrt{\left(t(p2) - t(p1)\right)^2 + \left(asn(p2) - asn(p1)\right)^2}$$

6.3.1.11 False Positive/Negative

False Positive and Negative error rates provide an important metric to quantify the effectiveness of security control. However, the definition of False Positive and False Negative in Information security is different, than in some other areas such as Statistics and Medicine. The following definition is used to describe the False Positive and False Negative rates in this work.

- False Positive – Legitimate data is marked as negative (not allowed access).
- False Negative – Illegitimate data is marked as positive (allowed access).

6.3.1.12 Training/Testing

Four classification algorithms (Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Artificial Neural Networks (NN) and Decision Tree (DT)) and one regression model (linear regression) are evaluated in following experiments. However, a few other techniques, including ensemble methods are examined to improve the result. If not otherwise noted, all experiments are based on 80/20 sample ratio for training and testing process.

The accuracy of a linear regression model can be determined by calculating the R-squared (R^2) value (coefficient of determination). R-squared is used to analyze how differences in one variable can be explained by a difference in the second variable. R-squared measures how accurately data can be fitted into the regression line.

R-squared = Explained Variation/ Total variation

R-squared is represented by a percentage value and it can be calculated using the following formula.

R-squared $= [(\mathrm{n}(\sum t * asn)\text{-}(\sum t)(\sum asn))/(\sqrt{(n\sum t^2 - (\sum t)^2)(n\sum asn^2 - (\sum asn)^2)}))]^2$

Where
t = timestamp
asn = Absolute Slot Number
n = number of samples in test data set

Classification algorithms are able to classify numerical, categorical and continuous input data into a finite number of classes. However, regression models are based on ordinal, continuous inputs and outputs. Different techniques such as likelihood, n-neighbours and square-error threshold can be used to determine a potential class for a corresponding input vector. In this experiment, Euclidean distance and predetermined threshold values are used to determine potential output values. Furthermore, the input vector for Euclidean computation is manipulated, using a scalar parameter to investigate the individual input parameter's impact on prediction accuracy. Potential candidate selection process is described in the following diagram.

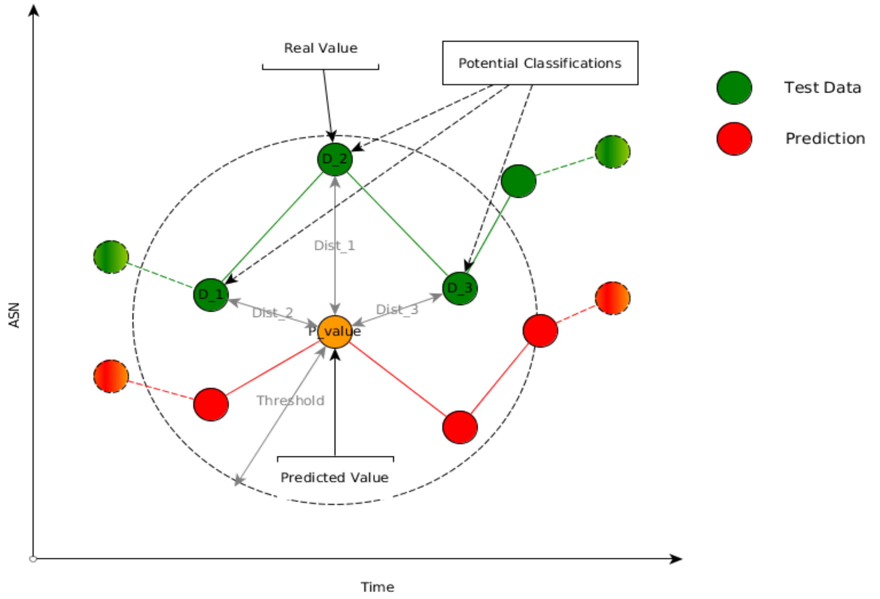


Fig. 10. Classification process of the regression model (ASN vs time)

The above diagram demonstrates the output selection technique used in a model based on linear regression. While green nodes represent the values of test data and red nodes represent predicted values for the corresponding test data. The orange node represents (P_value) the predicted value for input data D_2. For the most accurate prediction, P_value and D_2 should be aligned. However, regression models are based on approximation and a secondary mechanism is necessary to determine the most fitting output value for a particular input. For instance, in the above example, one of three nodes (D_1, D_2 and D_3) could be a potential match. Dist_1, Dist_2 and Dist_3 represent distance (modified Euclidean) from P_value to each node D_1, D_2 and D_3. A weight is applied to each feature (scalar_effect) to examine the impact of each input parameter.

$$\text{Distance} = \sqrt{\sum_{i=0}^{n} w_i * (p_{0i} - p_{1i})^2}$$

Where w is the weight added to each distance and n is number of input features.

The following diagram describes the four different classification conditions that can be observed in this experiment. The following definitions and names are used in this work.

- Prediction – value computed by the prediction model
- Correct Data Point – output value associated with corresponding inputs
- Wrong Data Point – value determined by the distance calculation formula

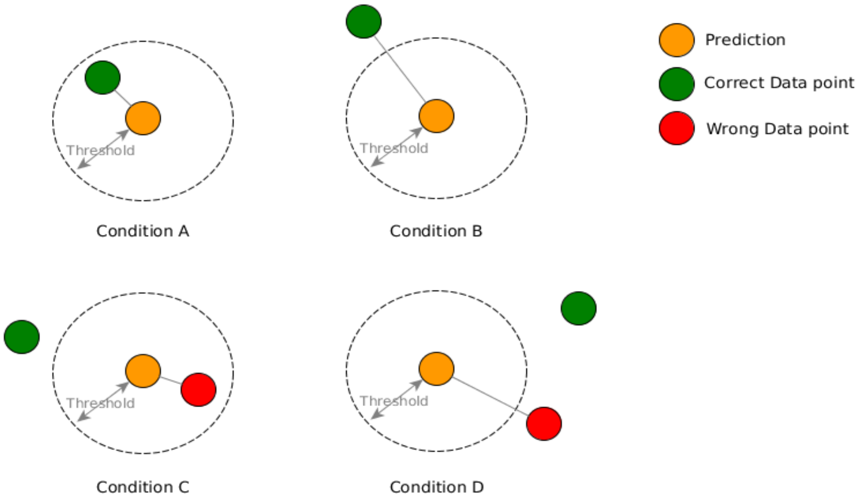


Fig. 11. Classification conditions

Condition A satisfies the correct prediction inside the threshold value. Condition B (False Positive) also determines the correct prediction. However, it locates outside of the comfort zone (threshold). In condition C (False Negative), an Absolute Slot Number (ASN) can be found within the threshold range, but the ASN number does not belong to the correct 802.15.4e/TSCH beacon. In condition D, no packet with an ASN number can be found within the threshold range.

The prediction accuracy for each condition is calculated using the following method.

$$\text{prediction\_accuracy} = \frac{1}{n}\left(\sum_{i=1}^{n} predict\left(p_{labelled_i}, p_{pred_i}\right)\right)$$

where n is the number of samples in the test set and

$$predict\left(p_{labelled_i}, p_{pred_i}\right) = \begin{cases} 0 if p_{labelled} \neq p_{pred} \\ 1 if p_{labelled} = p_{pred} \end{cases}$$

Experiment results for the linear regression models can be found in the corresponding section in Experiment Results.

6.3.2 Physical Layer (Wireless Attributes)

6.3.2.1 Background

Implementation of low powered wireless networks, operating on IEEE 802.15.4e, relies on several other supporting mechanisms. The open systems interconnection (OSI) model conceptually separates network architecture into seven interconnected layers and each layer is responsible for a certain set of operational responsibilities. For instance, open standards, such as TCP/IP (network/transport layer), 802.15.4 (MAC/physical layer) and COAP (application layer), working together to provide end-to-end communication between two nodes. In the following section, common attributes of the physical layer (wireless) including signal strength (RSSI), signal-to-noise ratio (SNR), link quality (LQI) and link distance (LD) are evaluated for potential a traffic anomaly detection model. (Those four attributes are denoted by Attributes $^{[wireless\_common]}$.)

6.3.2.2 Approach

Previous experiments, related to attributes of IEEE 802.15.4e/TSCH, have been performed in a simulated environment due to the limited accessibility of hardware operating in 802.15.4e/TSCH mode. However, the Attributes $^{[wireless\_common]}$ (RSSI, SNR, LQI, and LD) are common for most wireless networks operating in similar frequencies. Furthermore, Attributes $^{[wireless\_common]}$ are dependent on several factors including environmental characteristics, such as interference, multipath fading and temporal variables, such as wind, storm, snow and abscission. However, to simulate these external influences in a simulated environment is a challenge. Furthermore, both IEEE 802.15.4 and 802.11 networks are operating in the same unlicensed frequency range and similar environmental conditions. Therefore, data collected from an operational outdoor wireless network, operating in IEEE 802.11, has been used in the following set of experiments. More details regarding environmental settings and data collection methods are described in Topology & Configuration subsection.

6.3.2.3 Related Attack Vector

Due to the deterministic behaviour of low powered networks, a significant portion of traffic anomalies on low powered networks can be detected by analyzing the traffic flow. For instance, MAC-spoofing is a common attack launched in both wired and wireless networks. The MAC address is a 6-byte number, used by communication protocols to uniquely identify network nodes. Every communication interface, operating in IEEE 802.x networks are uniquely identified by a MAC address and a MAC address is assigned and embedded into the network interface device by the hardware manufacturer, during the manufacturing process. Theoretically, the MAC address assigned to a network interface shouldn’t be modified; however, modern operating systems are able to overwrite the original MAC address for various reasons. The MAC address is used by upper-layer protocols, such as the Internet Protocol (IP), to accommodate application-level communication between peers. With MAC-spoofing attack, the intruder high-jacks a MAC address of a legitimate entity to initiate an unauthorized activity. In some variances, the intruder may conceal his own MAC address to avoid detection. Different techniques such as DHCP snooping can be utilized to mitigate MAC address spoofing attacks.. If the operational behaviour of a low powered wireless network can be mapped, spoofing attacks can be detected by comparing a suspicious flow with normal operational flows. Other techniques such as replay attacks and flood attacks are used by adversaries to launch unauthorized activities, including denial of service attacks. These activities can be detected by comparing with normal-operation- map in a particular network environment. Especially in low powered environments, where the number of legitimate operations are restricted and can be defined using a finite number of tasks, traffic anomalies can be identified with higher accuracy by carefully examining the activities of each low powered node.

6.3.2.4 Assumptions

Data collected from a wireless network with unconstrained resources are used in the following experiments. Even though, both low powered and unconstrained networks are utilizing similar frequencies, a few fundamental differences between these two networks can be observed. The following table summarizes some of the assumptions taken into account in the proceeding experiments of this section.

Channel width	Channel width used by a wireless network operating in unlicensed frequencies can be adjusted as per a particular requirement. If a network comprises higher node density and smaller data transfer, the network can be adjusted to have a smaller channel width to accommodate a higher number of non-overlapping channels. However, most networks are operating in standardized channel widths to provide inter-operability. For instance, IEEE 802.11 network, operating on mode ‘b’ (2.4GHz) use 11-14 overlapping channels the size of 22 Mhz. Yet, low powered wireless networks operating in IEEE 802.15.4, 2.4GHz frequency, utilize 5MHz channels to accommodate 16 non-overlapping channels in North America. In this experiment, data from a wireless network operating in 20Mhz channel width has been utilized and led to the assumption, that the behaviour of input parameters (Attributes [wireless_common]) are not significantly different from the behaviour of data collected from a low powered wireless network.
Radio power	The power usage of the transceiver is considerably different between radios operating in a low powered environment and radios with unrestricted power settings. The amount of transmit power dictates how far a radio signal will reach. Furthermore, the available power determines how much noise can be absorbed. Wireless transceivers used in the following experiments are able to penetrate further and able to filter noise more

	effectively than the low powered devices. Even though, data is collected from a live network operated in similar frequencies, the parameters such as Signal-to-Noise-Ratio (SNR) and Link Quality (LQ) may interpret differently by low powered wireless devices. The following experiments are performed with the assumption that the behaviour of attributes used in the following experiments are not significantly influenced by the transmit-power of the radio transceiver, since the receive sensitivity of the incoming signal in both low powered and conventional network is comparable
Distance	Low powered wireless devices use low powered radio units and their transmission range is limited to few meters. However, wireless devices used in the following experiments are deployed a few kilometers apart and the wireless nodes are able to approximate the link distance quite accurately. The distance calculation mechanism is based on transmission delay between nodes. However, the calculation of the distance between two low powered wireless nodes could be a challenge. This experiment is based on an assumption that the distance calculation mechanism of low powered wireless networks is fairly accurate.
Interference	Low powered wireless networks mainly operate in private zones, with deterministic perimeters values. However, in conventional wireless networks, radio signals can penetrate through unknown zones, where unpredictable noise can be

	experienced. Furthermore, interference and the multipath effect caused by different objects can be controlled in a private zone. Data used in the following experiments are prone to a higher variance of interference. Therefore, it is safe to assume that variance of interference and noise in low powered wireless networks is lower than the noise variance of data used in the following experiments. The radio frequency (RF) interference of a particular node is determined by the SNR value of the corresponding node.
802.15.4e/TSCH	IEEE 802.15.4e networks operating in TSCH mode utilize channel hopping mechanism to prevent excessive energy consumption due to interference, re-transmission and packet loss. However, a fairly complex computational mechanism has to be implemented to determine the RSSI, LQ and SNR values, since those values can be varied among different channels utilized in TSCH mode. However, the following experiments are based on the assumption that a mechanism is in place to measure the reading of Attributes [wireless_common] and they are not significantly varied in different channels utilized by the TSCH mode.

Tab. 7. Assumptions (Physical layer feature set)

6.3.2.5 Settings

6.3.2.5.1 Topology & Configuration

Experimental data is collected from an operational, fixed, wireless network from a wireless service provider (Routcom Inc.) operating north of Toronto. This particular wireless network is deployed over a 600 square kilometer area using both licensed and unlicensed frequency bands to provide Internet access to rural customers. Routcom’s wireless network is built in a partial-mesh topology and from the data-link layer point of view, the network is operating in star and tree topology. The following diagram depicts the segment of Routcom network topology used in this work. It is important to emphasize that regardless of the power consumption or transmit power of individual nodes, wireless networks operating in similar frequencies and modulations, require similar receive sensitivity to operate reliably. For instance, for both conventional and low powered network operate in 2.4Ghz frequency, the reliability of a given radio link is not determined by the transmit power, but the receive sensitivity of the receiving node. Consequently, while conventional network with higher output power able to travel longer distance, low powered transmitter with low power, able to travel shorter distance while keeping similar receive sensitivity.

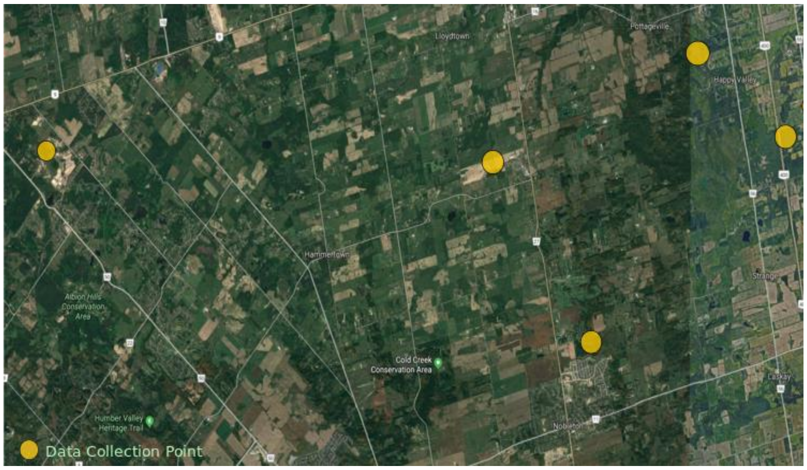


Fig. 12. Routcom wireless network segment

In the Routcom network segment used in this thesis, end nodes are connected to an Access Point (AP) using the star topology. Some APs are configured as bridged devices to connect multiple segments, using tree topology and others are configured as routed devices, where nodes belonging to one access point are not able to communicate with nodes from another access point using layer 2 (MAC) address.

The Routcom wireless network is designed using both licensed and unlicensed frequencies (900MHz, 2.4Ghz, 3.5Ghz, 5.8Ghz, 24Ghz). Furthermore, several wireless techniques, such as point-to-point (ptp), point-to-multipoint (ptmp) and wireless distribution mode (WDS) are utilized to provide reliable service to their customers. As previously noted, the wireless network is partitioned using layer 3 routing and each partition is further segmented using layer 2 VLANs. End units are connected to APs using both open standards such as 802.11a/b/n and proprietary systems (Nstreme, NV2) operating in unlicensed frequencies. Data used in the following experiments are collected from multiple wireless network segments operating on 802.11 mode.

The Routcom Wireless network comprises over 25 telecommunication towers situated in environmentally, diverse locations. Some towers are surrounded by dense trees and others are located in residential or commercial areas with higher interference presence. Furthermore, the physical structure of the towers and the different wireless equipment react differently to different weather conditions. For instance, compared to standalone towers, radios installed in guyed-towers (towers with guy-wires anchored to the ground for stability) may provide higher resistance to heavy wind. The following table describes some of the factors that could influence the attributes (RSSI, SNR, LQI, and LD) tested in the following experiments.

Factor	Description
Trees	If a transmitter is surrounded by trees, different weather conditions could affect the Attributes [wireless_common]. The heavy wind could obstruct the signal path between the transmitter and the receiver. The radio signal passing through wet leaves seems to have higher attenuation. Snow on tree leaves also impact the above-mentioned attributes. If a transmitter is surrounded by non-evergreen trees, different times of the year could produce different measurements.
Terrain	Surrounding terrain could also affect the readings of tested attributes, especially if the radio signal bounces before reaching its destination. If a transmitter is surrounded by farmland, with different crops, different readings can be expected.
Residential Area	In residential neighbourhoods, different wireless devices operating in the same frequencies, deployed for different intentions, including data communication and ISM related tasks can be found. Those devices generate a significant amount of noise and as a consequence, inconsistent

	measurements can be expected.
Commercial Area	Radio devices located on commercial properties could experience a similar effect to residential areas. However, noise generated on commercial properties can be significantly, reduced during off-hours.

Tab. 8. List of factors influencing the feature-set

6.3.2.5.2 Data Collection

Five locations (network segments) with different environmental settings are used to collect the data used in the following experiments. The following table describes the details of the locations used in data collection.

Location	Description
Loc_A	Remote nodes are located in heavily populated residential areas. The central station (access point) is mounted on a 120" guyed tower in an isolated location
Loc_B	Several industrial establishments are situated between most wireless nodes and the central station and the AP is mounted on a 120" guyed tower
Loc_C	Most remote stations are surrounded by trees. The central station is mounted on a standalone tower in a residential area. A significant amount of mixed trees are located between the tower and the remote stations
Loc_D	A central station is located in a guyed tower, surrounded by irregular terrain. A tower is

	heavily affected by high wind and blowing snow
Loc_E	The central station is mounted on a guyed tower, in an industrial area, located on flat terrain. Around 20 percent of the remote nodes are affected by wet leaves and snow, during the winter

Tab. 9. Wireless segments details

As previously mentioned, a different time of the day, week, month or year, can generate significantly different measurements for Attributes [wireless_common]. Several time intervals have been used to collect data to produce a comprehensive data map for each node. A detailed comparison of data collected at different times can be found in the Seasonal Effects subsection in the Conclusion section. The following table describes the data collection process.

Dataset Name	Frequency
DataSet_1	Every 1 minute
DataSet_2	Every 1 Hour
DataSet_3	Every 4 Hours

Tab. 10. Data collection frequencies

Data, collected over an extensive period of time (over a six month period) is used to build a detailed data set, representing various impacts, including seasonal effects, interference and maintenance downtimes.

Although, the following section is limited to examining four attributes (also known as Attributes [wireless_common],), a brief statistical analysis has been performed to observe the behaviour of aforementioned time-series attributes (packets, frames, bytes and uptime). Despite the fact that packets, frames and bytes represent three different numerical values in the above sample, they constitute the same property in three different layers. Hence, the following diagram is generated using the 'packets' attribute to describe the regression of received packets, over a period of time. Each line (different color) represents the regression of a different node.

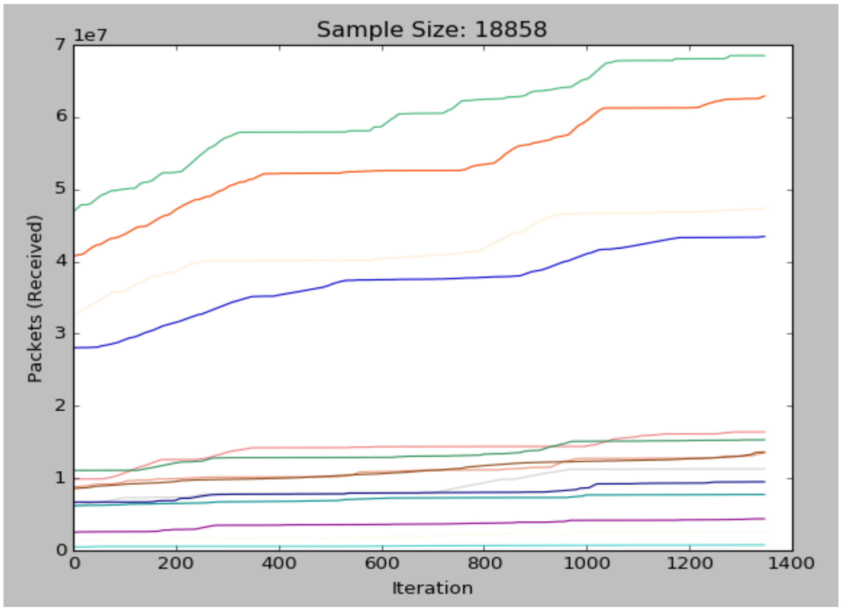


Fig. 13. Packet flow of individual nodes

The diagram confirms a potential positive correlation (Pearson r: 0.96) between the packets' flow and the time. However, a different correlation coefficient can be observed for the packet flow of each node. Furthermore, common trends can also be observed in the above diagram. For instance, during the iteration period of 800 – 1000, higher packet flow can be experienced in most nodes. Several factors, including interference could contribute to such behaviour and these occurrences can provide strong properties, determining the operational behaviour of corresponding nodes. Subsequently, these properties can be used to define normal operational conditions and to build complex prediction models to identify anomalies in low powered wireless networks.

6.3.2.6 Input parameter set

The feature vector (Attributes [wireless_common]) used for experiments in the following section include both numerical and categorical attributes. The following table summarizes the range of values used by each input parameter.

Parameter	Range	Unit
LQI	0 – 100	Percentage
Distance	0 – 100	KM
MAC	00:00:00:00:00 – FF:FF:FF:FF:FF	Hexadecimal (categorical)
Signal(RSSI)	-120 – 0	dB
SNR	0 – 100	dB
Uptime	Integer(64bit)	Seconds

Tab. 11. Input feature-set details

The following table is generated using data originated from a single node and this summarizes the descriptive statistics of input parameters. 18000 records are used to generate the following report.

	RSSI	SNR	LQI	Distance
Mean	-71.49	19.77	74.21	2.37
STD	5.06	5.20	18.13	0.48
Min	-91.00	0.0	0.00	2.0
Max	-58.00	34.00	99.00	3.0
Median	-71.00	20.00	77.00	2.0

25%	-75.00	16.00	69.00	2.0
50%	-71.00	20.00	77.00	2.0
75%	-58.00	34.00	99.00	3.0

Tab. 12. Descriptive statistics of input features

The above diagram demonstrates a higher variance (standard deviation: 18.13) of the LQI (Link Quality) compared to the RSSI (Signal Strength) and the SNR (Signal/Noise Ratio). Furthermore, statistics indicate no strong relationship between the RSSI and the LQI (Pearson r: 0.001) and a strong relationship between the RSSI and the SNR (Pearson r: 0.89). However, both the RSSI and the SNR are evaluated in the following experiments to obtain higher prediction accuracy.

The following diagram is generated using 18000 samples captured from a 14-node wireless network and this demonstrates the signal (RSSI) variance for each node.

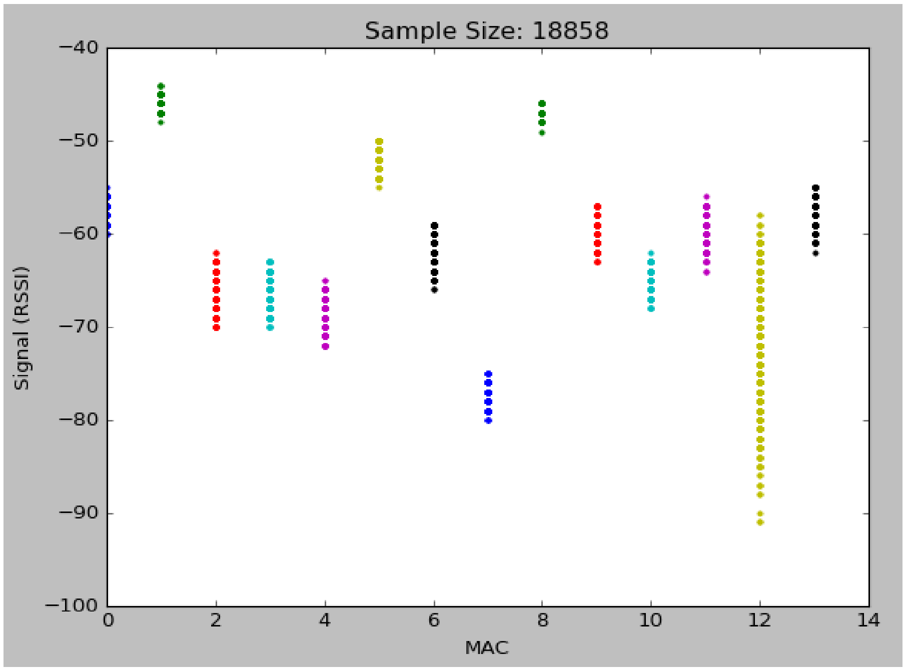


Fig. 14. Signal variance of individual nodes

The above diagram demonstrates a higher variance of RSSI for node #12 (yellow) and node #8 seems to have a relatively stable link. This effect could be caused by several factors including the physical stability of the radio mount and various obstacles, including trees between the two nodes. However, the above diagram confirms that a majority of nodes in this network segment are able to generate consistent signal strength (RSSI) during the test period.

The following diagram describes the frequency distribution of each input parameter.

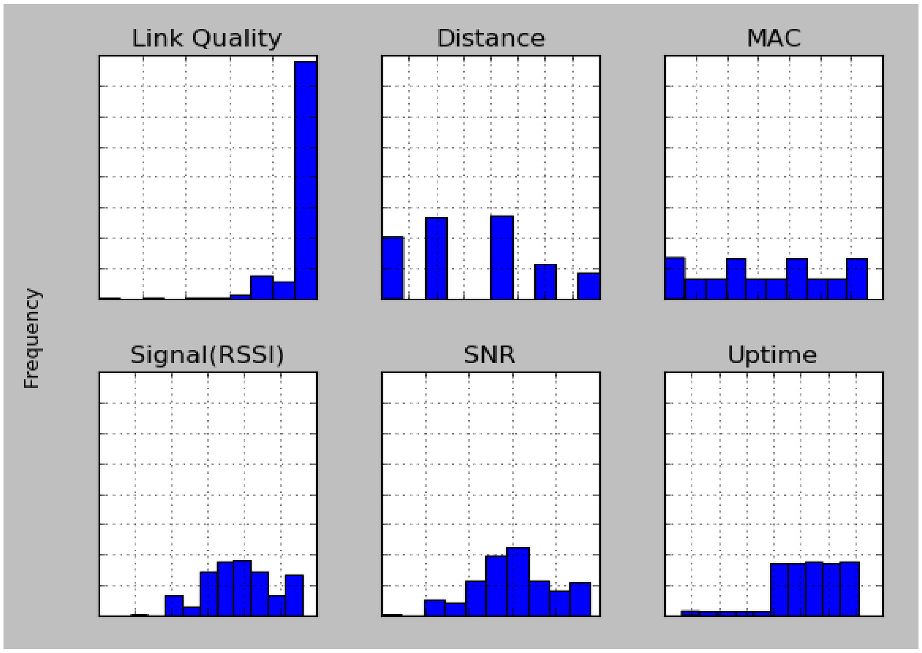


Fig. 15. Frequency distribution of input parameters

The above diagram demonstrates non-parametric properties of normally distributed models for the RSSI, SNR and distance attributes. Furthermore, the LQI (Link Quality Indicator) also demonstrates the characteristics of a half-fold normal distribution model. The MAC address is represented as a categorical attribute and it can be classified into a finite number of classes.

Prediction models based on supervised learning require a finite number of classes (labels) before being trained using a classification algorithm. However, the non-automated determination of classification groups could degrade the objectives of machine learning. In the following, machine learning methods are used to evaluate whether clustering techniques can be used to classify wireless data into multiple groups, based on similar characteristics. Sample data used in the following experiment is collected from a wireless segment with 14 nodes and following diagrams have been generated using 1000 samples. The mean-shift machine learning algorithm is used to identify potential clusters in the sample set. Several parameters are evaluated and the below diagram confirms that, none of the attributes is significantly contrastive to identify 14 different classes when using the mean-shift algorithm.

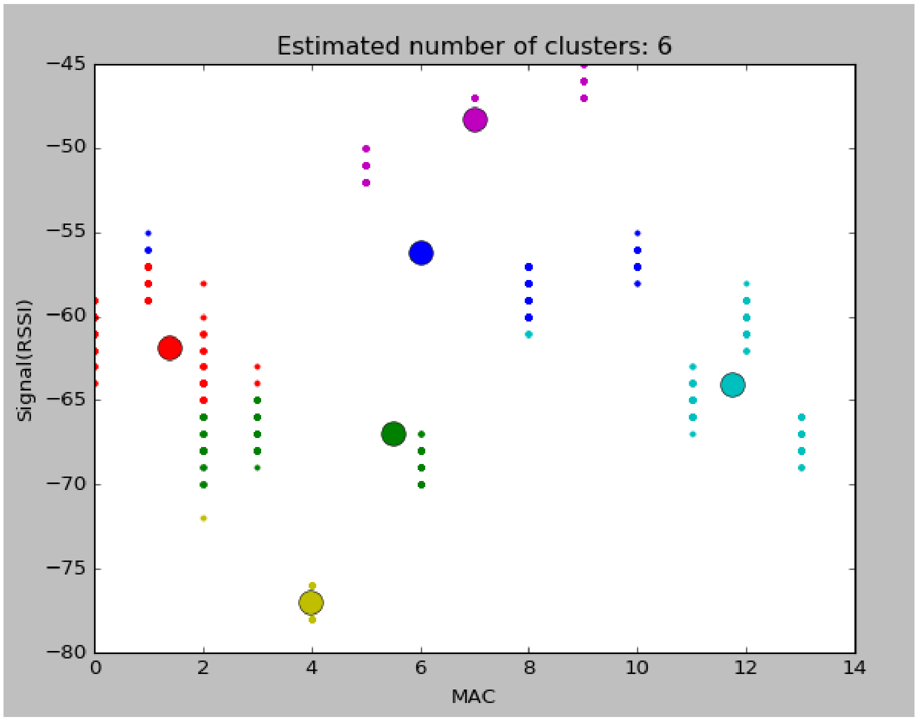


Fig. 16. Dynamic cluster detection [RSSI, MAC]

The above diagram depicts six different classes, using parameter list containing RSSI and the MAC address.

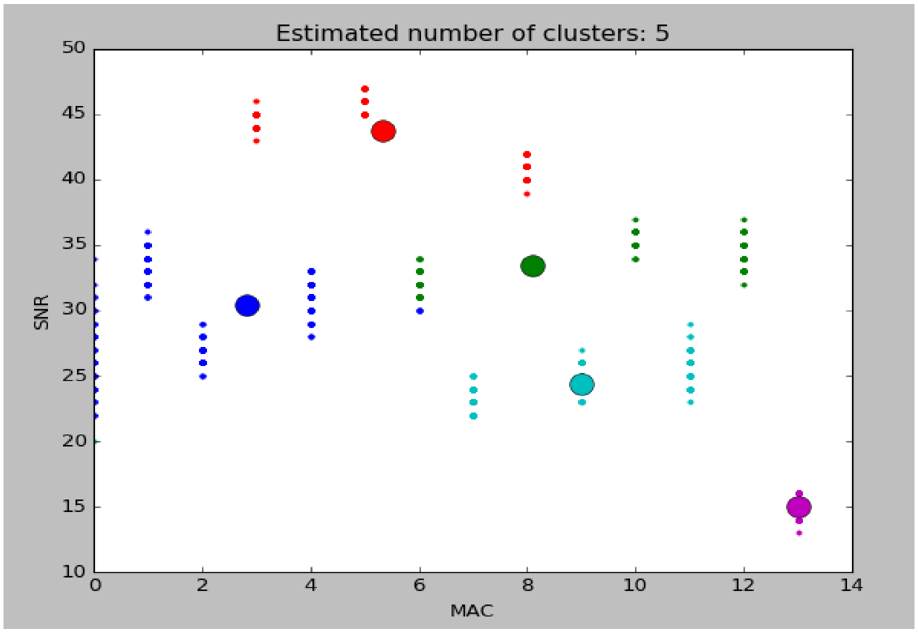


Fig. 17. Dynamic cluster detection [SNR, MAC]

The above diagram confirms there are five different clusters, when using the SNR and the MAC as input parameters.

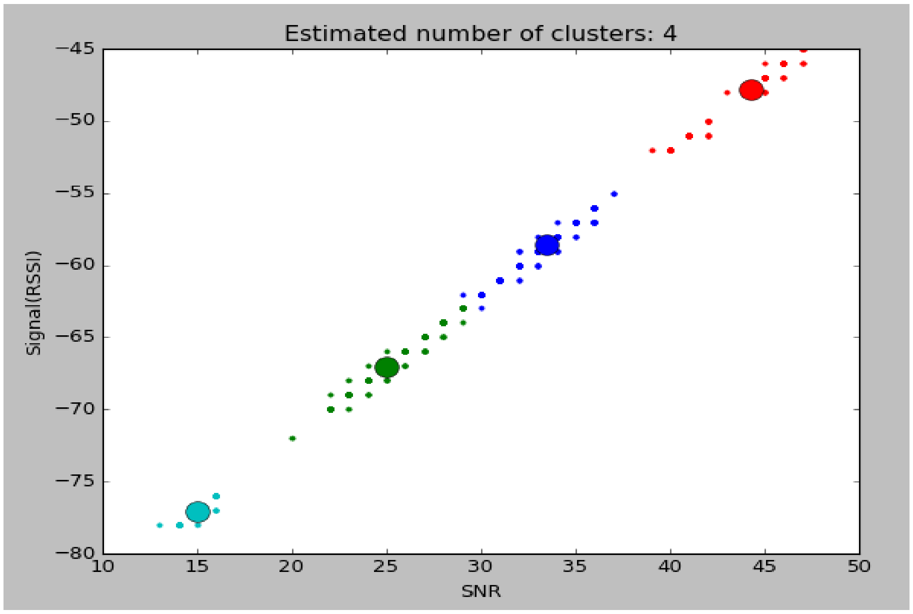


Fig. 18. Dynamic cluster detection [RSSI, SNR]

The above diagram demonstrates a strong positive relationship between the RSSI and the SNR. However, a combination of these parameters is only able to identify four different classes.

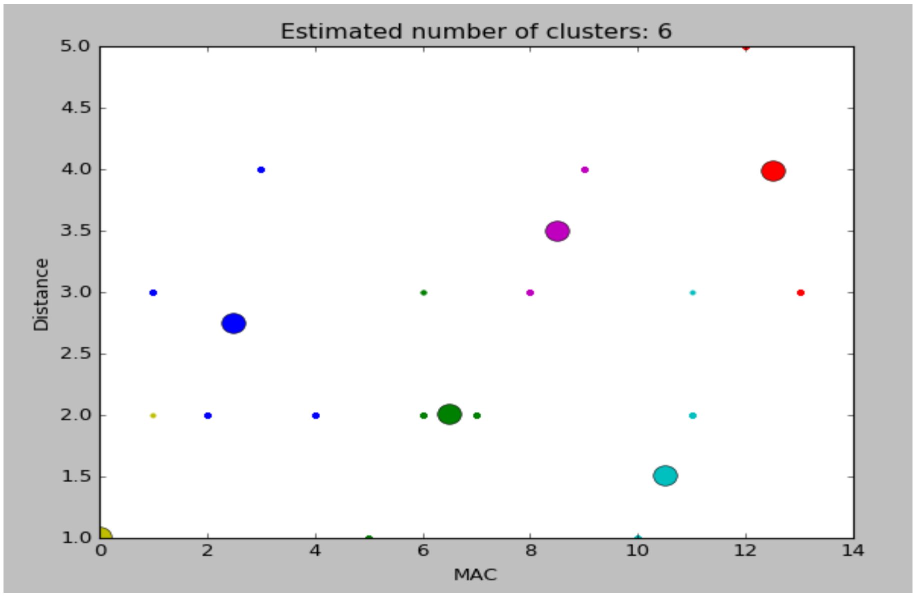


Fig. 19. Dynamic cluster detection [LD, MAC]

The input feature vector containing the MAC address and the Link Distance (LD) (above diagram) as well as MAC address and the LQI (below diagram) are able to identify six different clusters.

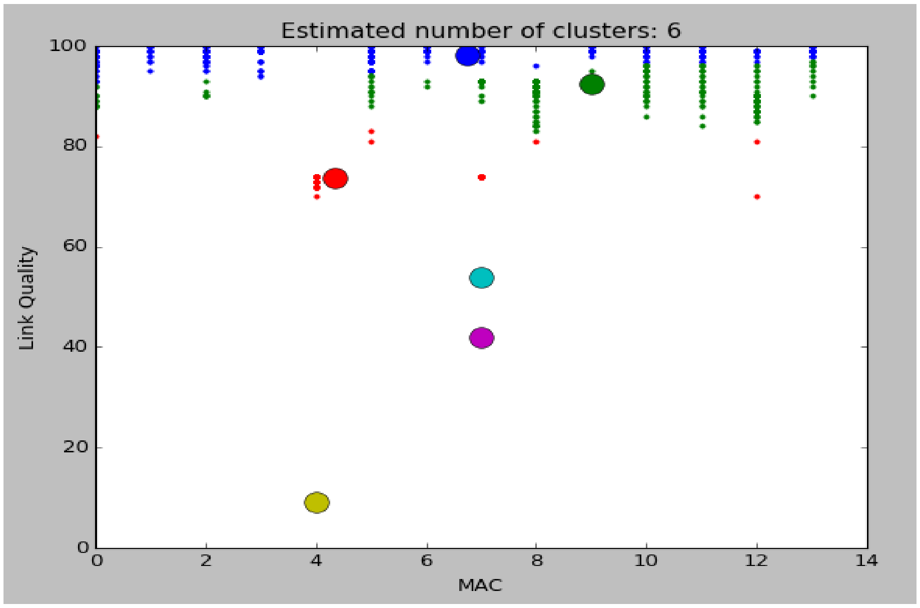


Fig. 20. Dynamic cluster detection [LQI, MAC]

6.3.2.7 Data Extraction

In the experiments, all input parameters, but the MAC address, are numerical attributes, represented by different numerical formats. However, the MAC address is a categorical attribute and represented by a hexadecimal representation of a 5-byte value. A few different techniques can be used to transform the MAC address to a numerical input value, understood by machine learning methods.

The following two techniques are used to convert the MAC address into a numerical representation.

- By converting the hexadecimal value to the corresponding 64-bit, integer value
- By collecting all the MAC addresses in the dataset and labelling each MAC address with a corresponding integer value

Both options have several advantages and disadvantages. In the first option, the MAC address is converted to a corresponding integer value and provides a unique identification for each MAC address. However, a 64-bit integer value of the MAC address doesn't provide any numerical significance in the training or prediction process. Furthermore, operating on a larger data set with a 64-bit integer value could slow down the training and prediction process. In the second option, the MAC addresses are labelled using integer values. However, the MAC-to-Label matching process has to be deterministic or preserved for the prediction process. Furthermore, if a new node joins the wireless network, it could require a re-training of the model, after adding a new MAC address to the MAC-to-Label list. Furthermore, prediction models based on the MAC-to-Label technique could require careful planning in order to prevent false negatives and to improve the generalization factor. Both above-mentioned techniques are tested to evaluate performance and resource utilization related indicators.

6.3.2.8 Sample Sets

The number of samples in a training set has a significant influence on prediction accuracy and generalization. 10 sample sets (50, 100, 500, 1k, 5k, 10k, 20k, 50k, 100k, 150k) are used to train prediction models to understand the relationship between training set size and prediction accuracy. Wireless data is collected using several segments of a complex wireless network and blended into a single data set to generate a diverse data set. The dataset also contains several subsets, including segment-specific data sets and time-specific data sets. Time-specific data sets are based on different probing intervals. Furthermore, sample data is collected over a six month period (April – October) to address some of the concerns related to seasonal effects. The objective of using several diverse datasets is to examine the influence of different localized factors on performance.

6.3.2.9 Labelling

The labelling mechanism utilized in experiments associated with Physical layer attributes is significantly different than in the experiments completed in the IEEE 802.15.4e/TSCH characteristics section. In experiments associated with Physical layer attributes, the MAC address is used as a label. Therefore, the number of potential output values are determined by the available nodes in a particular network. However, it is important to mention a potential concern, when using this labelling technique. In the sequence permutation technique, the MAC addresses in the anomaly data set belong to a legitimate node in the corresponding network. Considering the fact that a single parameter (MAC address) is used in permutation process, there is a higher probability that, the randomization process could label a legitimate flow as an anomaly; this could lead to potential higher false-positive rates. This occurrence can be minimized by using multiple parameters in the randomization process or utilizing random MAC addresses for the anomaly data set. However, anomaly could be caused by a trusted or a hostile node and such a model is able to detect both types of anomalies including IP, MAC and identity spoofing attacks with higher accuracy. Furthermore, with the use of the MAC address as a potential label, the corresponding prediction models can be used with different security controls, including data origin authentication (The outcome/result of an anomaly detection model for a particular traffic flow originated from a particular node can be used by authentication mechanism to determine the validity of the corresponding flow). In essence, prediction model should be able to determine which particular source (MAC address) owns a particular packet, based on the RSSI, SNR, LQI and LD values, retrieved from a corresponding packet.

6.3.2.10 Training/Testing

Training datasets are generated using data collected from individual wireless segments and accumulate data collected from multiple wireless segments. Different factors including training set size, classification algorithm, the aging process and the noise influence are examined using diverse data sets. Multiple prediction models are evaluated, including models trained with accumulated data sets as well as prediction models based on segment-specific-data to examine the impact on performance. Furthermore, corrupted and incomplete input data is assigned a null value (0.0) and used in the training and testing process.

6.3.3 LR-WPAN characteristics (Power Consumption)

6.3.3.1 Background

Contrary to conventional networks, wireless networks operating within constrained resources tend to generate predictable and repeatable actions. Low powered wireless devices are configured to reduce the amount of data being exchanged between nodes, by utilizing techniques such as data summarization, buffering, scheduling and local processing. Low powered wireless nodes are mainly powered by batteries and the power consumption should have a strong relationship to the number of tasks executed by the node's processor and peripherals, such as wireless transceivers and other I/O interfaces. Furthermore, the activity of a particular low powered node may be defined by a finite number of repeatable tasks. These repeatable tasks may form recurring behaviour patterns over time, and this may be directly related to the energy consumption of a low powered device. The objective of this portion of experiments is to determine the feasibility of using the battery usage of a low powered wireless node, to identify a set of operational contexts that a particular node belongs. Subsequently, to use the contexts information to design a prediction model that will determine security violations, including traffic anomalies.

6.3.3.2 Approach

The energy consumption of any device including low powered wireless devices is dependent on several factors, such as the hardware architecture, the operating system and the application running on a particular device. Furthermore, peripherals including wireless transceivers and sensor modules that are attached to a wireless device may also contribute to the energy consumption of a device. Furthermore, as previously discussed, the data transmission process consumes more energy compared to the power consumed by data processing. For instance, according to previous research on the average, 1:800 (computation to transmission) energy-ratio is used by a wireless device. Furthermore, average, low powered devices consume approximately 30 percent more energy, during transmission when compared to receiving [1]. Even though, a number of factors directly impact the energy consumption of a low powered device, data transmission and local processing contributes to a larger portion of energy consumption. Therefore, a positive relationship should be observed between data transmission/received/processed and the battery usage of a low powered wireless device.

Wireless devices, operating in IEEE 802.15.4e, may send periodic status updates using beacons. These status updates include the battery level and the PAN coordinator is obliged to accommodate a higher priority transmission for nodes with a critical battery level. Furthermore, periodic updates of the battery level may be used by routing protocols, such as RPL, to optimize the routing topology.

To accomplish the above-mentioned objectives, data collected from a simulated wireless network, operating in IEEE 802.15.4e/TSCH mode is used. Furthermore, the network is configured to send the battery level of a wireless node, using periodic messages. As previously noted, several factors contribute to the energy consumption of a wireless node; a generalized (See Below) formula is used to compute the energy consumption for each task (transmission, receive, buffer operations, computations, stack operations and interaction with I/O interfaces).

Generalized Formula = $\boldsymbol{\lambda}$*(0.5*Receiving process*U_x + 0.001*Buffer operations*U_x + 0.001*Packet encapsulation/decapsulation*U_x +0.1* Other operations*U_x + 0.8*Transmission*U_x)

($\boldsymbol{\lambda}$ = Constant to adjust the duty cycle power usage,
U_x = Data Unit)

The rate of the battery level change in a low powered wireless node is dependent on how actively the node is involved. For instance, a fully functional device (FFD), which is responsible for relaying data between peer nodes, may have elevated battery consumption. Furthermore, highly active nodes may experience a shorter battery life cycle. Maintenance procedures in a particular environment could also dictate the behaviour of the battery life cycle. For instance, if the battery maintenance procedure instructs replacement of the battery when a lower threshold is reached, a consistent, low battery value could be observed (Fixed lower battery value). However, if the battery replacement process is triggered by time (e.g. 1st of January or 1st day of summer); an inconsistent lower battery level value could be observed (Variable lower battery value).

To achieve these experimental goals, several different techniques are utilized in the corresponding experiments. Two different data sets are collected to satisfy both fixed and variable lower battery values. Several regression models are used to compare the performance, which correlate between the time and the available battery level. However, a primary objective of this portion of experiments is to determine whether the power consumption of a low powered wireless device can be used as an attribute to train a model to predict traffic anomalies in LoWSNs with higher accuracy. A few classification methods are evaluated in an effort to determine their usability as an anomaly prediction model. Several other factors, including training sample set size, labelling mechanism, false positive, false negative rate and prediction accuracy for unseen data, are also examined.

6.3.3.3 Related Attack Vector

As previously mentioned, the battery power level of a low powered node can be considered as time-series data and certain time-sensitive attacks can be detected, using a model based on time series data. For instance, replay attacks are common in both conventional and low powered networks, where the adversary uses previously captured packets to attempt unauthorized data access and disrupts data availability, by creating denial of service (DOS) attacks.

As previously noted, a transmission delay of EBs, caused by intentional and unintentional activities could hinder the synchronization of IEEE 802.15.4e/TSCH networks. Consequently, the corresponding wireless network may require re-initialization and it may lead to denial of service attacks and battery exhaustion.

In conventional networks, different techniques such as encryption, sequence number randomization, and comprehensive timestamp analysis can be used to mitigate certain re-play attacks. However, in low powered environments, computationally expensive operations such as encryption may be challenging. If a particular node generates a repeatable set of tasks, battery usage of the corresponding device can be used to predict the legitimacy of a particular operation. Time-sensitive battery usage can be used to identify replay attacks and other types of time-sensitive attacks including enhanced beacon manipulation attacks.

6.3.3.4 Assumptions

The outcome of the experiments related to battery usage relies on several assumptions and the following is a summary of these assumptions.

- Some nodes may contain actuators. However, the amounts of energy used by actuators are similar to the energy consumed by the sensor modules.
- A wireless node may operate as a sensor device or an actuator, but not as a multi-functional device (sensor and actuator).
- The amount of energy consumed by other operations, such as state transition, is negligible.
- The energy consumption rates in various hardware architectures, node OS's and I/O interfaces are consistent among different low powered nodes
- Each wireless node provides a reliable mechanism to compute energy consumption precisely in a timely manner.
- Operating Systems, running on low powered nodes, provide application program interface (API) to collect the battery usage information.
- The generalized formula used in experiments, to calculate the energy usage is reasonably formulated.

- The low powered wireless network topology allows a centralized node/device to collect data related to battery usage of each participating node.
- In a live environment, the battery usage data can be collected without exhausting the device and network resources.
- The operational guidelines are in place to recharge or replace battery sources in a systematic manner.
- The replacement batteries with similar parameter values (amount of joule, capacity, durability and other parameters) are used for individual nodes.
- Rechargeable battery drainage rates don't degrade significantly over a shorter period of time.
- Environmental changes such as severe weather conditions don't affect the battery life significantly for individual nodes.

6.3.3.5 Input parameter set

In the first set of experiments related to LR-WPAN attributes, regression analysis has been performed, to examine the relationship between battery usage and the time. In the second part of the experiment, labelled data is used with classification models to train classification models to identify anomalies.

Data is collected using a simulated wireless network operating in IEEE 802.15.4e /TSCH mode. The simulated network contains five low powered nodes and a single node is configured to transmit battery level values at fixed time intervals. 150,000 wireless packet samples are used and the following attributes are evaluated.

Parameter	Value
Battery Level (batteryValue)	0 -100 (percentage)

Battery usage	0 -100 (percentage)
Timestamp (frame.time_relative)	Milliseconds
Time Delta (frame.time_delta)	Milliseconds

Tab. 13. Input parameter details

The following diagram depicts the relationship between the total captured wireless packets and the number of packets containing details of the battery power level.

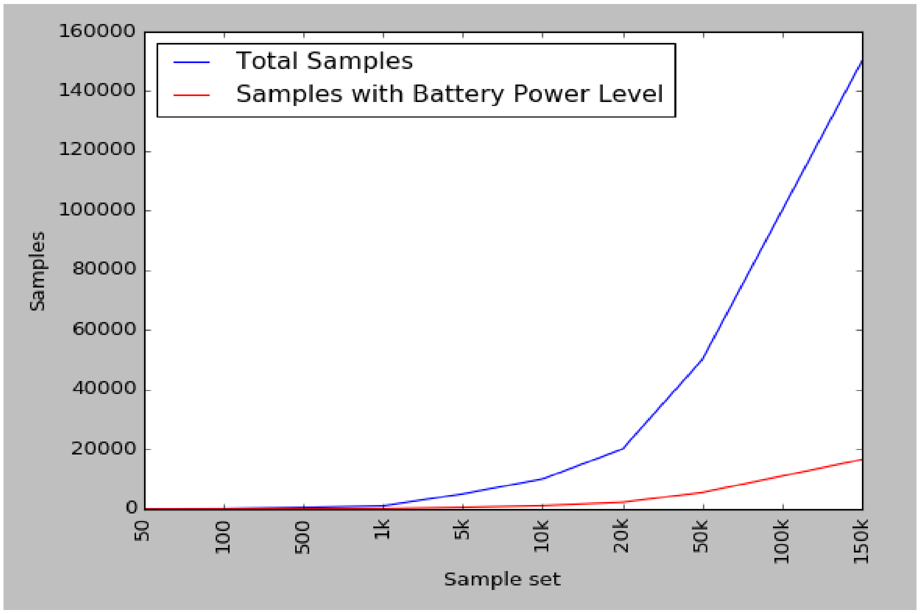


Fig. 21. Packet ratio (total vs packet with battery status)

This diagram demonstrates approximately 10 percent of packets containing details in respect to the battery power level. The captured data samples indicate the same packet is captured multiple times when a packet moves between different nodes. However, each packet contains a different timestamp. Furthermore, the battery life cycle of devices used in the following experiments is approximately, 1400 packets. In essence, the device battery is required to replace or recharge in about every 1400 operations to avoid battery drainage.

The following diagram depicts the frequency distribution of battery usage.

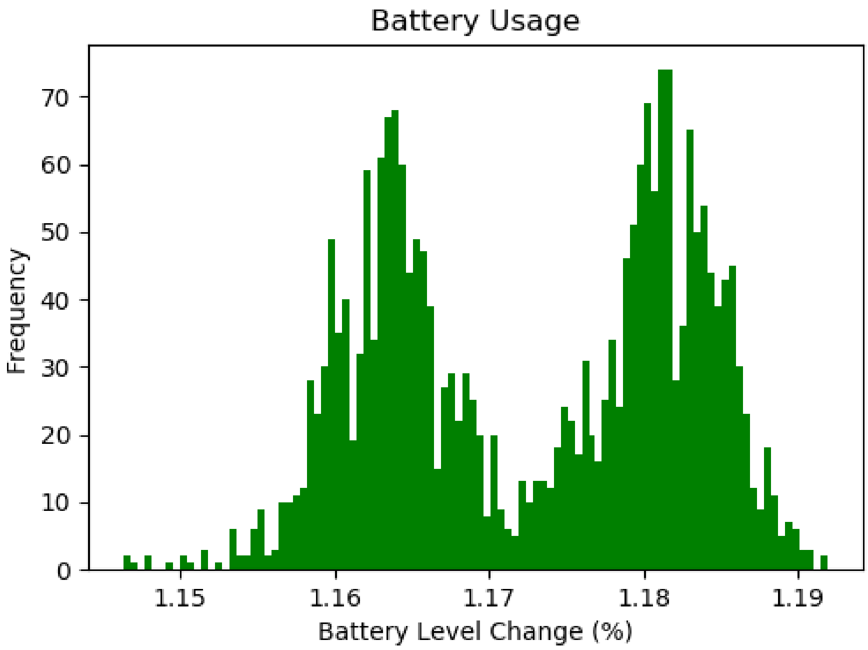


Fig. 22. Battery usage frequency distribution

This diagram confirms a normal frequency distribution for battery usage data. It is known that input parameters with normally distributed data sets, tend to produce higher prediction accuracies and generalized prediction models. Under the Experiment Results section, several prediction models, based on supervised classification methods, are evaluated.

6.3.4 Network Layer Characteristics

6.3.4.1 Background

In this proposed set of experiments, network layer attributes such as source/destination address, service type, packet length and flow control parameters (sequence numbers) are investigated to determine their usability as input parameters in creating prediction models to detect anomalies and also, to use as security controls, such as data origin authentication, in low powered wireless networks..

Source/Destination Identity: Any network operating in an open standard, such as IEEE 802.15.4, relies on certain attributes, such as source/destination identity, service type identifiers (ports) and flow control mechanism to provide efficient end-to-end communication. These attributes can be interpreted differently, in different standards. For instance, source/destination identity can be IPv4, IPv6, 6tisch, MAC or a label. These attributes are used to uniquely identify a node in a network.

Service Type: Service type defines which service/application owns a particular packet. For instance, network based on IPv4 uses 16-bit value (also known as port number) to associate a packet to an application.

Packet Length: Several different types of length measurements can be extracted from a network packet operating in the IEEE 802.15.4 standard. For instance, frame length indicates the size of a packet, including layer-2 (MAC) header. The packet length indicates the size of the packet at the network layer, including the network header. Payload size of a particular layer is calculated by reducing the header and footer (packet integrity check) portion of a corresponding layer.

Payload Data: The payload of a packet/frame carries data and belongs to adjacent, higher layer. Data generated in the application layer may include information, such as sensor readings or some command instruction set. In a conventional network, payload data could be encrypted or protected by other means. However, in low powered environments, where resources are constrained, the application layer information may transmit, unprotected. Furthermore, in low powered networks, wireless nodes are programmed to generate tiny instruction sets, such as temperature or pressure as an 8-bit value, switch control value as a binary (0 or 1), or actuator command which is an integer key from a list of key-value pairs. These application layer data may contain a repeatable, discrete, finite, set of values.

Timestamps: In a conventional wireless network operating in a star topology, the timestamp is calculated in source and destination nodes only. However, in mesh networks, where intermediate nodes are responsible for relaying packets between adjacent neighbours, the timestamp generated in intermediate nodes provide a significant amount of details regarding the data path and the flow-related information such as delays and packet loss (IEEE 802.15.4e/TSCH networks are synchronized to a PAN coordinator. However, if the network is not operated in TSCH mode, capturing application assigns a timestamp to each packet based on the captured-time). Those attributes (delays, packet loss, buffering rates) could define characteristics of the path a particular packet is taking. Intermediate timestamps can be used to discover these characteristics and consequently, different flows can be identified by evaluating timestamps of the traffic path. Once a legitimate set of flows is identified, the flow map can be used to identify normal and anomaly data.

The above-mentioned network layer characteristics, such as source/destination address, source/destination ports (service type), frame/packet payload length and flow control data (various sequence numbers), could be potential attributes to identify different packet flows correctly. In this section, the above mentioned, well-known network layer attributes are analyzed to design a potential prediction model to identify anomalies in low powered networks.

6.3.4.2 Approach

The traffic behaviour of data generated in a conventional network, with unlimited resources, is significantly different than the traffic patterns of low powered wireless networks. For instance, in a low powered environment, different techniques such as data summarization, data buffering, header trimming and local processing are used to minimize the network load. As a consequence, recurring behaviour, lower overhead and larger time intervals can be observed, in low powered wireless traffic. Similar to previous experiment settings, OpenWSN is used to build a simulated network environment operating in IEEE 802.15.4e/TSCH mode. Selected nodes are configured to generate sensor data and transmitted to a remote destination in a predetermined, time interval. Five low powered nodes, operating in a mesh topology, are used to collect 50000 wireless packets, used in corresponding experiments.

Four machine learning classification algorithms are used throughout this work to compare the performance under different conditions. However, in some experiments, ensemble models are used to improve the prediction accuracy and other performance indicators (False Positive/False Negative). Several characteristics such as:

- Classification method
- Training sample size
- Labelling technique
- Labelling delimitation
- Binary/multi-label prediction models
- Prediction accuracy with unseen (future) data
- False-positive/false-negative rates are examined

6.3.4.3 Related Attack Vector

Use of network layer attributes is common in various security controls, including firewalls, masquerading, intrusion and anomaly detection. One of the main reasons behind the common use of network layer attributes is that these attributes represent a significant amount of information about a particular data flow or communication. Consequently, these attributes are able to produce detailed information with respect to a particular flow.

A number of security attacks are based on spoofing an identity of a legitimate node. A significant portion of identity-related attacks can be identified by correctly identifying the legitimacy of a particular packet or flow. The traffic flow in low powered wireless networks is highly predictable and a repeatable, finite number of data flows can be observed. Those flows can be classified into a number of contexts, based on data origin or destination. By correctly identifying a finite set of contexts, it is possible to identify whether a particular packet or a flow, is a part of a particular context or deviation. Consequently, anomaly detection outcome can be used as an input for a secondary security control mechanism such as data origin authentication in low powered environments. For instance, if a sensor node is configured to send an alert when the temperature value reaches 16 Celsius or 23 Celsius, data initiated from the particular node can be classified into two distinctive groups (a finite number).

6.3.4.4 Assumptions

The data used in the following experiments are collected from a simulated environment. Therefore, a number of assumptions must be made to adapt experimental results to a network operating in constrained resources. A few assumptions worth mentioning are listed below.

- Experimental results generated using a small number of nodes should be able to produce comparable results within larger networks.
- A finite number of contexts can be determined in wireless networks, operating with constrained resources.
- The topology and layout of the low powered wireless network must be static, and unpredictable movements of individual nodes are negligible.
- Low powered wireless network is operating in an open standard such as IEEE 802.15.4 where, fundamental network information such as IP, labels, MAC addresses, service type (port) are readily available.
- Number of active nodes, in the low powered network is considerably unchanged over a longer period of time.
- IP, label and MAC address allocation is consistent and fixed among nodes.
- Service types, communication ports and destination addresses are fairly immutable during a longer period.
- Configuration settings such as sensor data retrieval interval for each node are fairly static.
- During the training process, the network is properly synchronized and unexpected communication delays, congestions, packet loss and retransmissions are consistent or negligible.
- Training data is collected under normal operating conditions and no adversary effects or anomaly behaviour should influence the training data collection process.

6.3.4.5 Input parameter set

The following 13 attributes of IEEE 802.15.4 are evaluated in this section.

Attribute	Description
frame.number	A relative number assigned by data capturing software
frame.time_relative	Timestamp assigned to a packet when it is received
frame_time_delta	Time difference between two consecutive receives
Frame.len	Frame length
ipv6.plen	length of IPv6 packet
ipv6.nxt	Service type
ipv6.src	IPv6 source address (6tisch)
ipv6.dst	IPv6 destination address (6tisch)
wpan.src64	Source identity used by MAC layer
wpan.dst64	Destination identity used by MAC layer
zep.channel_id	Channel id used to transmit packet (allocated by PAN)
wpan.frame_len	Layer 2 frame size
wpan.seq_no	Sequence number used in MAC header

Tab. 14. Input feature list

The above attributes are readily available for networks operating in IEEE 802.15.4 networks. 50000 data samples collected from a simulated network are used and the following set of diagrams depicts the frequency distribution of each attribute. For the following illustration, a 10000 sample set has been used. (In following diagrams, Y-axis of frequency distribution diagrams represent the frequency of occurrence of the “Title” parameter)

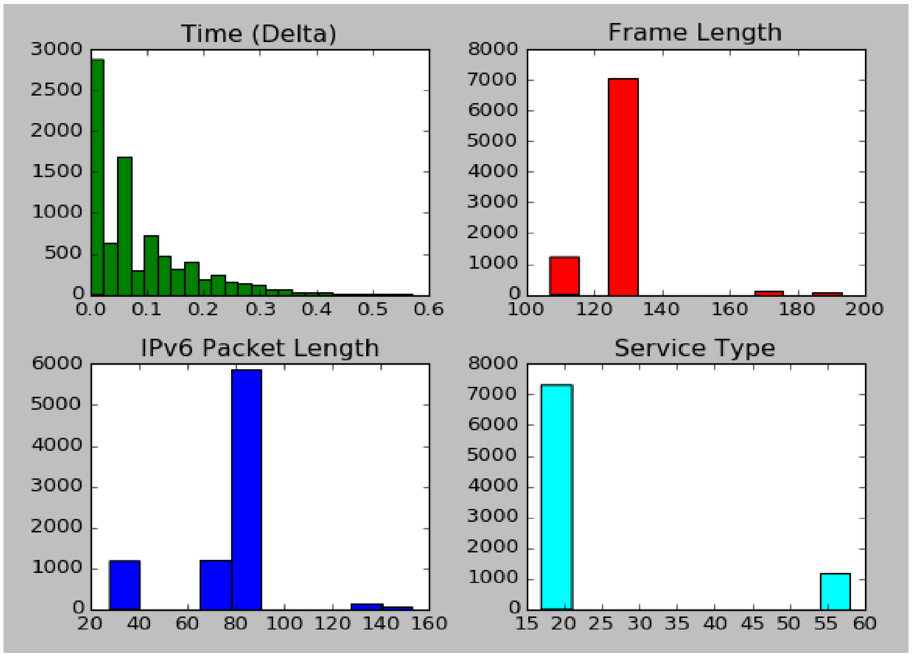


Fig. 23. Frequency distribution of input features (ordinal)

The above diagram confirms that time-delta (frame.time_delta) is a perfect fit for a half-normal distribution model. However, frame length (Frame.len), IPv6 packet length (ipv6.plen) and service type (ipv6.nxt) can be classified into a finite number of discrete classes (categorical data).

The following diagram depicts the frequency distribution of source/destination-related information, extracted from IEEE 802.15.4 packets. The source IP contains a uniform distribution of all but one address. This particular address is the source address used by PAN coordinator that sends periodic enhanced beacons (EB), containing network schedule and other network management information. In the destination address (IP) there are two values. The value with a higher frequency is the broadcast address (bbbb::1) and used by EBs. The second IP is assigned to the PAN coordinator (ff02::1a) and it confirms that the rest of the nodes communicate only with the PAN coordinator (no node-to-node direct communication on layer 3). However, as far as the layer 2 communication is concerned, each node equally (intermediate nodes have a bit higher number of packets) participates in the packet exchange process. However, in the destination MAC address table, significantly higher packet rates can be experienced for the broadcast MAC address.

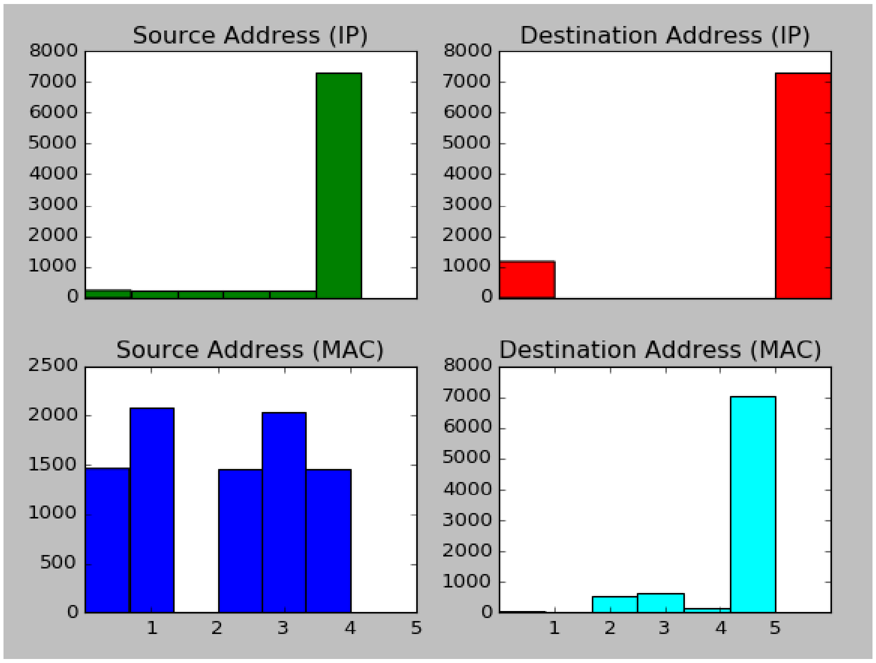


Fig. 24. Frequency distribution of input features (identity, categorical)

The above diagram demonstrates several categorical attributes. The source IP address can be classified into six groups. While five groups contain similar occurrences, the group with PAN coordinator's IP address seems to generate over 70 percent of the total distribution. Irrespective of filtering data initiated from PAN coordinator, it is possible to classify data, based on source IP address, into five uniform groups. In essence, both these parameters, and the destination MAC address can be used to identify traffic flows. However, the destination IP address doesn't seem to provide enough clusters for use in the classification process.

The following diagram demonstrates the frequency distribution of the frame number (frame.number), the time elapsed (frame.time_relative), the channel allocation (zep.channel_id) and the sequence number (wpan.seq_no). The diagram confirms that all parameters but channel allocation, demonstrates non-parametric properties of a continuous data set.

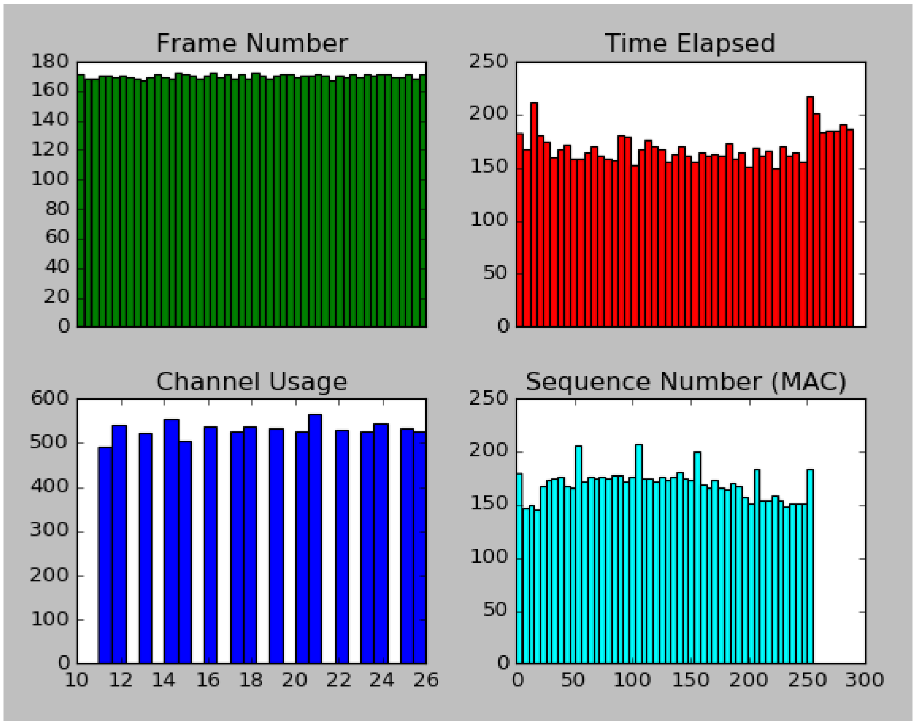


Fig. 25. Frequency distribution of input features (continuous)

Wireless networks operating in IEEE 802.15.4e TSCH mode utilizes a dynamic channel hopping mechanism to communicate between two nodes, within a specific time slot. The below frequency distribution graph describes the behaviour of channel usage (zep.channel_id) in the simulated network, used in data collection. The graph confirms that frequency is hopping between channels ranging from 11 to 26. The following diagram demonstrates the channel hopping sequence for the first 200 samples.

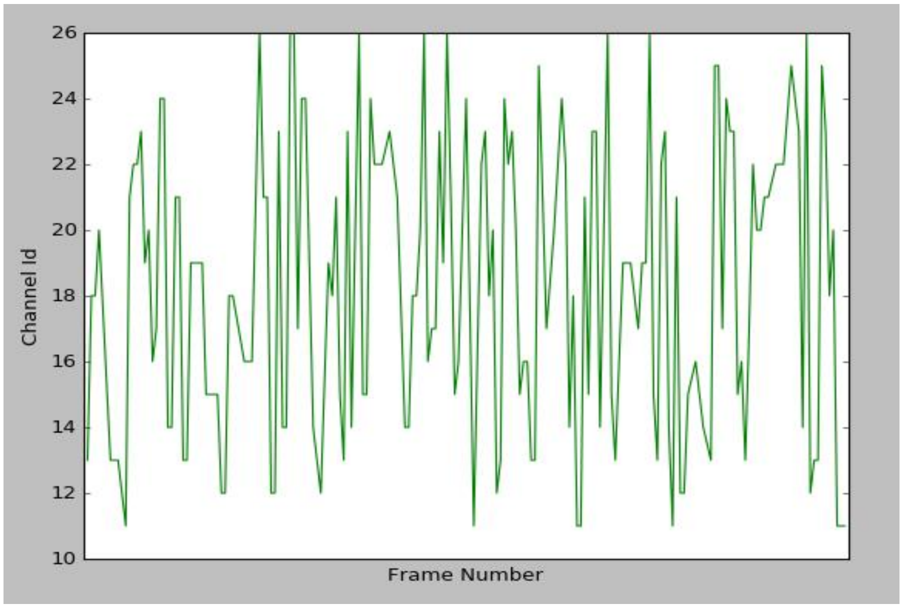


Fig. 26. Channel utilization (802.15.4e/TSCH channel hopping mechanism)

Even though TSCH mode utilizes a deterministic algorithm to compute a channel number for a particular time slot, the above diagram confirms a random channel allocation mechanism in the simulated wireless network. This may be caused by the implementation of the scheduling algorithm in the simulation software.

The above frequency distribution diagrams confirm that some IEEE 802.15.4 parameters examined in this work, can be utilized in building different types of prediction models. For instance, frame.number, and wpan.seq_no can be used in regression analysis. However, attributes such as time.delta, frame.length and packet.length demonstrate non-parametric properties of a normal distribution model and these attributes may be ideal input features for a classification model. Furthermore, node identity-related information, such as source/destination IP and MAC can be used with clustering methods to identify possible classification groups. In the following section, several different mechanisms are used to further analyze the aforementioned network layer attributes of IEEE 802.15.4.

In the following experiment, the relationship between time, frame number and the sequence number are examined. The following diagram is generated using two sets of samples. While the first set contains 100 samples, the second set contains 500 samples.

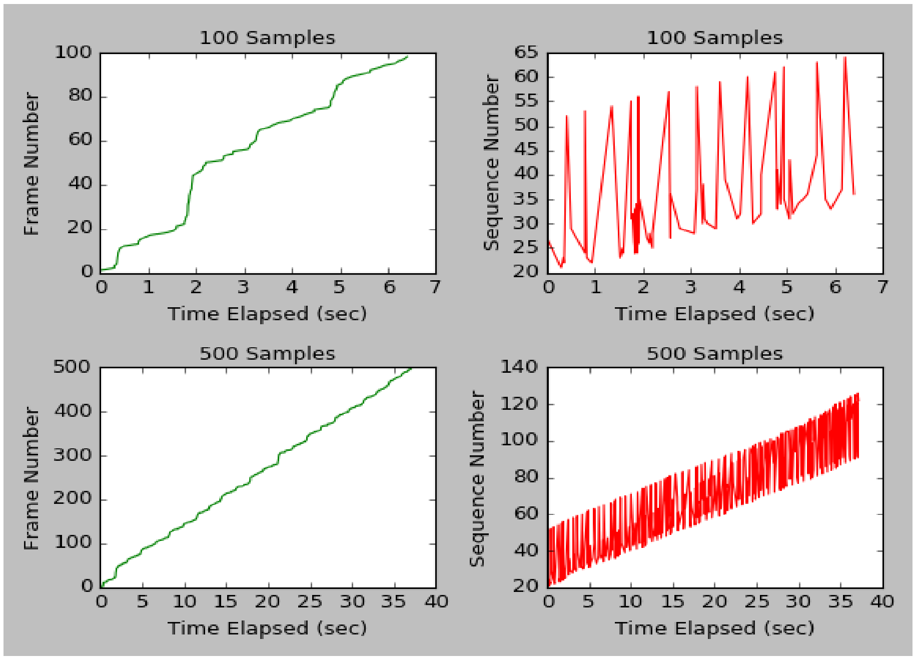


Fig. 27. Behaviour of the frame number and MAC layer sequence number

The above diagram demonstrates a strong positive relationship between frame.number and the time. In essence, the above diagram confirms that wireless networks, operating in IEEE 802.15.4e/TSCH, can observe wireless packets, in fairly fixed intervals. However, no linear relationship between MAC layer sequence number (wpan.seq_no) and the time (frame.time_relative) can be observed for these data sets. Yet, different techniques, such as moving-average can be used to transform data, before performing regression analysis with such attributes.

Several tests, such as the Pearson correlation coefficient, the Spearman's rank correlation coefficient are available to quantify the relationship between different parameters. The following Pearson correlation (also known as r) values are obtained for the above test.

Sample set	Pearson_r(Frame Number)	Pearson_r(Sequence Number)
50	0.914	0.043
100	0.981	0.428
200	0.995	0.696
500	0.999	0.905
1000	0.999	0.947

Tab. 15. Correlation coefficient comparison: frame number vs sequence number

The above table confirms that larger datasets could generate a positive, linear relationship between the sequence number and the time elapsed. However, wireless networks operating in IEEE 802.15.4e, use one byte to store the wpan.seq_no value and re-initiate to zero, when sequence number reaches the ceiling of one byte (255).

The following diagram summarizes the behaviour of four IEEE 802.15.4e attributes (packet length (wpan.frame_len), service type (ipv6.nxt), channel utilization (zep.channel_id) and time usage (frame.time_delta)).

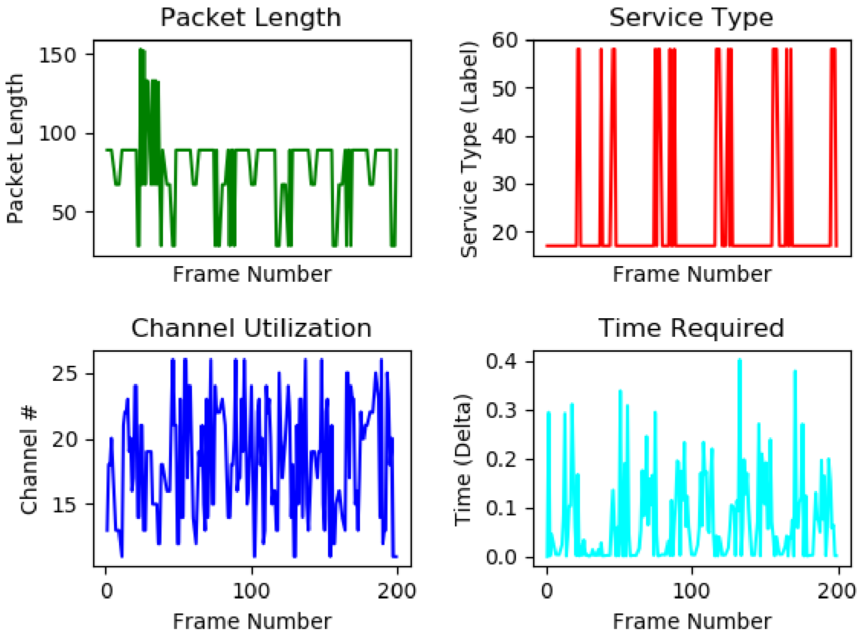


Fig. 28. Behaviour of input features against the traffic flow

The above diagram is generated using 200 packets. The first graph demonstrates a significant fluctuation in packet size at the beginning. This behaviour may be caused by an excessive amount of control data during the formation and convergence of the wireless network. This trend seems to stabilize once the network is fully formed and routing is converged. A repeatable pattern in a regular interval can be observed once the network is stabilized. The second graph describes the service types used by wireless packets. The graph confirms that the service type fluctuates between two values. Service type 17 is used by the UDP, to exchange data between nodes. Service type 58 is used by the ICMP to manage RPL and other network management related services. The channel utilization diagram confirms that the channel hopping procedure seems to be a stochastic process and doesn’t adhere to any pattern. However, in network, operating in IEEE 802.15.4e/TSCH mode, the channel is determined by few parameters, including absolute slot number (ASN) and channel offset values. Even though, the IEEE

802.15.4e/TSCH standard defines a range of channels for a particular frequency, the selection of the number of channels is at the sole discretion of particular network implementation. Therefore, the channel list, used by a particular network may provide a unique identity to a corresponding network. However, further evaluation of the IEEE 802.15.4e/TSCH channel hopping mechanism is restricted in the version of simulation software used in this work, due to limited access to the configuration settings.

The following diagram is generated using the attributes related to node identity. Both, the source IP address and the source MAC address seem to follow a certain pattern. However, the destination IP address is fluctuating between two values (PAN IP and the broadcast address).

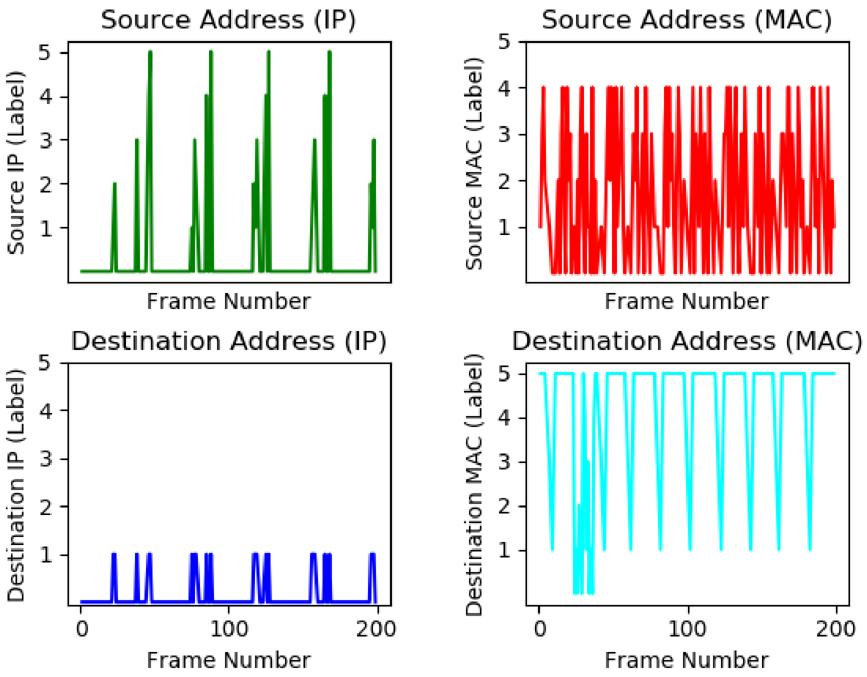


Fig. 29. Activities of wireless nodes in IEEE 802.15.4e/TSCH mode

Supervised machine learning models are dependent on labelled data. Yet, number of labels is implementation-specific. For instance, with binary classification, two labels are used to identify two different outcomes. If a particular prediction model is designed to authenticate the source node or the data origin, the output could include a number of nodes available in the network. However, a supervised learning model should identify the list of possible outputs before the training process. Instead of manually determining the number of nodes, clustering algorithms can be used to dynamically determine the possible number of classification groups. The most important aspect of the clustering algorithms is in its ability to operate on un-labelled datasets and to determine possible classes, based on the characteristics of input data. For the Network Layer experiments, a mean-shift clustering algorithm is evaluated to examine how accurately, a sample dataset would be clustered, based on the characteristics of input data. Sample data used in this experiment is collected from a simulated wireless network with five nodes. The following combinations of input parameters are evaluated to determine the potential clusters.

- Ipv6.plen, frame.len
- Ipv6.plen, wpan.src64
- Wpan.frame_length, ipv6.plen, wpan.dst64
- Wpan.src64, frame.time_delta, ipv6.plen

The following diagram confirms that several combinations of input parameters are able to successfully classify input data into five different classes. These classes can be used as possible outputs (labels) to train prediction models with multiple outputs using supervised learning.

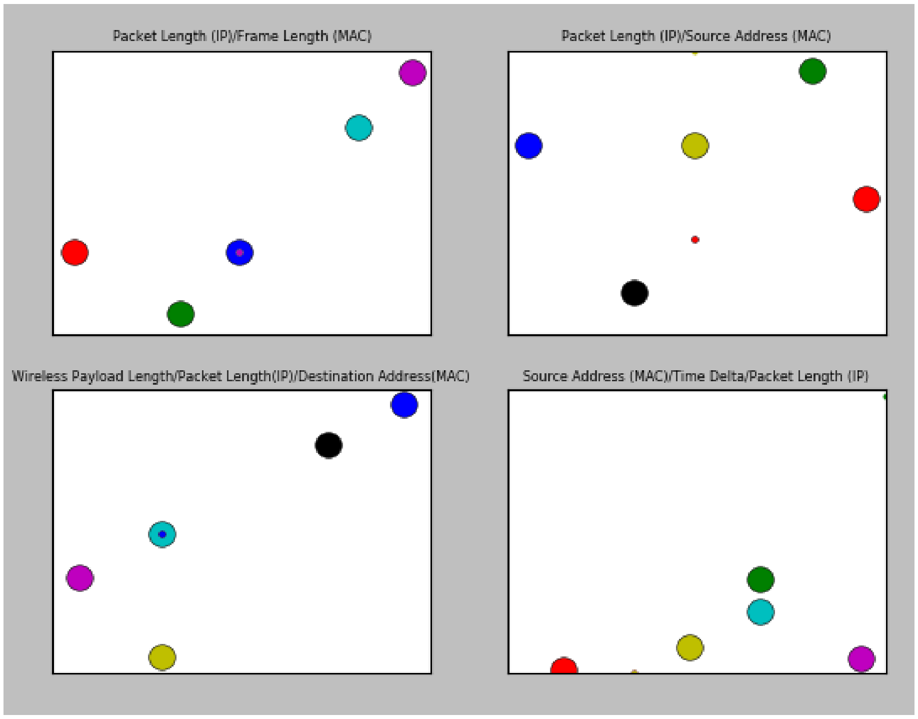


Fig. 30. Dynamic classification based on different input feature vectors

Chapter 7. Experiment Results

7.1 IEEE 802.15.4e/TSCH Characteristics

7.1.1 Linear Regression

In the following experiment, the relationship between ASN and timestamp (frame.time_relative) is investigated using several sample-sets. Correlation parameters can be used to approximate unseen time-slot values (ASN), based on a given timestamp and its usability can be further enhanced by predicting traffic anomalies, using pre-defined threshold values.

The following diagram depicts the expected values and corresponding predicted values calculated using a regression model. The graph is generated using normalized input data.

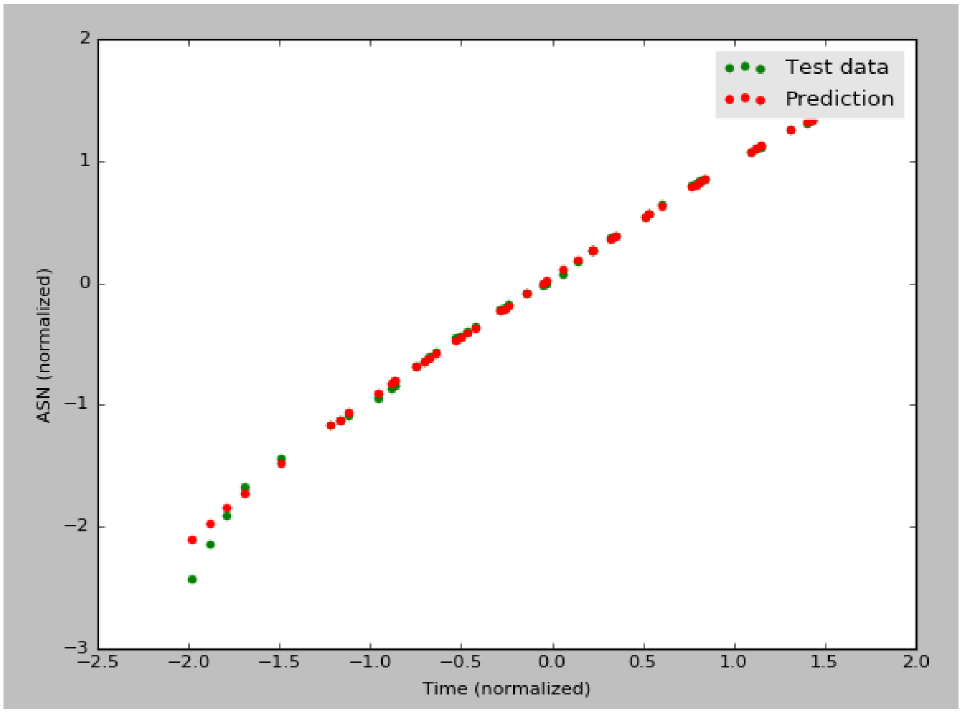


Fig. 31. Regression analysis (ASN vs Time)

The above experiment was able to produce 0.983 correlation coefficient value for the Pearson's r test. However, the correlation coefficient does not directly reflect the accuracy of a prediction model, based on linear regression and different metrics are required to determine the prediction accuracy of a model based on linear regression.

7.1.1.1 Sample Size

In the following experiment, sample size indicates the number of samples used to train regression models and 13 different sample sets (50, 100, 200, 500, 1k, 2k, 5k, 10k, 15k, 20k, 25k, 30k, 40k) are used to compare the behaviour of prediction accuracy. Several performance indicators including prediction accuracy and the correlation coefficient are used to quantify the findings.

R-squared value provides a general indicator of how closely prediction values fit in the regression line. The following chart describes the r-squared values for different sample sets.

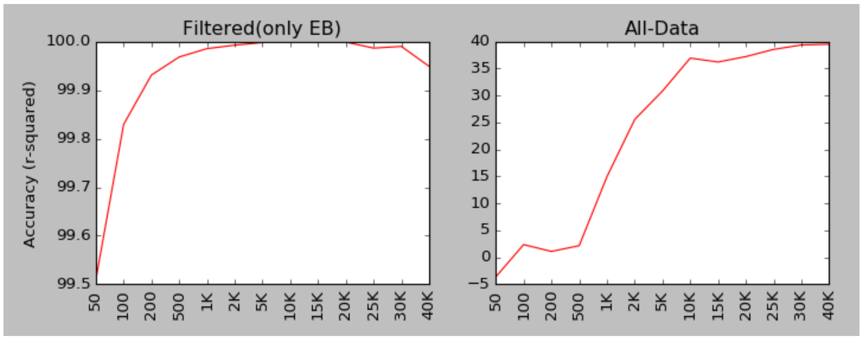


Fig. 32. Training set size vs accuracy (correlation coefficient)

The above diagram compares the correlation coefficient (r-squared) values for regression models, trained with enhanced beacons (packets with ASN number) and a model with unfiltered data. The diagram confirms a significant improvement to the accuracy, by using an only-EB based model. Furthermore, over 98 percent accuracy (correlation coefficient) can be obtained in a model trained with 500 or more samples. However, with unfiltered data, the R-squared value asymptotes to around 40 percent, regardless of sample size. As previously noted, un-filtered data contains around 60% of EBs and consequently, 40% of data in training and validation set has ASN value set to 0. This action could contribute to the lower prediction accuracy in un-filtered models.

The main objectives of this work are to identify anomalies using inherited attributes of low powered wireless networks, yet r-squared is unable to provide an adequate assessment as to the reliability of an anomaly detection model. However, metrics, such as acceptance rate, rejection rate, false positives and false negatives can be used to quantify the performance of a prediction model. These matrices are dependent on a threshold or a baseline value to classify data into corresponding groups. In the following section, threshold values are used to assess four different conditions (condition A, condition B, condition C and condition D) as defined in the previous section. More details about threshold values are discussed in the Training/Testing subsection in the Experiment Settings section.

7.1.1.2 Noise Threshold

While higher false-positive rates lead to a rejection of positive data, false negatives cause the acceptance of invalid data. False positives and false negatives' rates have a different impact, in different environments. For instance, higher attention is required for false-negative rates when a prediction model is used to control access to highly secured data. However, higher false-positive rates may also lead to severe consequences in certain applications, including personal health detection systems. Anomaly detection systems should be able to adjust these parameters to satisfy the security requirement of a particular environment.

In the following experiment, different threshold values are used to determine the behaviour of the four above mentioned conditions. The following result is based on a model trained with 1000 samples. The X-axis represents threshold variance, based on the percentage increment of the ASN number. For instance, if experiment data consists of ASN numbers from 50 to 250, one percent increment indicates a block of two ASN values ((250-50)/100).

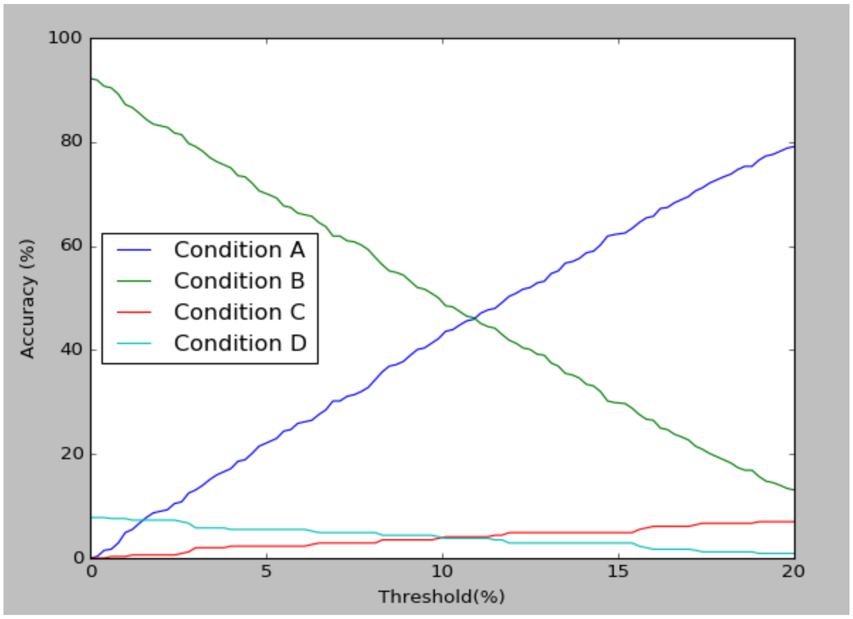


Fig. 33. Classification conditions vs threshold (boundary)

The above diagram confirms the increase of the threshold value, increases the accuracy of detection of correct ASN numbers (condition A), as expected. (Larger threshold (boundary) value increases the tolerance value for the transmit-time and it allows more time for packets to reach the destination). Condition C (false negative) also provides a positive linear relationship with the threshold value. However, the above diagram demonstrates a stronger relationship between condition A (correctly identified) and threshold value compares to condition C (false negative) and the threshold value. Therefore, it is safe to say, that, by fine-tuning the threshold value, it is possible to obtain higher prediction accuracy, with a marginal increment of false negatives. For instance, it is possible to obtain over 80 percent prediction accuracy while keeping the false-negative rate below eight percent and by applying a 20 percent threshold value. However, larger threshold values may increase the time window for an intruder to construct an unauthorized beacon frame. Therefore, the threshold value has to be adjusted to prevent a larger attack window.

The above result has been consistent with different sample sets. The following diagram was generated, using six different sample sets, and it demonstrates the relationship between the threshold value and different performance indicators (four conditions). Regularization component value 4.0 has been used to generate the following result and more details about regularization are discussed in Regularization subsection in the Experiment Settings section.

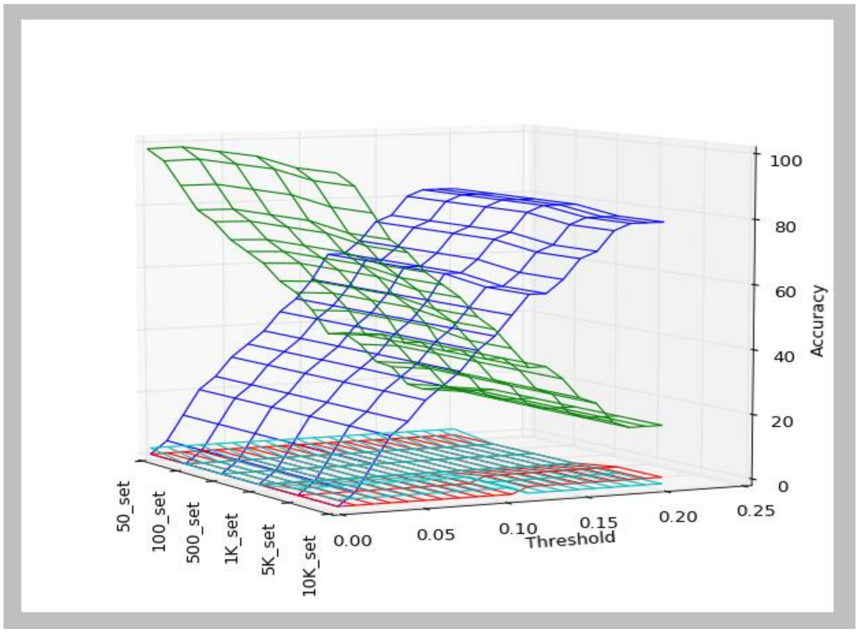


Fig. 34. Threshold vs sample size vs accuracy (corr. coef.)

The above diagram confirms consistent results for all four performance indicators (condition A, B, C and D), regardless of the sample set size, used in the training process.

In an experimental data set, the ASN number is represented by a five byte hexadecimal number and time is represented by a fraction of seconds. The data has been normalized (between {-2.0 … +2.0},) before use by linear regression models. However, in this particular model, the time delay should have a higher negative impact on prediction and it could be manually regulated. In the following, the distance calculation formula, utilizes a regularization component (λ), to produce the above discussed impact. More details about the regularization technique used in this work can be found in the section 7.3.1.10.

Regularized_distance(P_1,P_2) = $\sqrt{\lambda(t) * \left(t(p2) - t(p1)\right)^2 + \left(asn(p2) - asn(p1)\right)^2}$

The following diagram is generated using a 5000 sample set and the threshold value is set to 0.2. This demonstrates the effect of a manually adjusted regularization component.

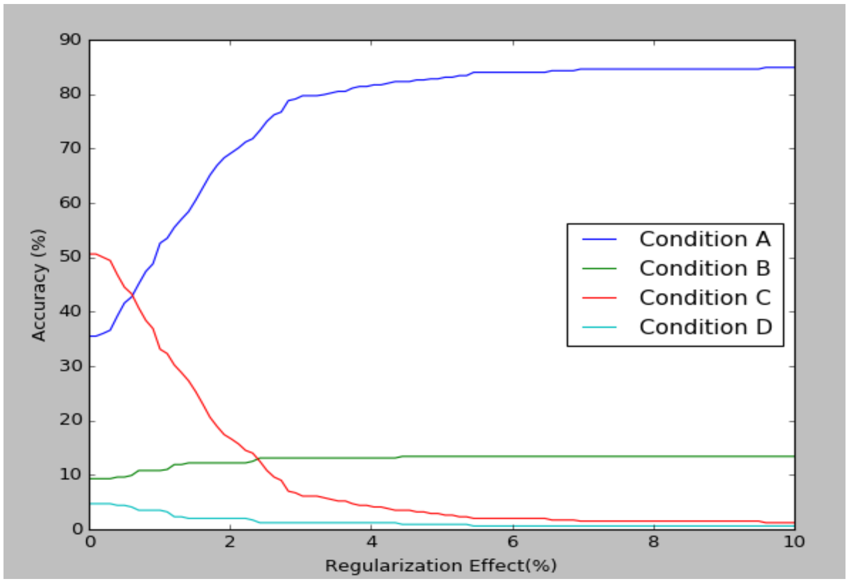


Fig. 35. Regularization vs condition accuracy

The above diagram confirms that higher accuracy can be achieved by, carefully selecting the regularization component. The satisfactory parameters for regularization can be determined by determining the acceptable levels for each condition. However, different sample sets demonstrate different accuracies. For instance, with a 5000 sample training set, using 2.7 λ(regularization value), a 78 percent prediction accuracy (condition A) could be achieved. However, to achieve 78 percent of prediction accuracy with a 10000 training set, a 3.2 of regularization value (λ) was required. Regardless of the training set size, comparable readings for all four conditions (condition A, B, C and D) could be obtained by adjusting the regularization component value in the range of 4.5 – 5.0.

The following diagram depicts the regularization effect for six training sets (50, 100, 500, 1K, 5K and 10K) and it confirms that values for all four conditions asymptote when the regularization component is calibrated to 5.0. (Threshold value is adjusted to 0.2 to generate the following diagram). In essence, the regularization component can be used to adjust the time-window value to satisfy particular security requirement while keeping the prediction accuracy above the requested threshold.

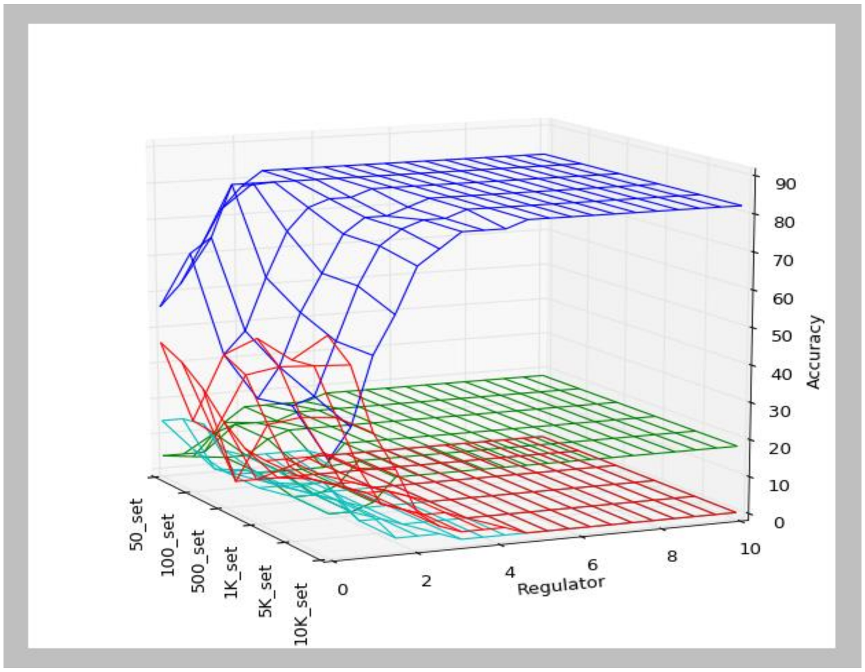


Fig. 36. Regularization vs sample size vs accuracy (conditions)

7.1.1.3 Higher-Order Polynomial

Prediction models based on machine learning are prone to both an over-fitting and under-fitting phenomenon which leads to unreliable predictions. Different techniques are used to address the above conditions including regularization, and principal component analysis. Higher-order polynomials can be utilized with different machine learning algorithms, including linear regression, to obtain models with higher prediction accuracy and generalization power. In the following experiment, the higher-order polynomials (1st order to 5th order) of ASN are used to assess the effect on prediction accuracy. The following diagram summarizes the findings of the experiment. (Threshold value 0.2 is used to generate the following result).

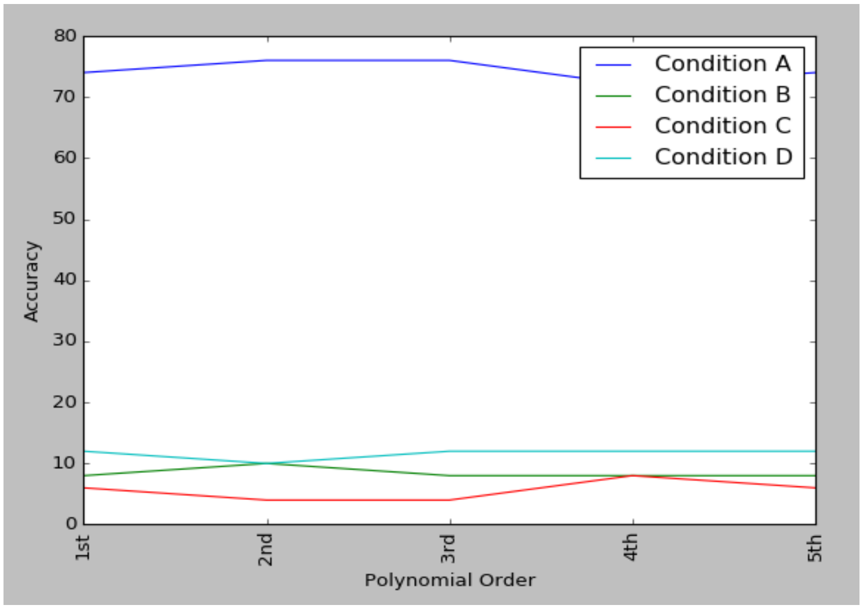


Fig. 37. Polynomial order vs accuracy (conditions)

Although training models based on higher-order polynomials, consumed substantial resources, corresponding prediction models were unable to improve the performance significantly.

7.1.2 Classification Methods

In the following, data collected from a wireless network operating in IEEE 802.15.4e/TSCH are tested with four classification algorithms (Support vector machines, K-nearest neighbours, neural networks, and decision tree) to build a prediction model using 802.15.4e attributes. Furthermore, a few different ensemble methods are also evaluated for further improvement of performance indicators, including prediction accuracy, false positive and false negative rates.

7.1.2.1 Input Parameter set

Contrary to regression analysis, where input data and corresponding outputs are a distribution of a continuous data set, classification models rely on input data retrieved from a normally distributed population. In previous experiments, frame.time_relative and slot number (ASN) have been used as input features for regression analysis. However, both frame.time_relative and time slot number (ASN) belong to continuous data sets and this is required to transform them into a normal distribution model, before being used with classification models. The following technique is used to derive two new variables (time_delta, asn_delta) using frame.time_relative and ASN.

time_delta$_i$ = frame.time_relative$_{i+1}$ - frame.time_relative$_i$
asn_delta$_i$ = asn$_{i+1}$ - asn$_i$

where $i \in (numberofsamples)$

The following diagram depicts the frequency distribution for time_delta and asn_delta.

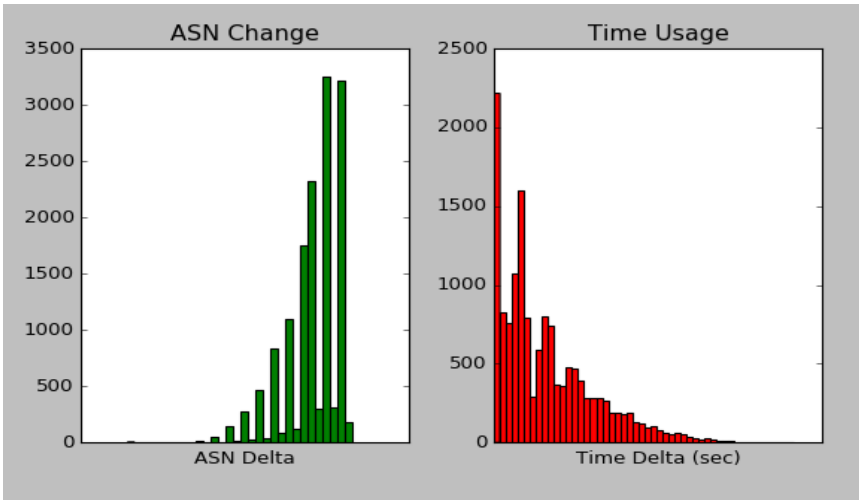


Fig. 38. Frequency distribution ASN (change) and Time (change)

The distribution of ASN-delta demonstrates non-parametric properties of a left-skewed distribution. The distribution of time-delta produces properties of a half-normal distribution model. Different techniques are available to normalize data before being used in machine learning. However, in this work, Scikit-Learn pre-processing libraries with default parameters (L2 – Norm) is used to transform data.

7.1.2.2 Sample Size

Similar to previous experiments with linear regression, 13 different sample-sets are used to investigate the relationship between prediction accuracy and training set size. All the following experiments are conducted using 80:20 train-validation ratios if not otherwise specified. Furthermore, a random sample set with 5000 samples retrieved from the same population is used to evaluate the performance indicators including prediction accuracy, false positive and negative rates.

The following experiment is performed using four default classifiers (SVM, KNN, NN, and DT) and several training sets. The experiment is based on a binary classification model and sequence permutation technique which is used for the labelling process. The following diagram compares the prediction accuracy of different training sets for each classification algorithm.

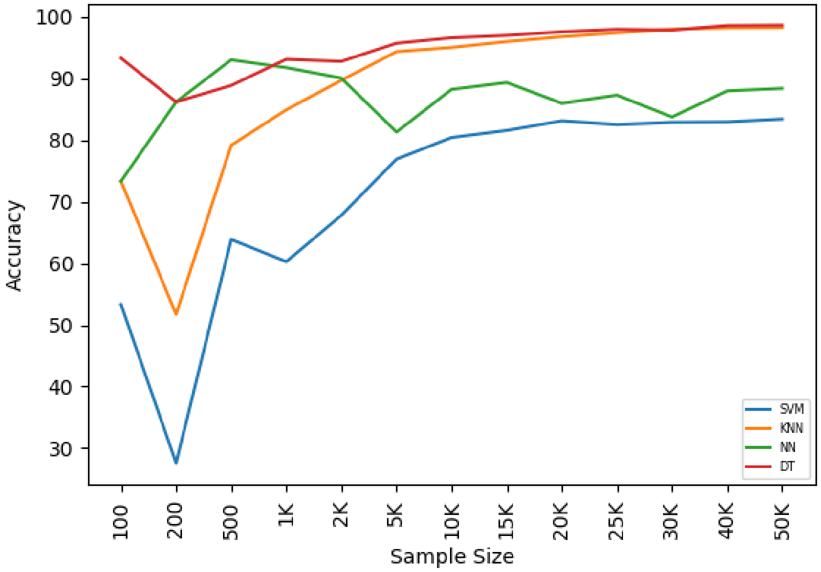


Fig. 39. Training set size vs prediction accuracy

The above diagram demonstrates a positive relationship between prediction accuracy and the training set size with prediction models based on all classification models but neural networks. However, neural networks based models are able to produce consistent accuracy with prediction models, trained with 500 or more samples. Furthermore, the prediction model based on the decision tree algorithm is able to produce remarkable prediction accuracy.

The following experiment is a continuation of the previous experiment, by replacing the SVM and NN with two ensemble methods (Bagged decision tree and random forest).

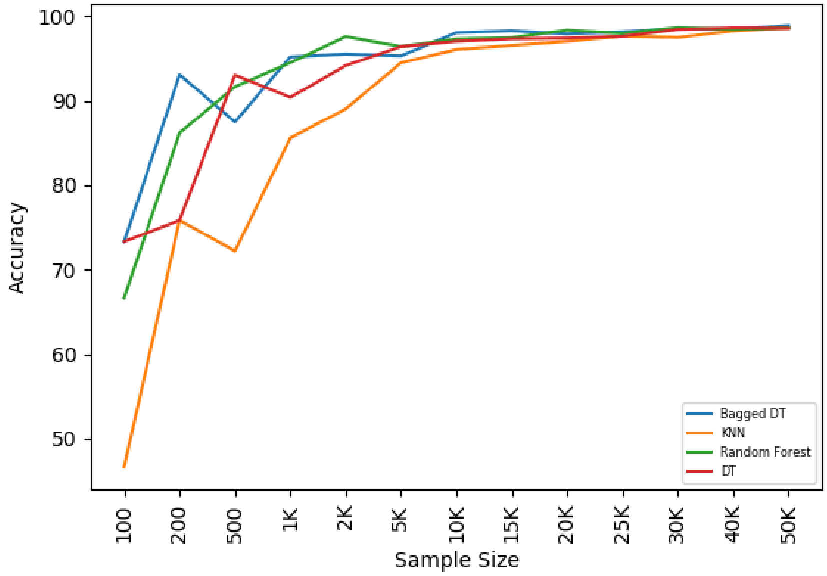


Fig. 40.Training set size (ensemble methods) vs prediction accuracy

The above result confirms a significant improvement of prediction accuracy with ensemble methods. The above result might be able to further improve by fine-tuning machine learning algorithm related, parameters and detailed evaluation of machine learning methods is beyond the scope of this work. Furthermore, the above diagram demonstrates a positive relationship between prediction accuracy and a training set size for all four classifiers.

7.1.2.3 Noise Threshold

The purpose of the following experiment is to determine the impact of an unexpected variance (noise) of data on prediction accuracy. Different causes may contribute to spikes in the data flow, including hardware failure, interference, network congestion and battery drainage. Recurrent impacts such as seasonal effects can be learned by classification models trained with larger data sets. While larger data sets produce a higher variance, smaller data sets are inclined to bias predictions. In the following experiments, a controlled random noise is introduced to valid data and performs a stress-test on the prediction model. Random noise is retrieved from a normally distributed noise-sample-set, parameterized (controlled) by the standard deviation. The following experiment is performed using four default classification algorithms, with a 5000-sample unchanged, training set. While prediction models are trained with unaltered data, test data is modified with a random-noise to examine the impact. The following diagram demonstrates the corresponding result.

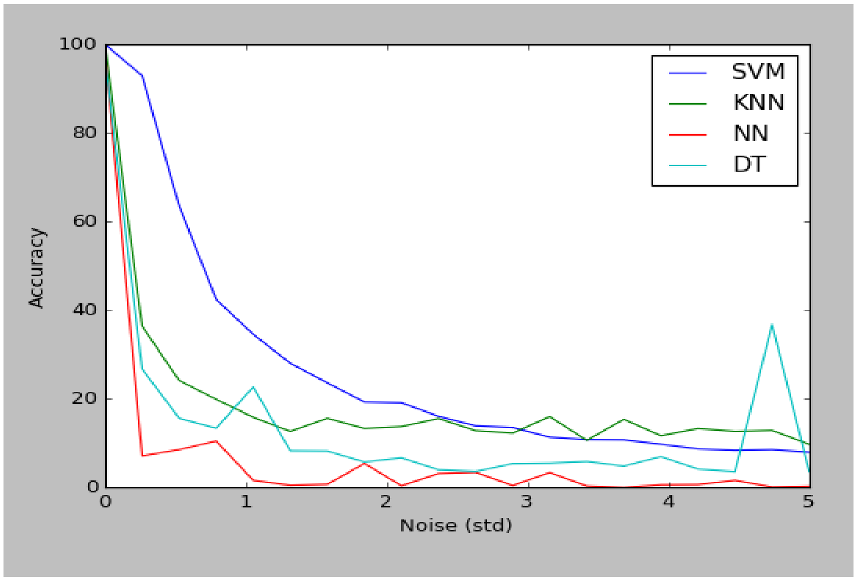


Fig. 41. Data variance vs prediction accuracy

The above diagram demonstrates a significant drop in prediction accuracy with a higher variance of test data and it confirms that prediction models, based on all four algorithms, are unable to identify unseen data with higher variance, accurately. Surprisingly, the model based on Decision Tree has produced a significant deviation from the expected result with higher noise. Hyperparameters associated with Decision Tree or algorithm implementation of Sci-kit Learn may have contributed to this anomaly. However, to come to a more specific conclusion to such a behaviour, further investigation regards to implementation of Decision Tree algorithm in Sci-kit learn may be required.

The following diagram is generated, using prediction models based on the Decision Tree. (The objective of this experiment is to observe the behaviour of prediction accuracy against data variance and sample size. Therefore a single classifier is utilized in the experiment). The purpose of the following experiment is to examine the relationship between the variance of unseen data and the prediction accuracy of models trained with different sized training sets.

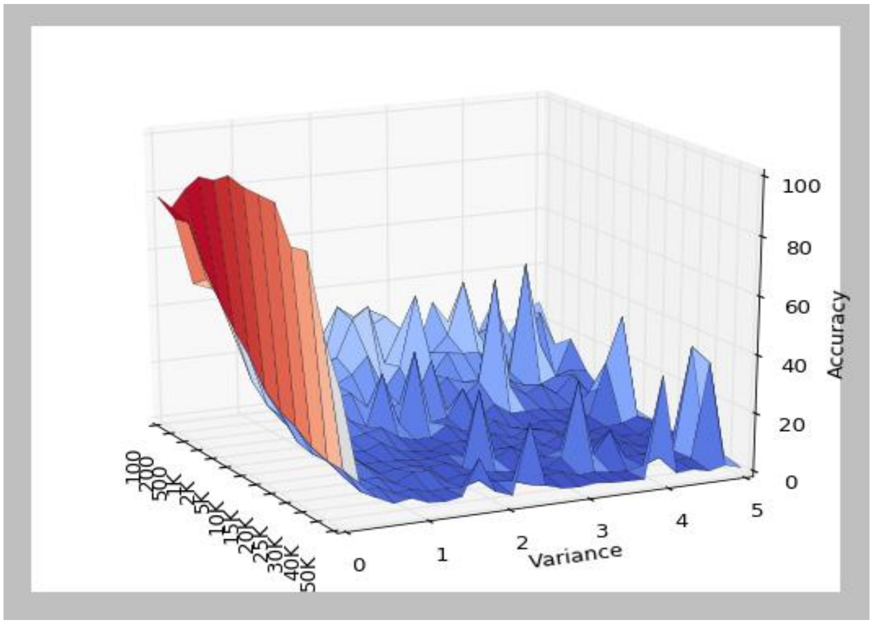


Fig. 42. Data variance vs training set size vs prediction accuracy

The above diagram confirms that regardless of the training set size, prediction accuracy drastically diminishes with a higher variance of unseen data.

In the previous experiment, the prediction accuracy, against the variability of unseen data is examined. In the following experiment, training data is modified with random controlled noise to simulate various, unspecific impacts. The random noise generation technique as previously discussed in the subsection of Labelling in the section of Research Methods & Design; the variability of noise is controlled using the standard deviation of a normally distributed, noise sample set. The following experiment is performed using a 5000-sample training set modified with a random noise ranging from 0.0 to 1.0 standard deviation. The following diagram demonstrates the corresponding result.

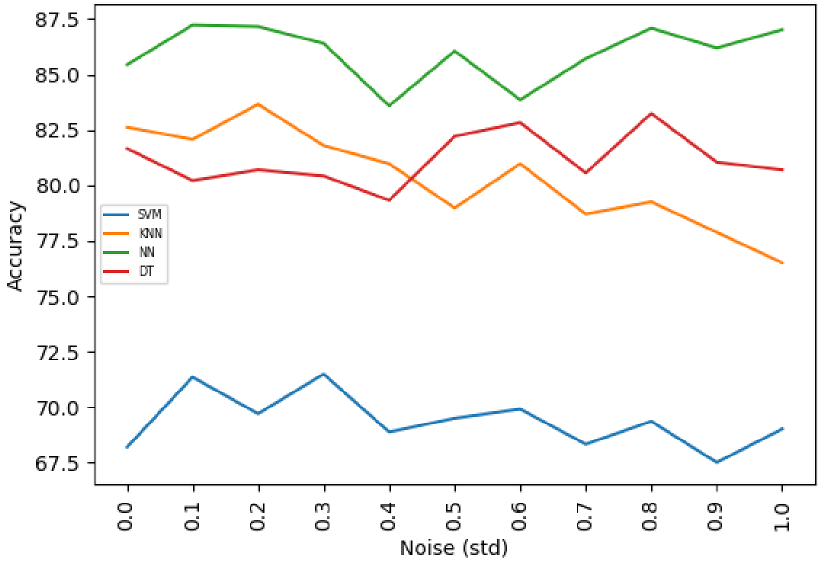


Fig. 43. Data variance vs classifier vs prediction accuracy

The above result indicates a decrement of prediction accuracy, with a higher variance of input data.

A follow-up experiment is completed to investigate the behaviour of ensemble methods with regards to higher variance of input data and result confirms that some ensemble methods are able to improve prediction accuracy, marginally, with higher-variance data sets and the above trend continues, with larger training sets

The following diagram is generated using prediction models based on the random forest algorithm. The objective of the corresponding experiment is to examine the relationship between input data variability (noise), training set size and the prediction accuracy.

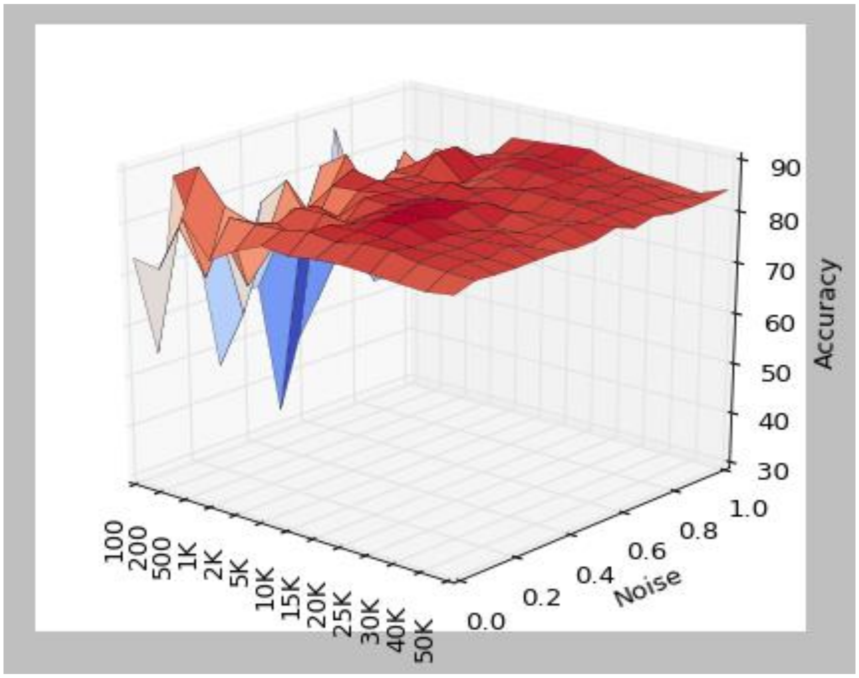


Fig. 44. Data variance vs training set size vs prediction accuracy

The above result demonstrates an inconsistent relationship between prediction accuracy and input data variability (noise) for models trained using smaller training sets. However, models with larger training sets are able to produce consistent accuracy, regardless, of input data variability.

7.1.2.4 Positive/Negative Ratio

Supervised learning methods rely on labelled data to build prediction models. While binary classification models are based on two output classes, multi-label classification models consist of three or more finite numbers of output values. Most anomaly detection models are built on binary classification models, where a particular packet or flow is identified as normal or an anomaly. The reliability, effectiveness and generalization of a prediction model based on machine learning are dependent on several factors, including proper selection of input features, training set size, training set selection criteria, classification algorithms and label ratio (number of normal vs number of anomalous samples). For instance, if a particular training set comprises only normal data, it is quite a challenge for a classification algorithm to determine baseline values for normal and anomalous data. Therefore, it is important to use a training set representing each output class that has an acceptable data ratio.

In the following experiment, label ratio for a binary classification model is examined. For the following experiment, a 5000-sample training set is used and a positive (normal) sample set size is regulated from 10 percent to 90 percent of total training samples. The following diagram describes the findings.

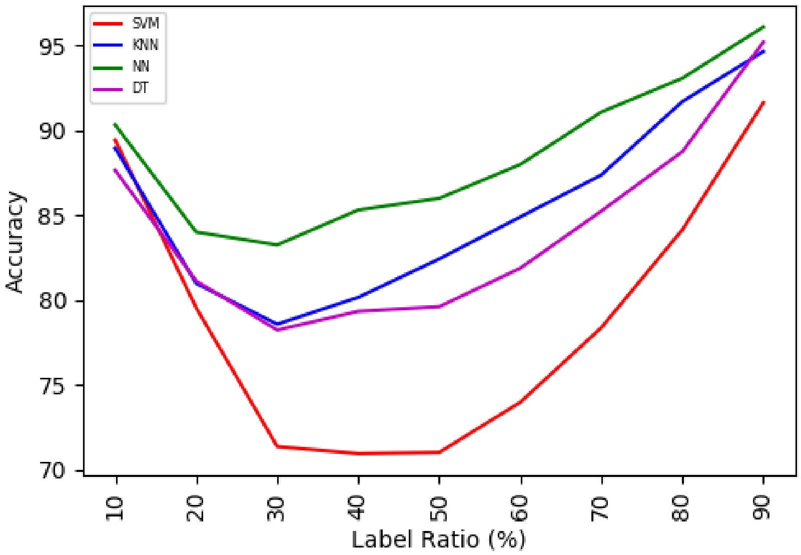


Fig. 45. Label ratio vs prediction accuracy

Surprisingly, in the above experiment, prediction models with smaller or larger positive sample ratios were able to produce higher accuracy, and sample sets with equally distributed labels tended to produce lower accuracy. A smaller training set size used in the previous experiment maybe a contribution to the above results. The following experiment is conducted using several data sets to further investigate the previous experiment result, and corresponding results are depicted below.

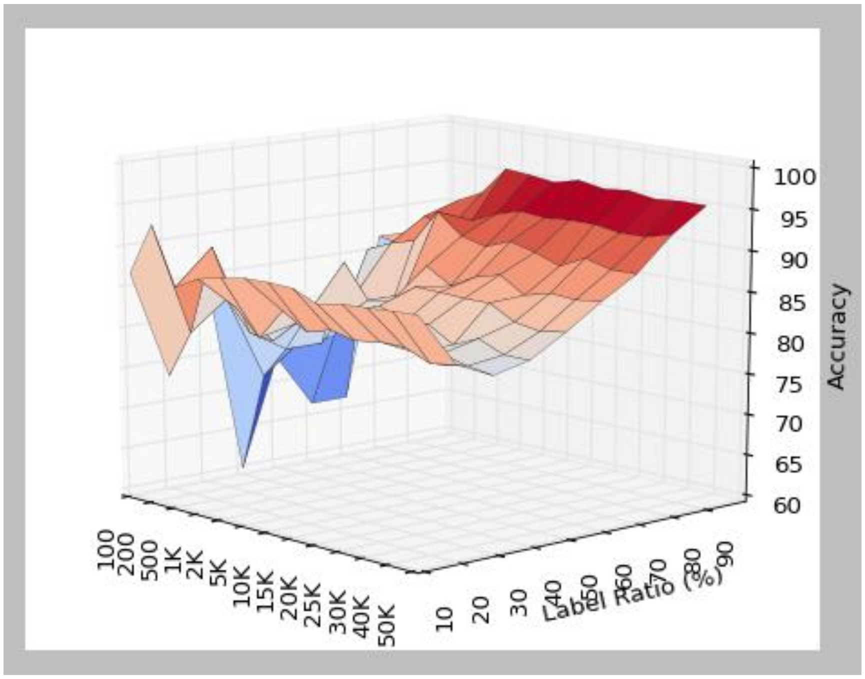


Fig. 46. Label ratio vs training set size vs prediction accuracy

The above result confirms that regardless of the sample set size, training sets with non-equally distributed labels are able to produce higher accuracy in binary classification models. Models based on higher label ratio might be able to classify data with higher number of particular label with a higher accuracy and that may contribute to the above behaviour.

7.1.2.5 Model Aging Process (Retention Factor)

An important characteristic of a reliable prediction model is strong generalization power. In essence, a reliable prediction model should be able to maintain higher prediction accuracy and other metrics such as false positive and negative rates that are fairly consistent in longer periods of time. In the following experiment, the prediction model's aging process is investigated. In this experiment, a number of test-sets with 2000 samples in each, collected in predetermined time intervals over prolonged periods, are used. Four default classification algorithms are used to train prediction models with a training set of 5000 samples, collected from the same network, at time zero (t0). Corresponding results are depicted in the following diagram.

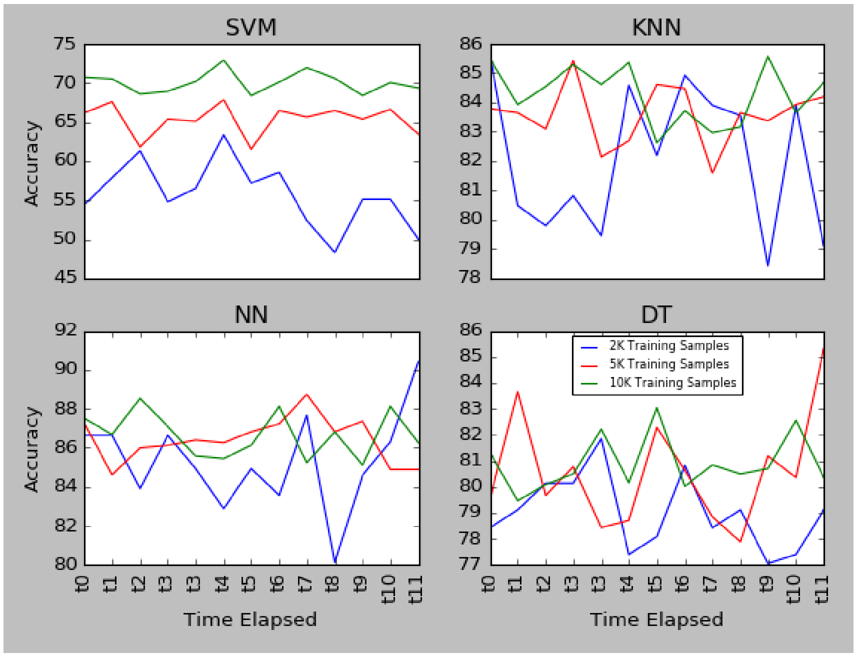


Fig. 47. Prediction model aging process

The above diagram does not demonstrate consistent behaviour with any classification model. However, all four classification models are able to maintain comparable, prediction accuracy during the test period. Furthermore, models trained with smaller datasets (blue line), demonstrate a higher variance of prediction accuracy and models trained with a larger data set (green line), gravitate to more stable accuracy rates.

7.1.2.6 Filtered/Non-Filtered

As previously noted, several types of packets including the enhanced beacons (EB), management, and application-specific data can be found in wireless networks operating in IEEE

802.15.4e/TSCH. The simulated wireless network used in this work contains around 62 percent EBs, 12 percent routing-related data, and the rest, including application-specific, unicast traffic. In previous experiments, only packets containing time slot numbers (ASN) are used to produce prediction models. However, a packet ratio of a particular network is dependent on several factors including number of nodes, node configuration, TSCH, RPL and other configuration settings. In the following experiment, the comparison between prediction models trained with filtered and unfiltered data is examined. In an unfiltered data set, non-numerical or unavailable data is replaced by the value 0. Four default classification methods and several sample sets with a different number of samples are used to examine any relationship. The corresponding result is depicted in the following diagram.

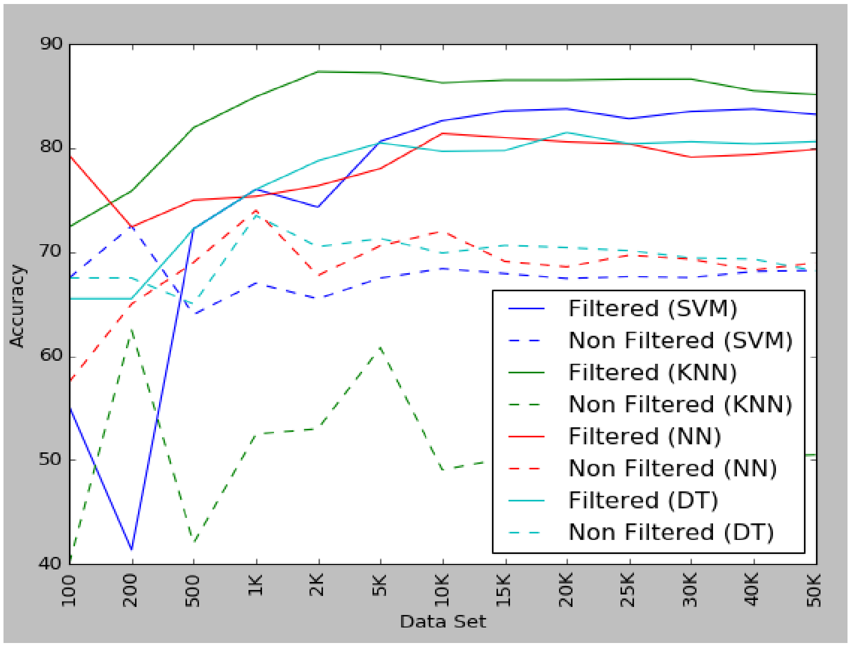


Fig. 48. Filtered, non-filtered data vs prediction accuracy

The above result demonstrates inconsistent relationships among classification models. For instance, a prediction model based on

KNN with filtered data is able to achieve the highest accuracy amongst tested classifiers. Yet, the same classifier produces the lowest prediction accuracy, with unfiltered data. However, three other classifiers (SVM, NN, and DT) are able to improve the prediction accuracy by around 15 percent, with the use of a filtered data set. KNN algorithm is based on determining the K nearest neighbours using Euclidean distance, while other algorithms use complex optimization methods to determine the decision boundaries. These characteristics may contribute to the above behaviour.

7.1.2.7 False Positive/Negative Rates

Although prediction accuracy provides a good indication about the effectiveness of a prediction model, a few other attributes, including false positive/negative rates, resource utilization and model aging rates, also contribute to selecting an acceptable solution. Acceptable rates for false positives and negatives are very implementation-specific. In the following experiments, the relationship between false positive/negative rates and several attributes, including training set size, the model aging process and noise factors are investigated.

In the following experiment, the relationship between false positive/negative rates and a training sample set size is investigated. Similar to previous experiments, several sample sets, with different numbers of samples, are trained, using four default classification algorithms. The sequence permutation technique is used to produce the negative data set used in the following experiment. A data set with 5000 samples extracted, from the same population, is used to determine the false positive/negative rates. Experimental results are depicted in the following diagram.

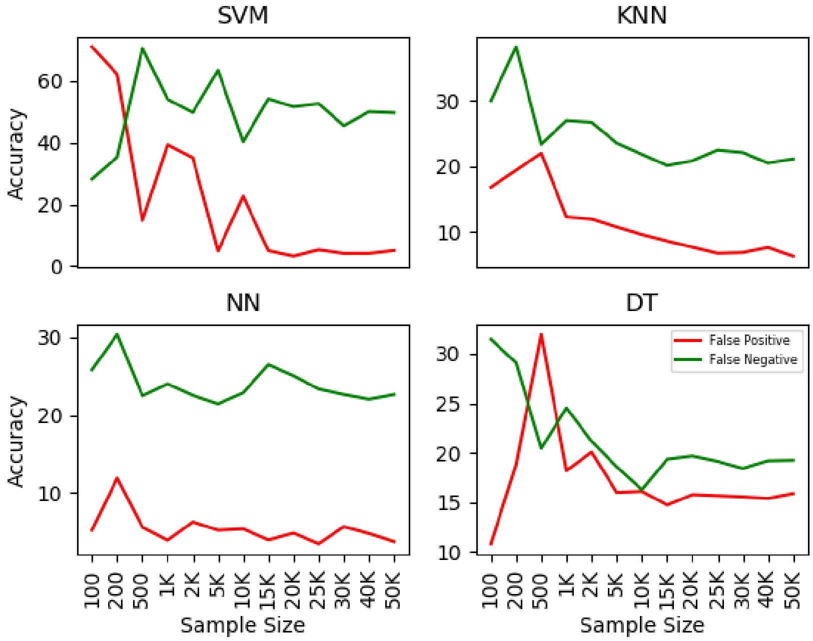


Fig. 49. Training set size vs false positive/negative rates

The above diagram demonstrates the different characteristics of false positive/negative rates among classifiers. For instance, models based on SVM, produce extreme false positive/negative rates with prediction models trained using smaller data sets. However, neural networks models are able to keep false positive/negative rates, at a consistent level, regardless of training set size. Models based on both DT and KNN are able to minimize false positive/negative rates in larger training set models.

The objective of the next experiment is to investigate the relationship between the model aging process and false positive/negative rates. The prediction models are trained with 5000 samples collected at time t0, and several sets with 2000 samples each, collected during a prolonged period, using fixed time intervals that are tested for false positive and negative rates. The corresponding result is described below.

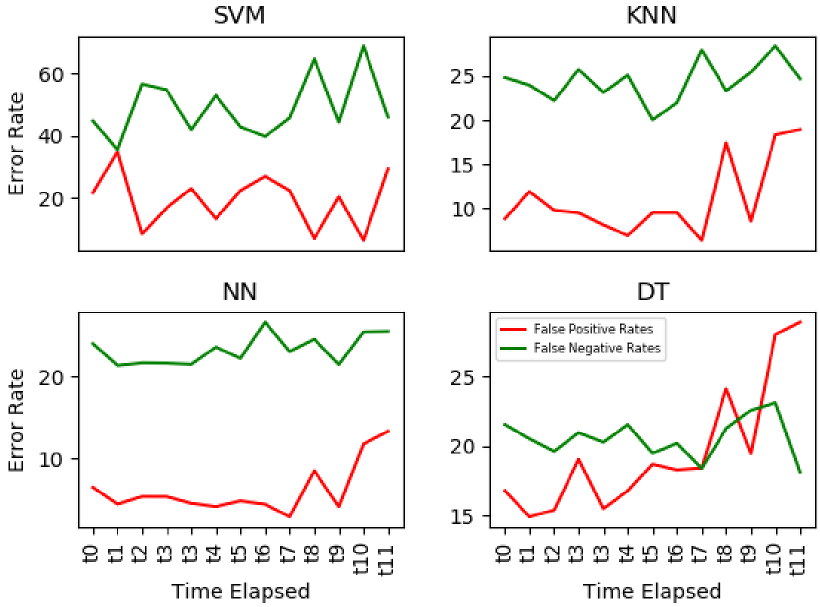


Fig. 50. Time elapsed vs false positive/negative rates

The above result indicates that there were no significant changes in false positive and negative rates over time in all four models but DT. However, model based on the Decision Tree algorithm demonstrates a positive relationship between false positive rate and the time elapsed while maintaining a consistent false-negative rate error during the test period.

The purpose of the final experiment in this portion of the work is to examine the relationship between false positive/negative rates, training set size and noise thresholds. The following diagram is generated using several training sets, with adjustable noise threshold values, to determine the corresponding false positive/negative rates. The prediction model is based on the random forest and these false positive/negative rates are calculated using 5000 samples.

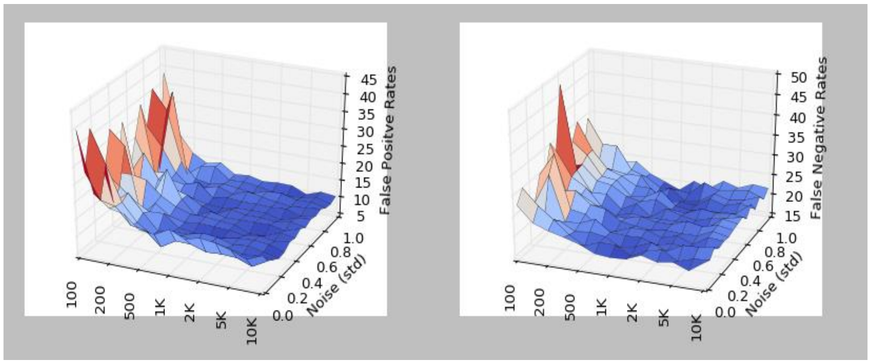


Fig. 51. Data variance vs sample size vs false positive/negative rates

The above diagram confirms that regardless of the noise variance, prediction models trained with larger sample sets are able to maintain low false positive/negative rates.

7.1.2.8 Time/CPU/Memory Utilization

It is important to examine the resource utilization costs by different classification models. In the following, four default classification algorithms are examined for resource consumption. Three performance indicators (CPU usage, time, memory usage) are tested and the following diagram describes the corresponding result. The usage values of the following diagram are based on a training process of a prediction model using 10000 samples and ASN-delta and time-delta as input parameters. The usage values are further generalized by averaging the result of 10 consecutive training processes. Y-axis of the following diagram describes the usage of individual attributes (CPU: Percent, Time: Seconds, Memory: Megabytes). For instance, a model based on SVM has used 30 percent of CPU usage, 120 seconds time and 58 MB of memory). However, prediction models are trained in centralized systems and resource usage values can be further reduced by using modern state-of-art techniques such as cloud based systems and distributed machine learning tools.

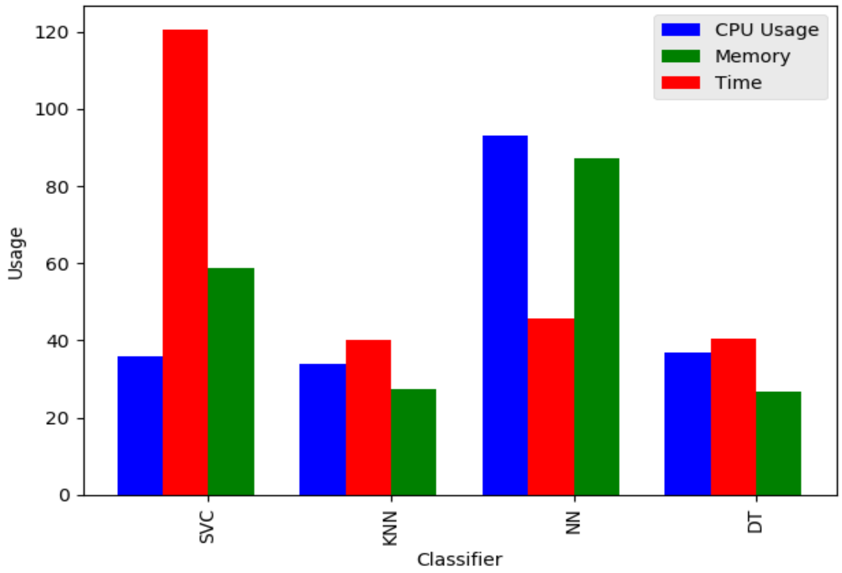


Fig. 52. Resource utilization by classifiers

The above diagram confirms that the prediction models based on support vector machines require a significantly longer period for the training process. In the meantime, the neural network based models utilizes higher computational power during the training process. However, with protocol specific attributes, models based on K-nearest neighbours and decision tree, are able to utilize fewer resources while providing similar prediction accuracy.

7.1.2.9 Result Summary

In the above experiments, IEEE 802.15.4e/TSCH attributes were evaluated to build a model to identify traffic anomalies in low powered wireless networks. In the above experiments, the ASN and timestamp of a corresponding packet were examined for potential features to a machine learning based prediction model. Several characteristics including input parameters, labelling mechanisms, training set size, input data variance, and model aging processes were investigated. The most important findings of these experiments are listed below.

- IEEE 802.15.4e/TSCH Characteristics
  - Absolute slot number of a wireless network operating in IEEE 802.15.4e/TSCH can be found in enhanced beacons.
  - Each EB encapsulates active time slot number (ASN) in the payload portion of the information element in 802.15.4e/TSCH packet header.
  - A significant portion of IEEE 802.15.4e/TSCH packets is enhanced beacons. (Simulated data used in this work comprises over 60 percent EBs).
  - Absolute slot number and timestamp produce a strong positive correlation (Pearson r: 0.983).
  - Non-parametric tests indicate that absolute slot number change (ASN_{t+1}-ASN_t) and corresponding time usage ($time_{t+1}$-$time_t$) values demonstrate strong properties of a normal distribution model.
  -
- Regression Models
  - With the use of manually adjusted threshold values, regression data can be classified with higher accuracy (over 78 percent) while keeping false positive/negative rates low.
  - Regularization techniques could be used to improve prediction accuracy while maintaining low false positive/negative rates.
  - Input parameters based on higher-order polynomial, are unable to improve prediction accuracy, significantly.
- Classification Models

- Prediction models based on delta values (difference) are able to produce higher accuracy with larger training sets (e.g. a prediction model based on decision tree was able to achieve 98 percent accuracy with 20000-sample training set).
- All tested classification algorithms but SVM, trained with 500 or more samples were able to predict anomalies with 85 percent or more accuracy.
- Ensemble methods demonstrated significant improvement over some classification algorithms (e.g. RF over SVM) and were able to obtain over 98 percent accuracy with larger training sets.
- Higher noise in unseen data may drastically reduce the prediction accuracy. For example, unseen data with an average of one standard deviation variance may reduce prediction accuracy by 50 percent. (A consistent behaviour could be observed, regardless of training set size).
- However, if the prediction model is trained with higher variance data, prediction accuracy can be improved with unseen volatile data. For instance, a model based on DT was able to achieve 80 percent accuracy using data with one standard deviation (STD) noise injected.
- The higher variance of input data has a minor impact on prediction models trained with larger data sets. For instance, prediction models trained with 50000 samples were able to reach 85 percent consistent accuracy, regardless, of the noise level (tested for 0.0 to 1.0 STD)

- Labelled data ratio has a significant influence on prediction accuracy for binary classification models. For instance, models trained with very small or very large positive sample ratios were able to produce higher accuracy, compared to prediction models trained with balanced sample ratios.
- Prediction models trained with delta values (difference) seems to have a higher retention factor. For instance, all four tested classification models (SVM, KNN, NN and DT), trained with 10000 samples were able to maintain prediction accuracy during the test period.
- Models based on SVM incline to produce inconsistent false positive/negative rates during the aging process. However, models based on KNN, NN and DT were able to maintain false positive/negative rates throughout the test period.

7.2 Physical Layer (Wireless) Characteristics

7.2.1 Classification Methods

7.2.1.1 Input Parameters

In the following experiment, each input parameter (RSSI, SNR, LQI and LD) is individually examined to identify the influence on the prediction accuracy. Four classification algorithms are used to train prediction models, using 18800 samples collected from a single wireless segment. Furthermore, for the following experiment, the MAC address is used to identify different output classes (labelling). The following diagram describes the summary of findings.

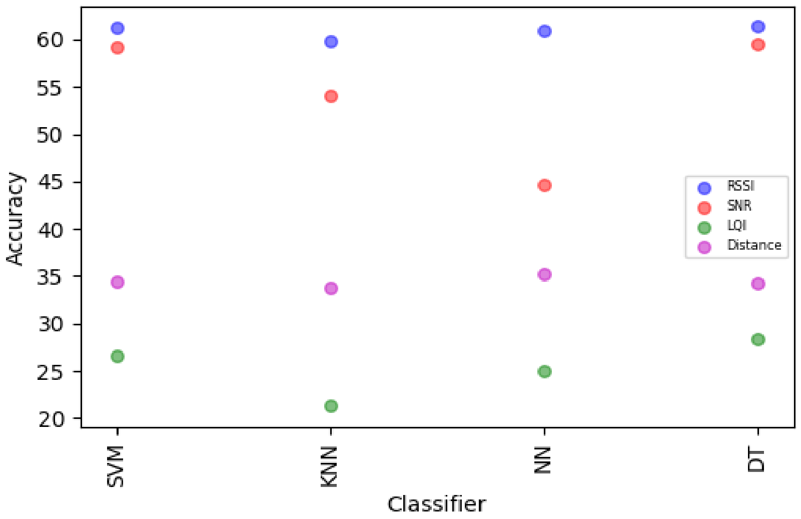


Fig. 53. Prediction accuracy of models based on individual features

The above diagram demonstrates marginally higher accuracy for prediction models based on RSSI and SNR, using DT as a classification algorithm. Yet, the above accuracy rates may not be satisfactory to use in an anomaly detection model. Furthermore, compared to the accuracies of most previous experiments, the above experiment result seems to produce much lower accuracy. However, it is important to emphasize that the accuracy of most previous experiments are based on binary classification models and the above results based on a classification model with 14 possible outputs (training data is collected from a wireless segment with 14 wireless nodes and the mean prediction accuracy is around seven percent (1/14)). In essence, the above model is able to improve prediction accuracy by over 50 percent. However, further experiments will be performed in the following section, to determine the relationship between prediction accuracy and the number of nodes in the network.

The following diagram depicts prediction accuracy rates when parameter pairs are used.

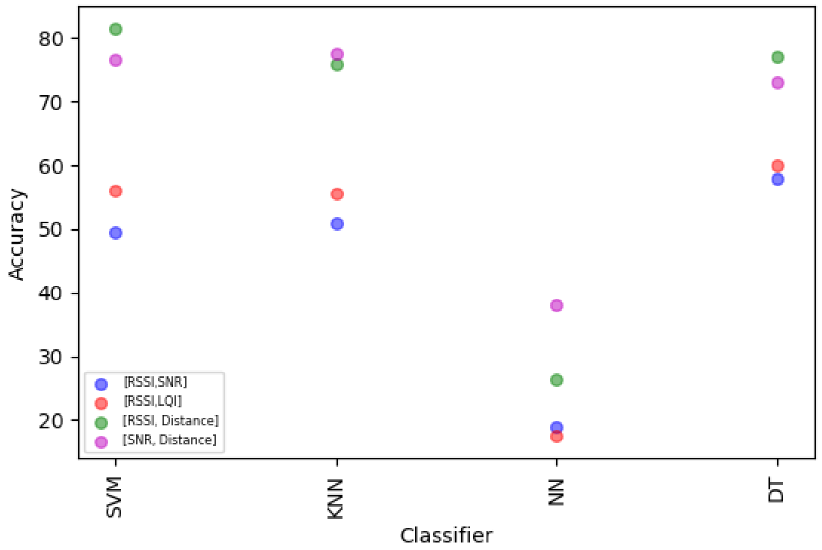


Fig. 54. Prediction accuracy of models based on a pair of features

The above result indicates a significant improvement of prediction accuracy with models using distance as an input feature. It also demonstrates an unexpected behaviour worth mentioning. Input parameter ‘distance’ measures the link distance between two nodes to the closest kilometer. Descriptive statistics confirm only small volatility (standard deviation = 0.48) in ‘distance’ compared to the other parameters used in experiments. Furthermore, ‘distance’ as a single input is unable to produce higher prediction accuracy (only 36 percent) due to the small variance in the radio link distance (STD = 0.48) of the operational network. However, pairing ‘distance’ with other attributes such as RSSI and SNR are able to improve prediction accuracy significantly. In essence, although individual attributes may not be able to produce higher information gain, higher prediction accuracy can be achieved by bundling multiple attributes.

All the above experiments are performed using a sample set collected from a single wireless segment containing 14 active nodes. However, with a mixed data set, prediction accuracy seems to be affected considerably, in a negative way. A detailed comparison of performance between uni-segment and multi-segment based prediction models are investigated in a separate section.

7.2.1.2 Sample Size

In this experiment, the relationship between the training sample set size and the prediction accuracy is investigated. In the following experiment, a sample set collected from a single wireless segment has been utilized.

Although, the primary objective of this experiment is to determine the relationship between the prediction accuracy and the training sample size; false positive and negative rates are also examined to support the findings. False-positive and false negative rates provide a strong quantification of a particular anomaly detection model and a comprehensive analysis of prediction models in this regards can be found in False Positive/Negative subsection in Physical Layer Characteristics section. All following experiments associated with Physical layer attributes utilize four input parameters (RSSI, LQI, SNR and Distance) unless otherwise indicated.

In the following, several training sets are utilized to determine the relationship between prediction accuracy and the training set size.

The following diagram is generated using eight different sample sets (50, 100, 200, 500, 1000, 5000, 10000 and 20000) and four machine learning methods (SVM, KNN, NN and DT).

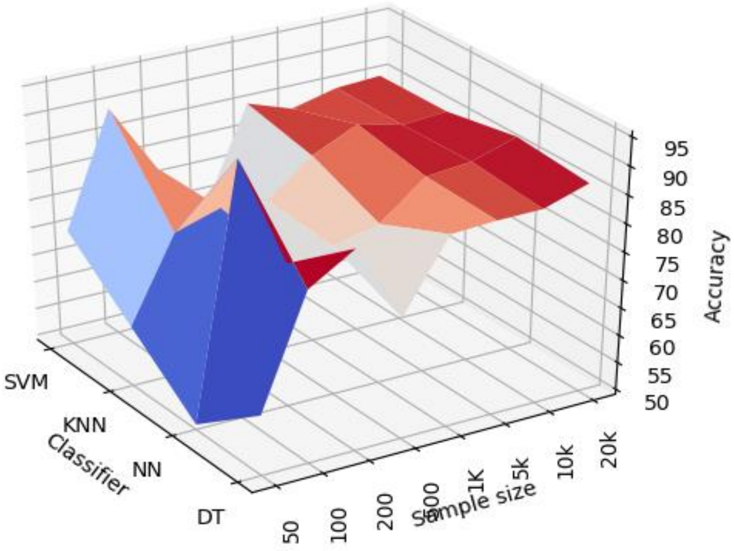


Fig. 55. Training set size vs classifier vs prediction accuracy

The above diagram demonstrates consistently higher prediction accuracy with models trained with larger sample sets. However, small training sets (less than 500 samples) are unable to produce satisfactory results for all four classifiers. It is worth mentioning that the number of samples required to train a model with a higher prediction accuracy is dependent on the nature of data and the strength of decision boundary between different classes.

The following diagram compares the behaviour of false positive/negative rates against training set size with four default classification methods.

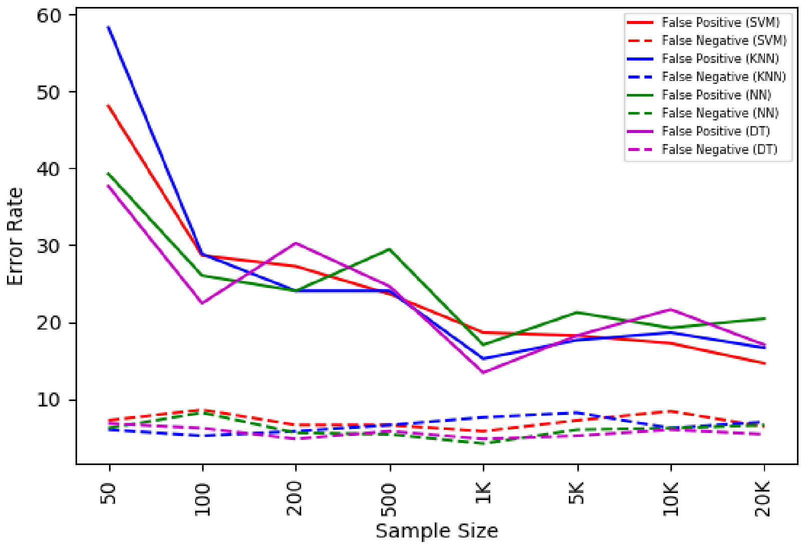


Fig. 56. Training set size vs false positive/negative rates

The above result confirms that regardless of the machine learning method being used, unreliable accuracy can be observed with models, trained with a smaller number of samples. However, models trained with over 500 samples are able to produce higher prediction accuracy and lower false positive and negative rates. Furthermore, a consistent positive relationship between the prediction accuracy and the sample size and a negative relationship between false positive/negative rates and sample size can be observed.

7.2.1.3 Number of Nodes

A typical low powered wireless network may comprise a few to a few hundred nodes. In the following experiments, the influence of number of nodes in the network, on prediction accuracy is investigated. The data set (aggregated) used in the following experiments is collected from several network segments. The first experiment is conducted using 1000 samples and four different machine learning algorithms.

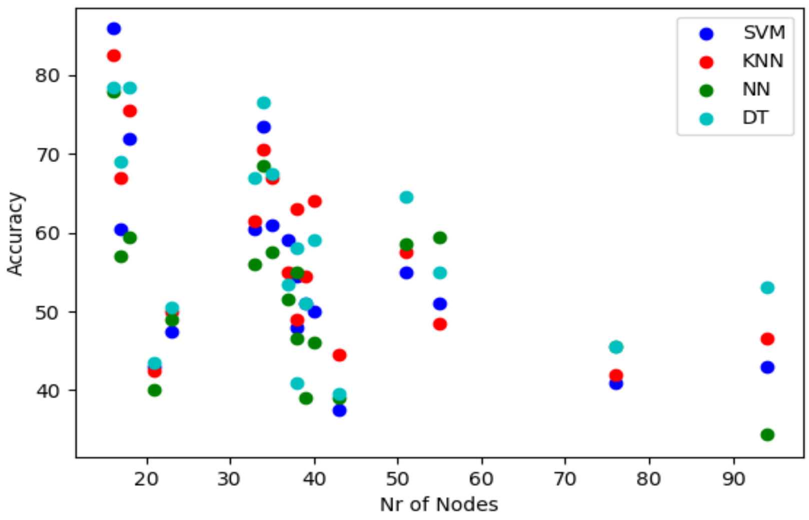


Fig. 57. Number of nodes vs prediction accuracy (small training set)

The above diagram demonstrates degradation in prediction accuracy when the number of nodes is increased. Similar behaviour can be observed with all four classification algorithms.

In the following experiment, the previous test is repeated using 5000 samples to train the prediction models; the result of the experiment is depicted in the following diagram.

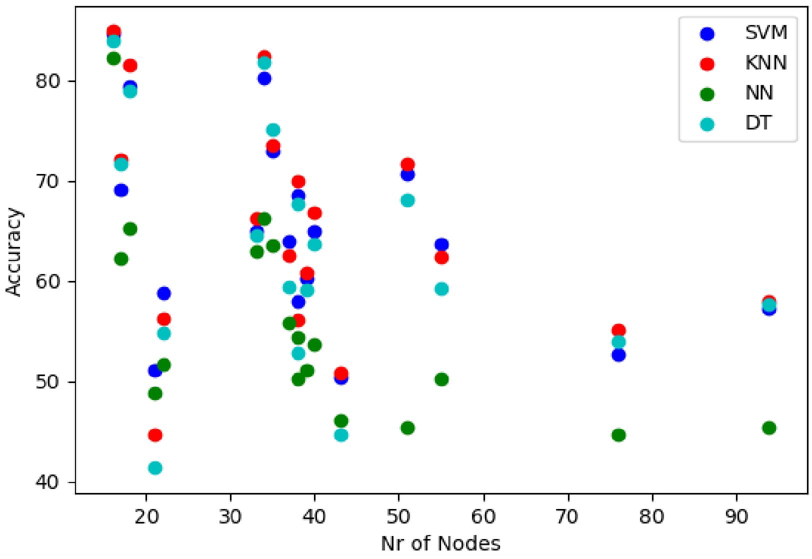


Fig. 58. Number of nodes vs prediction accuracy (larger training set)

Both experiments confirm a negative relationship between the prediction accuracy and the number of nodes in the wireless network. However, prediction models trained with larger sample sets are able to produce higher accuracy (around 5 percent improvement) for all four classification algorithms.

The following diagram summarizes the result obtained from the previous two experiments.

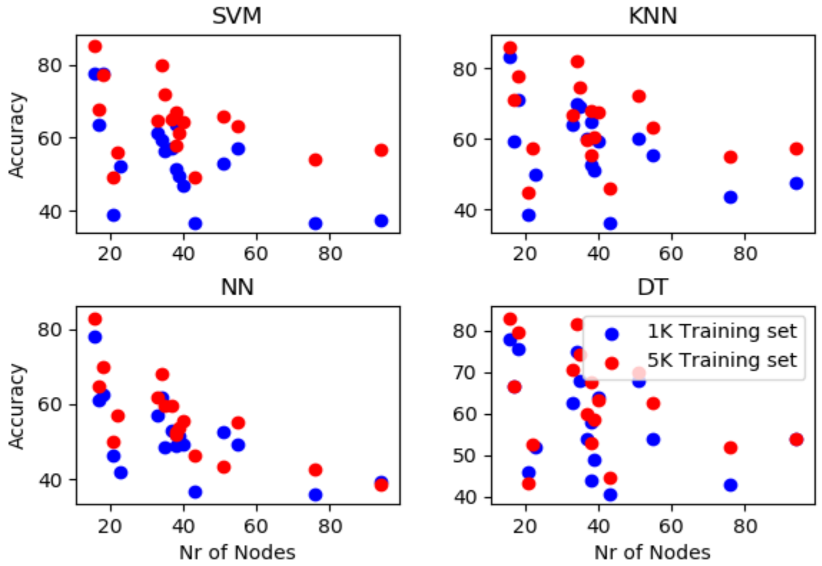


Fig. 59. Training set size vs number of nodes vs prediction accuracy

The above diagram confirms that regardless of the number of nodes in the network and machine learning algorithm, larger training sets are able to produce higher prediction accuracy.

The above diagram confirms that some classification algorithms are unable to produce higher accuracy with larger networks. For instance, a classification model based on NN is only able to produce 37 percent prediction accuracy with 94-node data set. However, as previously mentioned, since, the prediction model is not based on binary classification, 37 percent accuracy is still a significant improvement over the mean accuracy (1.06 percent) (Mean accuracy = 1/(nr of nodes) = 1/94).

The previous experiments are performed using wireless data collected from multiple wireless segments. However, different network segments can be influenced differently by different factors, including obstructions, interference and multi-path fading. As a result, wireless segment-specific effects may lead to a higher variance in aggregated data (lower prediction accuracy). In the following experiment, a relationship between the number of nodes and the prediction accuracy of a single wireless segment is investigated. In the following experiment, the dataset collected from a single segment is used and the number of nodes in the network is determined by removing packets, belonging to a particular MAC address. The following diagram demonstrates the relationship between the prediction accuracy and the number of nodes when a single wireless segment data is used to train the prediction model.

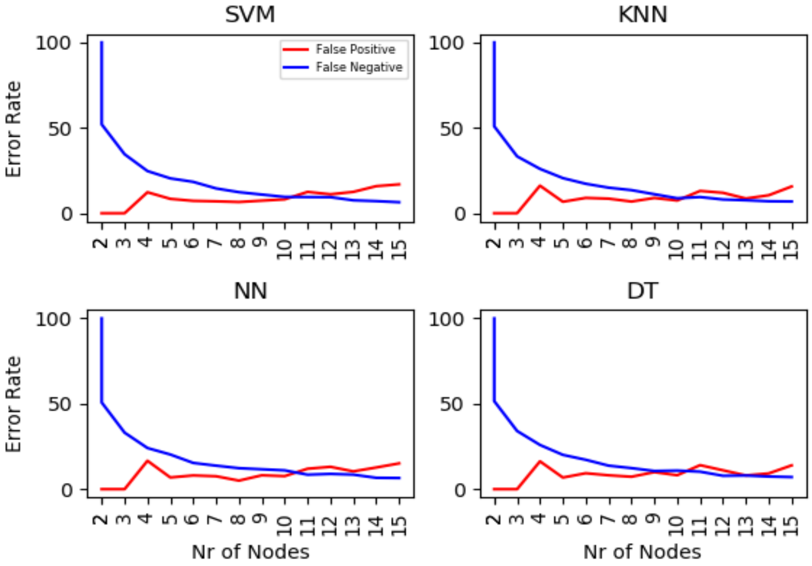


Fig. 60. Number of nodes vs false positive/negative rates (single seg)

The above results demonstrate that the number of nodes seems to have a slight positive relationship with false positives. Surprisingly, false-negative rates can be minimized by increasing the number of nodes in individual segments.

7.2.1.4 Labelling (MAC vs Binary)

Prediction models based on supervised learning rely on labelled data. Furthermore, the labelling process is also dependent on the objectives of the prediction model. For instance, if the objective of a particular prediction model is to determine anomalies, a binary classification model could be the right fit. However, if a prediction model is responsible for authenticating a source node, a more complex labelling mechanism, such as MAC addresses as labels, may be required.

In the following experiment, performance, including prediction accuracy and false positive/negative rates of binary and multi-label classification models are compared. With a multi-label model, (possible model for source authentication), the source MAC address is used as a label in the classification model. The following diagram compares the performance between binary and multi-label classification models.

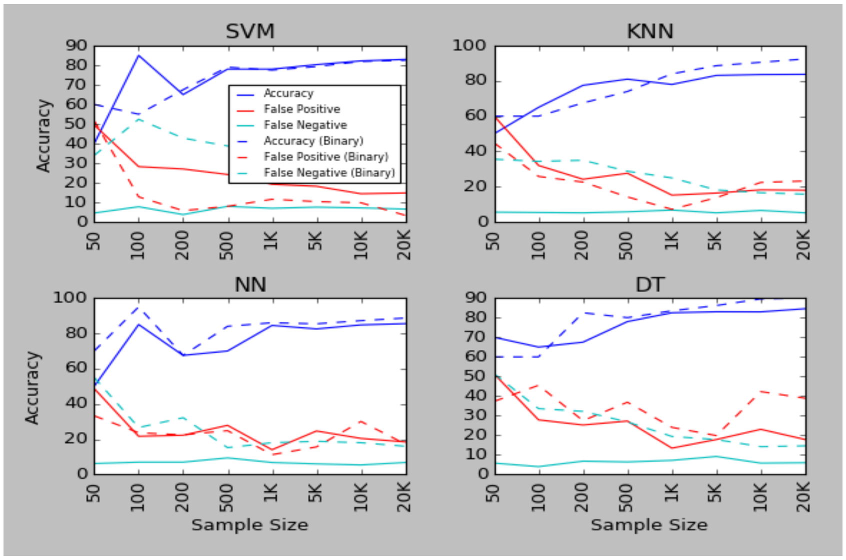


Fig. 61. Binary vs multi-label model comparison against sample size

The above result confirms a smaller gain in prediction accuracy with binary classification models. This trend continues with three (KNN, NN and DT) classification models trained with larger sample sets. However, mixed results can be observed in the false positive and negative rates. For instance, SVM has lower false-positive rates in binary classification models, while, DT seems to have higher false-positive rates in binary classification models with larger data sets. Furthermore, higher false-negative rates in binary classification models and consistent low false-negative rates with multi-label models can be observed for all four classification algorithms.

7.2.1.5 Noise Threshold

As previously mentioned, various factors including environmental conditions could significantly influence the behaviour of wireless networks. Although experimental data is collected from a live wireless network, during a prolonged period (six months), all possible factors may not influence the data set. For instance, heavy winds and storms in the late summer months and blowing snow during winter months can be experienced in North America. These seasonal effects significantly influence various properties of wireless networks, including the attributes (RSSI, LQI and SNR) tested in this work. In the following experiments, random noise is induced to experimental data to simulate severe environmental effects to investigate the reaction of prediction models.

In the first experiment, individual parameters are used in the prediction models. A random, controlled, synthetic noise, ranging from 0 to 3 STD (standard deviation), is injected to training and testing data to understand the response of the prediction model for input parameter variations. For the following experiment, 5000 and

10000 samples, collected from a operational wireless network with 16 nodes are used. It is worth mentioning, in the following experiment, synthetic noise is introduced to all data sets, before generating a positive and negative data set. The sequence permutation technique is used to generate the anomaly data set used in the following experiment. The following diagram is generated using a multi-label prediction model with a single-attribute-noise-injection.

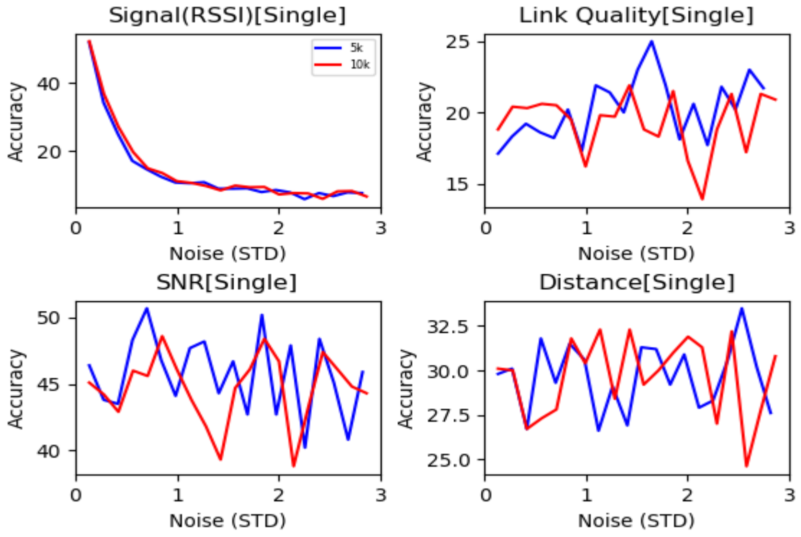


Fig. 62. Data variance vs prediction accuracy of a single-feature, multi-label model

The above result demonstrates inconsistent behaviour with all but RSSI when a single input parameter is used in the prediction model. Regardless of the training set size, model based on RSSI data has produced a non-linear negative relationship between data variance and the prediction accuracy. However, other attributes with higher variance were unable to produce consistent results.

In the following, the previous experiment is repeated, replacing a single input parameter with a multi-input-parameter prediction model. However, still, only a single input parameter is regulated using random-generated noise. The following diagram depicts the corresponding result.

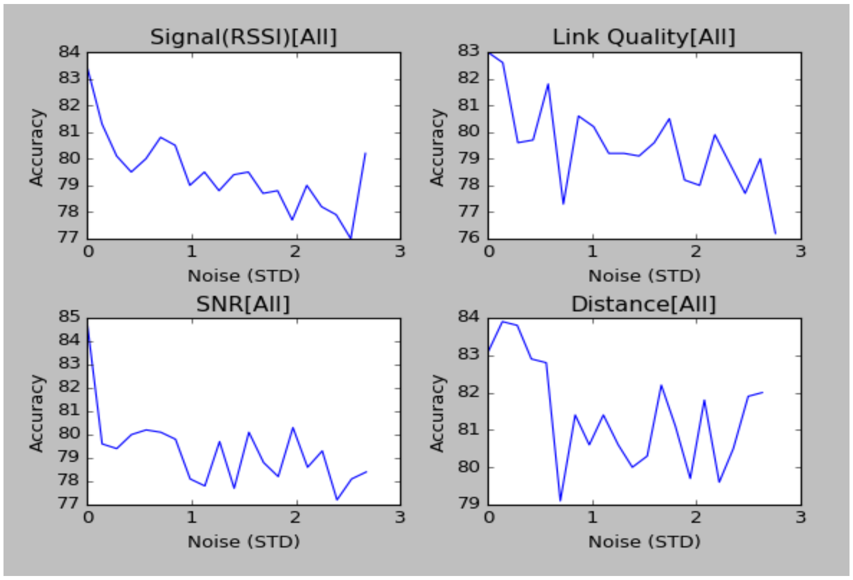


Fig. 63. Data variance vs prediction accuracy of a multi-feature, multi-label model

Above results demonstrate drastic improvement of prediction accuracy with multi-parameter prediction models. It is worth reminding that only a single parameter is manipulated in the above experiment. In essence, by carefully selecting the non-correlated input parameter set, highly generalized prediction models can be obtained. In data science, several techniques including principal component analysis (PCA) can be used to identify the most critical input parameters.

7.2.1.6 Model Aging Process (Retention Factor)

In the following experiment, the prediction model retention factor is investigated. 5000 samples collected from a single wireless segment at a relative time at t0, are used to train the prediction model. 13 sample sets, containing 2000 samples, collected from the corresponding wireless segment over multiple months, are used in the testing process. Each test set is collected in a fixed time interval over a few months to examine the long term effect on the prediction accuracy. The long term effect on other performance, including false positive and false negative rates are investigated in False Positive/Negative Rates subsection.

The following diagram depicts the retention factor in prediction accuracy.

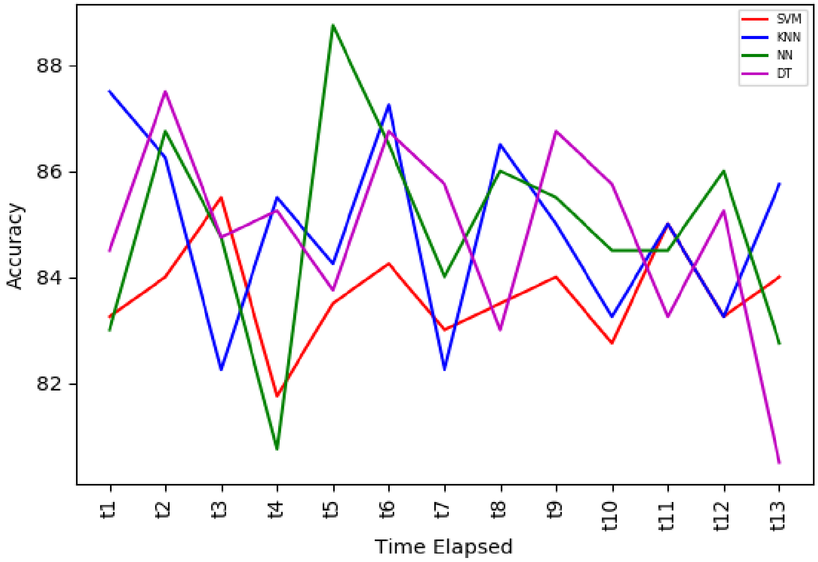


Fig. 64. Retention Factor vs classification algorithm

All four classification models demonstrate a satisfactory aging process in the prediction accuracy. While models based on NN tend to produce a small decline in the prediction accuracy against time, the other three algorithms are able to demonstrate consistent accuracy over a longer period of time. It is worth emphasizing that, data used in this work is collected during the summer months and seasonal influences may not be represented accurately by the data set.

7.2.1.7 False Positive/Negative Rates

In the following, several experiments are performed to understand the relationship between false positive/negative rates and various factors, including the number of training samples, the number of nodes, input parameters, classification algorithm, noise influence, and the labelling mechanism.

In the first experiment, the behaviour of false positive/negative rates against a number of samples in a training set is examined. The following experiment is completed using several sizes of training sets and a 1000-sample dataset is used to evaluate the prediction accuracy. Furthermore, training and testing data, used in the following experiment are collected from a single segment of a wireless network.

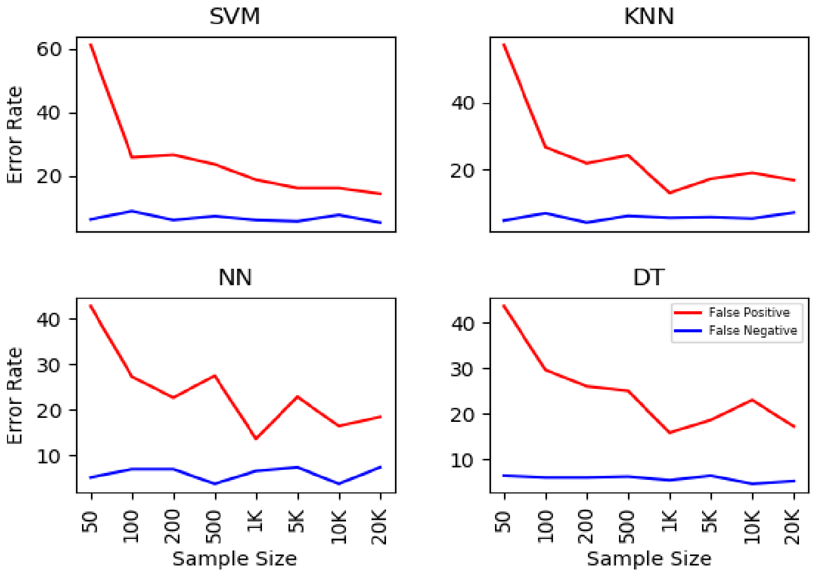


Fig. 65. Training set size vs prediction accuracy (multi-label model)

The above diagram demonstrates low false-negative rates, for all four classification models, with different sizes of training sets.

However, significantly higher false-positive rates can be observed for all classification models. Furthermore, the diagram confirms a significant decline of false-positive rates with larger training sets.

The objective of the following experiment is to determine the relationship between binary/multi-label classification models and the false positive/negative rates. Binary classification models can be used in network security to detect anomaly based on a particular packet or a flow. However, with models based on multi-label classification are able to detect anomalies as well as such models can be used with other security controls such as data origin authentication. In the following experiment, two prediction models are generated using the same data set, but with two different classification methods (binary and multi-label). With binary classification, two sample sets (positive (normal), negative (anomaly)) are merged and randomized to build a training data set. In a multi-label classification model, the training data is labelled using the source MAC address of a corresponding packet. In both models, a sequence permutation technique is used to produce a negative (anomaly) data set. The following diagram compares the false positive/negative rates of a binary and a multi-labelled classification model.

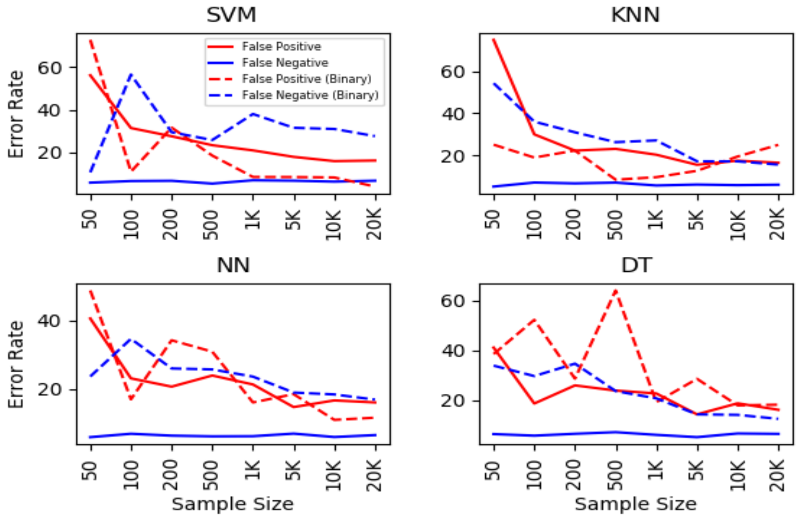


Fig. 66. Training set size vs false positives/negatives (multi-label)

The above diagram demonstrates fairly inconsistent results with all four classification models. However, all four multi-label classification models are able to produce consistent false-negative rates with different training sample sets. Furthermore, all but NN multi-label models depict a declining trend of false-positive rates with larger training sets. However, it is difficult to identify any obvious pattern between false positive/negative rates and the training sample set size for binary classifiers.

Number of Nodes

False-positive/negative rates may also depend on the number of labels used in the classification model. However, with multi-label classification models, the number of labels is directly related to the number of nodes in the corresponding network. As previously noted, data collected from five independent network segments are used in the experiments of this section. In the following, two separate experiments are conducted to determine the relationship between the number of nodes and false positive/negative rates. In the first experiment, false positive/negative rates for each wireless segment are examined. Each wireless segment contains a different number of nodes (16, 17, 18, 21 and 22). Multi-label prediction models are trained using 5000 samples and a separate data set with 5000 samples are used to calculate the false positive and negative rates. The following diagram demonstrates the corresponding result.

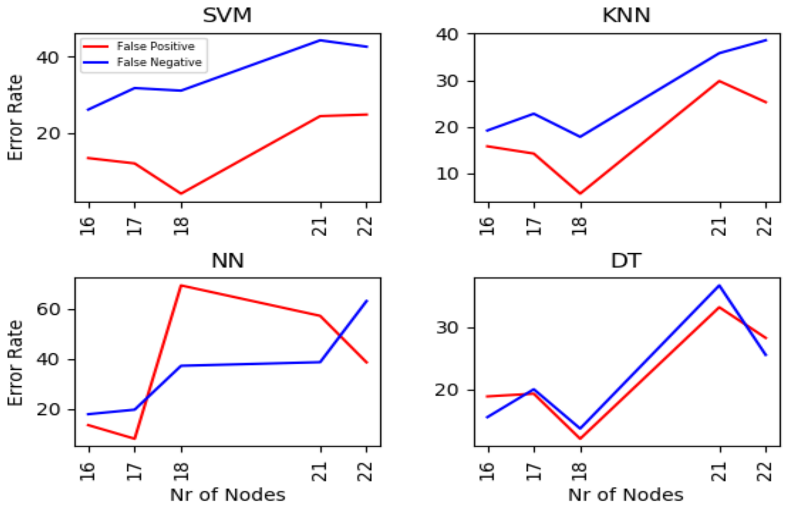


Fig. 67. Wireless segment vs false positives/negatives (multi-label)

The above diagram demonstrates similar behaviour with all four classification models. With each model, both false positive and negative rates increase with the number of nodes. However, as previously noted, each wireless segment could be affected by different environmental factors and this could lead to higher false-positive/negative rates and an inconsistent relationship between number of nodes and false positive/negative rates. Therefore, a second experiment is conducted with a single segment data set to further examine the relationship between number of nodes and false positive/negative rates. In the following experiment, several sample sets, extracted from a data set collected from a single segment is used. In this experiment, data belonging to selected nodes are removed from the dataset to reduce the number of nodes in the network. 5000 samples from each wireless-segment are used in the training process and a separate data set containing 5000 samples, collected from each wireless-segment, are used to evaluate the false positive/negative rates. The following diagram describes the findings of the above experiment.

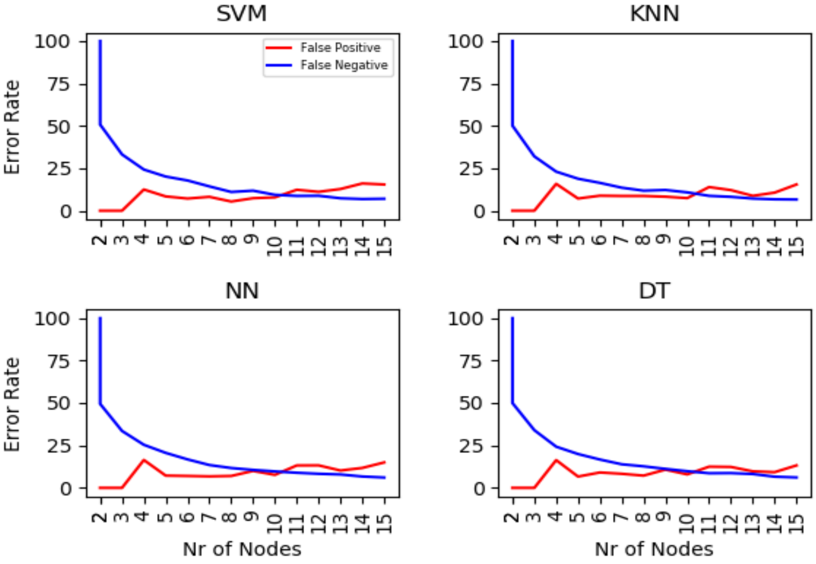


Fig. 68. Number of nodes vs false positives/negatives (single segment)

The diagram confirms a steady decline of false-negative rates with an incremental of number of nodes. However, consistent false positive rates with a marginal increment, responding to the number of nodes, can be observed.

Input Parameter Variance (Noise)

As previously mentioned, low powered wireless networks deployed in different environmental conditions are prone to a higher variation in measurements. Several factors including weather, frequency utilization and seasonal effects can contribute to this inconsistency. In the following experiment, the relationship between false positive/negative rates and the variance of input parameters are examined. 5000 samples collected from a single segment are used to train the prediction model. Four different input parameter settings are used in the following experiment and a random, controlled synthetic noise is injected into a single input parameter, while other parameters are left untouched. The following diagram summarizes the findings.

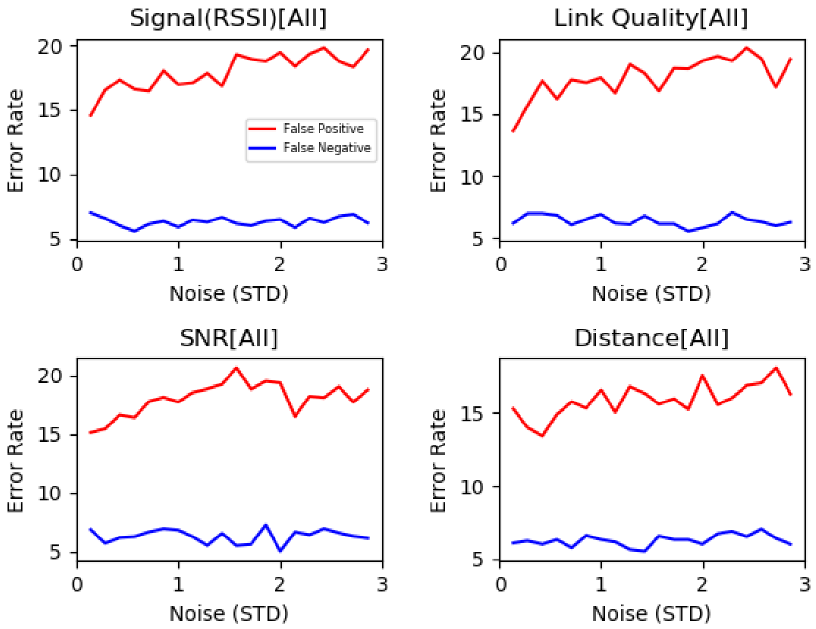


Fig. 69. Data variance vs false positive/negative rates

The above diagram demonstrates consistent false-negative rates for all four parameter sets. However, a small positive relationship can be observed between false-positive rates and the magnitude of synthetic noise. The above experiment confirms that the higher variation on a single attribute may not be able to influence the false positive/negative rates significantly.

Prediction Model Aging Process (Retention Factor)

In the following experiment, the behaviour of false positive/negative rates against unseen data is investigated. A prediction model is trained, using 2000 samples, collected from a single network segment at relative time t0. False-positive and negative rates are calculated using several test data sets (2000 samples each) collected from the corresponding network segment, over six months, in a fixed time interval. The observed results are depicted in the following diagram.

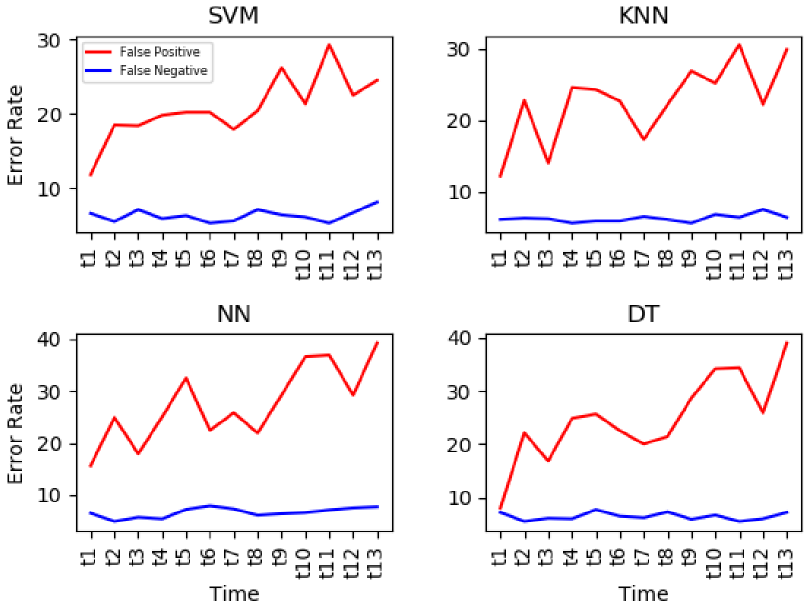


Fig. 70. Model Retention Factor (5k training set)

The above results demonstrate a remarkably consistent false-negative rate, in all four classification models, with unseen data. However, a positive relationship between false-positive rates and time elapsed can be observed in the above diagram. Sample size of the initial training set (5000 samples) may lead to such an inconsistency in false-positive rates. In the following, the above experiment is repeated with a larger training data set (10000 samples) to confirm such a theory.

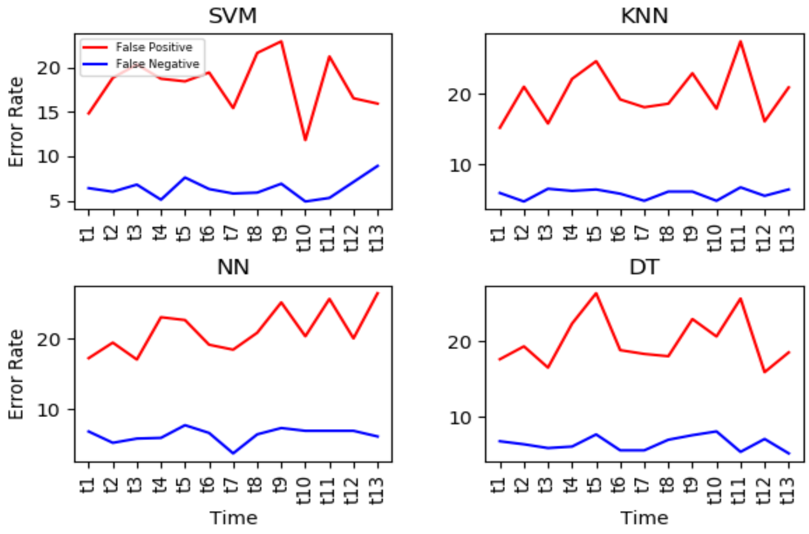


Fig. 71. Model Retention Factor (10k training set)

The above diagram confirms more stable false positive rates with the models trained using larger sample sets.

7.2.1.8 Time/CPU/Memory Utilization

The resource utilization data provides an overview of resource usage by different elements, including the machine learning algorithms, the sample size, the classification modes (binary/multi-label) and the complexity of network (number of nodes). More importantly, resource utilization data provides a measuring tool to determine a suitable prediction model for a particular environment. In the following, a few critical resources, including time, memory and CPU utilization are examined.

The following table describes the average time (in seconds) required for the training and the prediction process based on the number of nodes in the sample set. For the following test, the prediction model is trained with 5000 samples and a 5000-sample test-set is used.

Nr of Nodes	SVM	SVM (BIN)	KNN	KNN (BIN)	NN	NN (BIN)	DT	DT (BIN)
2	4.26	4.96	4.26	4.28	4.49	41.54	4.24	4.29
4	4.74	5.94	4.26	4.29	36.47	43.22	4.25	4.27
6	4.55	5.96	4.18	4.30	21.87	42.86	4.24	4.30
8	4.50	5.92	4.25	4.28	12.50	41.70	4.24	4.28
10	4.56	5.85	4.24	4.30	29.81	41.07	4.27	4.32
12	4.71	5.80	4.25	4.27	52.79	42.57	4.17	4.27
14	4.83	5.83	4.26	4.27	52.60	42.59	4.27	4.27

Tab. 16. Resource utilization (time) - number of nodes vs [binary, multi-label]

The above table demonstrates comparable, time utilization values for all, but the NN classification models. The prediction model based on neural networks consumes significantly higher processing time. Furthermore, the number of nodes doesn’t impact the time consumption of the training and the testing process significantly. Furthermore, in most cases, regardless of the network size, binary classification models utilize slightly more time.

The following table provides training and testing time consumption report for prediction models using different training set sizes and different machine learning algorithms.

Size	SVM MAC	SVM BIN	KNN MAC	KNN BIN	NN MAC	NN BIN	DT MAC	DT BIN
50	3.63	3.70	3.59	3.64	3.74	3.75	3.58	3.63
100	3.72	3.91	3.70	3.73	3.89	3.89	3.67	3.72

200	3.76	3.92	3.69	3.79	5.47	4.20	3.77	3.79
500	4.06	4.19	3.83	3.89	7.64	6.10	3.81	3.91
1k	4.52	4.82	4.20	4.32	8.92	7.47	4.21	4.26
5k	8.48	10.17	7.06	7.15	46.32	35.90	7.00	7.09
10k	14.31	19.48	10.97	11.02	91.54	31.18	10.66	10.72
20k	29.15	45.63	20.48	20.59	183.7	144.5	20.46	20.47

Tab. 17. Resource utilization (time) – sample size vs [binary, multi]

The above table demonstrates comparable time usage for prediction models based on SVM, KNN and DT. However, similar to the previous report, prediction models based on NN demonstrate significantly higher time usage with all sizes of training sets.

The following table describes the CPU usage of different prediction models. Calculations are based on averaging three CPU usage measurement values. (The CPU utilization has been computed using the "top" command in Ubuntu system. However, in multi-core systems, Ubuntu calculate the CPU usage by adding the usage of individual CPU by a particular application)

Size	SVM	SVM BIN	KNN	KNN (BIN)	NN	NN (BIN)	DT	DT (BIN)
50	108.93	108.88	109.08	108.83	380.31	315.35	108.89	108.78
100	108.96	108.32	108.88	108.89	402.98	347.15	108.87	108.94
200	108.97	108.81	108.87	109.14	404.56	372.50	108.85	108.91
500	108.88	108.91	108.88	108.64	409.24	418.54	108.95	108.91
1k	109.56	108.87	108.79	108.99	420.28	419.77	108.89	108.88
5k	108.82	108.82	108.92	108.92	419.04	418.03	108.80	108.81
10k	108.96	108.85	109.02	108.89	416.66	397.66	108.84	108.97
20k	108.78	108.87	108.86	108.76	413.76	416.28	108.96	109.17

Tab. 18. CPU utilization – number of nodes vs [binary, multi]

Similar to the time usage report, the above table demonstrates comparable CPU usage for prediction models based on all tested machine learning algorithms, but neural networks (NN). Both binary and multi-label classification models based on NN consume significantly higher CPU resources.

The following table describes the memory usage of the prediction models trained with different training set sizes and different classification algorithms.

Size	SVM	SVM BIN	KNN	KNN BIN	NN	NN BIN	DT	DT BIN
50	128.3	129.0	128.7	129.0	133.5	134.2	129.1	129.2
100	129.0	129.3	129.6	129.0	133.3	133.6	128.6	129.6
200	129.2	129.1	128.9	129.3	134.3	133.5	129.3	129.7
500	129.2	130.0	128.9	129.4	133.7	134.8	129.0	129.1
1k	129.5	131.4	129.2	129.6	134.1	135.7	129.6	129.8
5k	130.3	132.7	129.9	131.5	136.4	137.0	130.0	132.2
10k	134.6	136.1	131.9	135.1	139.7	140.5	131.7	135.5
20k	134.9	142.5	134.0	140.0	143.2	156.3	134.0	140.0

Tab. 19. Memory utilization – number of nodes vs [binary, multi]

Unsurprisingly, all four classification algorithms have similar memory consumption tendency. Furthermore, with all four algorithms, the binary classification models seem to have higher memory consumption with larger training sets. However, models based on KNN and DT were able to utilize minimum resources to train anomaly detection models based on Physical layer attributes.

7.2.1.9 Result Summary

In this portion of experiments, the IEEE 802.15.4 physical layer attributes have been investigated to build a prediction model to detect anomalies in low powered wireless networks. Several of the

IEEE 802.15.4 physical layer properties, including signal strength (RSSI), noise ratio (SNR), link quality (LQI) and link distance are evaluated. Furthermore, different factors, such as the training set size, the number of nodes, classification algorithms, noise effect, and the prediction model aging process are examined. In the following, important findings in the above experiments are listed.

- Characteristics of Physical layer attributes:
    - Experimental data demonstrated a smaller variance with the signal strength (RSSI) and the noise (SNR) data. However, link quality (LQI) data produce a significantly higher variance.
    - A negligible correlation (Pearson r: less than 0.01) could be found between signal strength (RSSI) and the link quality (LQI). RSSI and LQI produce two strong independent input features for classification models.
    - Variations of the individual attributes, among nodes were not directly correlated. Different factors including the location, physical stability, signal path and the local interference can contribute to this instability.
    - Input features (RSSI, SNR, LQI and LD) were unable to differentiate individual nodes into unique classes. In essence, clustering algorithms with the use of individual or combined input parameters were unable to accurately classify data into multiple groups based on data origin.
    - By segmenting larger networks into smaller clusters based on common properties including location, environmental influences and link attributes

(distance, link path), higher prediction accuracy could be achieved

- Classification models based on single segment:
    - With the use of multiple attributes in the training process, the prediction accuracy could be improved notably (85 percent prediction accuracy could be achieved with the use of all four attributes in input feature vector).
    - All four classification models (SVM, KNN, NN and DT) were able to produce a higher prediction accuracy (over 85 percent), with the prediction models trained with 20000 samples, collected from a single network segment.
    - Surprisingly, the false-negative rates of a single-segment prediction model declined with a higher number of nodes.
    - Consistent false positive rates could be observed in models trained with single segment data.
    - In individual wireless segments, with increments in the number of nodes, a smooth decline of false-negative rates could be observed. However, under similar conditions, false-positive rates tended to increase steadily.
    - The false-positive/negative rates of a prediction model based on single-segment-data were significantly lower than a prediction model trained with multi-segment data.

- Classification models based on multi-label:

    - Classification models based on the multi-label technique were unable to produce higher accuracy (max 60 percent with RSSI), with a single parameter as an input. However, the above accuracy is a significant improvement over the mean accuracy rates.
    - The prediction accuracy can be significantly, inconsistent in models trained with smaller data sets (less than 500 samples).
    - Regardless of the training set size, classification models based on the multi-label technique were able to maintain consistent, low false negative rates. However, higher false positive rates could be observed with smaller training sets. False positive rates could be significantly reduced in models trained with larger data sets.
    - In a multi-label classification model, the prediction accuracy has declined incrementally, with the network size (number of nodes). Similar behaviour could be observed with all four evaluated classification models.

- Classification models - General
    - Compared to multi-label classification models, binary classification models were able to gain a smaller improvement in prediction accuracy with most classification algorithms.
    - The prediction accuracy of fluctuating, input parameters could be improved by combining multiple input parameters together.

- All four classification algorithms performed well in a model aging process evaluation test. Each prediction model was able to maintain the prediction accuracy, within a three percent margin over a longer period of time. However, experimental data was collected during the summer months and seasonal influences may not be accurately represented by the data set.
- When the prediction model is based on multi-segment data, the false positive/negative rates did not correlate, directly, to the number of nodes in a particular network segment. For instance, an experiment confirmed that with prediction models based on all four algorithms, the wireless segment with 18 nodes, were able to produce smaller false-positive rates, compared to wireless segments with 17 and 21 nodes. Other factors, including environmental and higher data variance, may contribute to the above inconsistency.
- When multiple input parameters were used, a higher variance of a single parameter didn’t impact the false positive and negative rates significantly.
- The prediction model aging process evaluation test confirmed that with time, false-positive rates increase steadily. However, consistent false-negative rates could be observed for a longer period of time.
- As far as resource utilization is concerned, all classification algorithms, but neural networks (NN) consumed similar processing time in training the prediction models. However, classification models

based on NN required an average of six times, the processing time, required by other classifiers.

- The number of nodes in the wireless network, didn't impact the processing time significantly. However, in most cases, binary classification models required more processing time when compared to multi-label models.
- During the training process, all four classification algorithms had similar memory allocation. However, in larger data sets, binary classification models required slightly higher memory usage.

7.3 LR-WPAN Characteristics

7.3.1 Regression Methods

In the following experiment, the battery level of a low powered wireless node is evaluated as time-series data using a few different regression models. The main objective of this experiment is to determine, how different regression models are able to approximate the battery level using the timestamp as an input parameter. The battery level of a low powered wireless device can deplete linearly with regard to the time, during a single life cycle of the battery life. The following diagram is generated using a portion of a battery life cycle data and this demonstrates a negative relationship between the power level of a battery and the time.

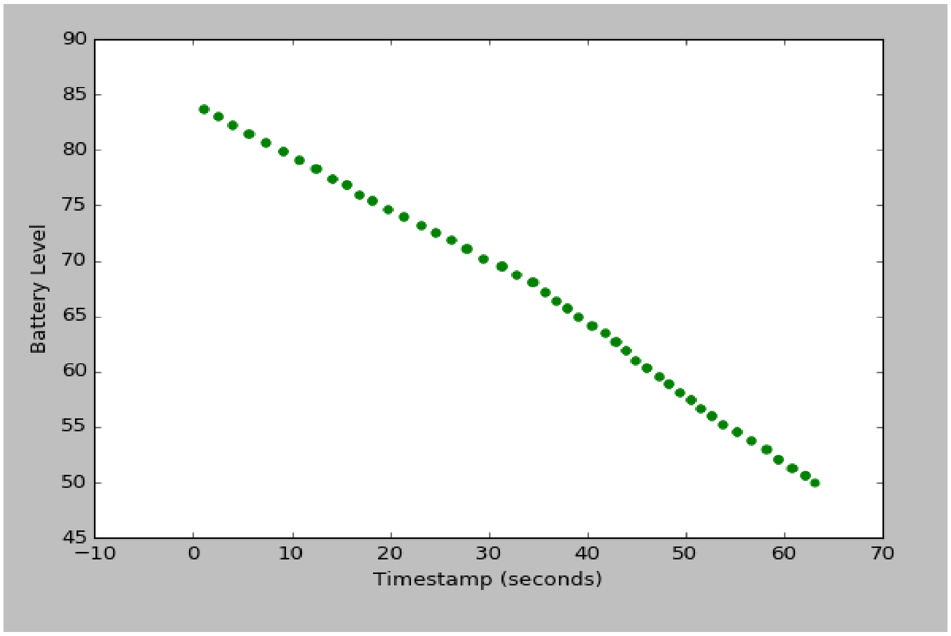


Fig. 72. Regression of a single battery life cycle

The following diagram is generated using a linear regression model and it confirms a strong negative linear relationship with time (correlation coefficient (Pearson R) = -0.99). The prediction model has delivered over 99 percent accuracy (correlation coefficient) with sample data, collected from a single life cycle of the battery.

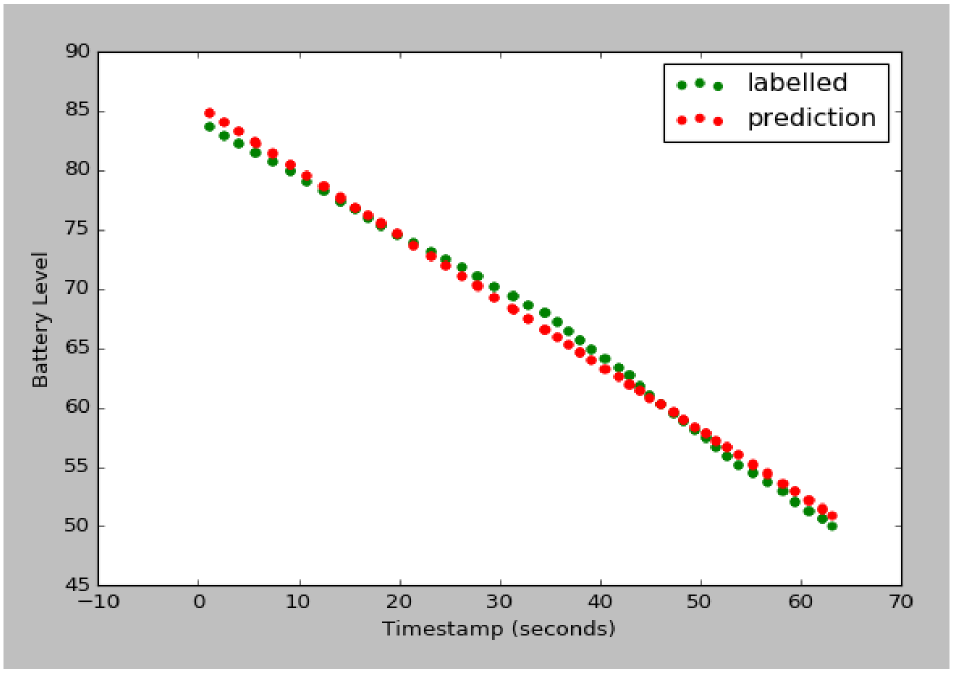


Fig. 73. Regression analysis: expected vs predicted (single cycle)

However, the life cycle of a battery repeats at both consistent and inconsistent time intervals. The following diagram confirms a non-linear relationship with a sample set, containing multiple battery life cycles. The red line indicates the approximation line, determined by a prediction model based on linear regression. However, the corresponding model was only able to produce dismal results (accuracy – below five percent).

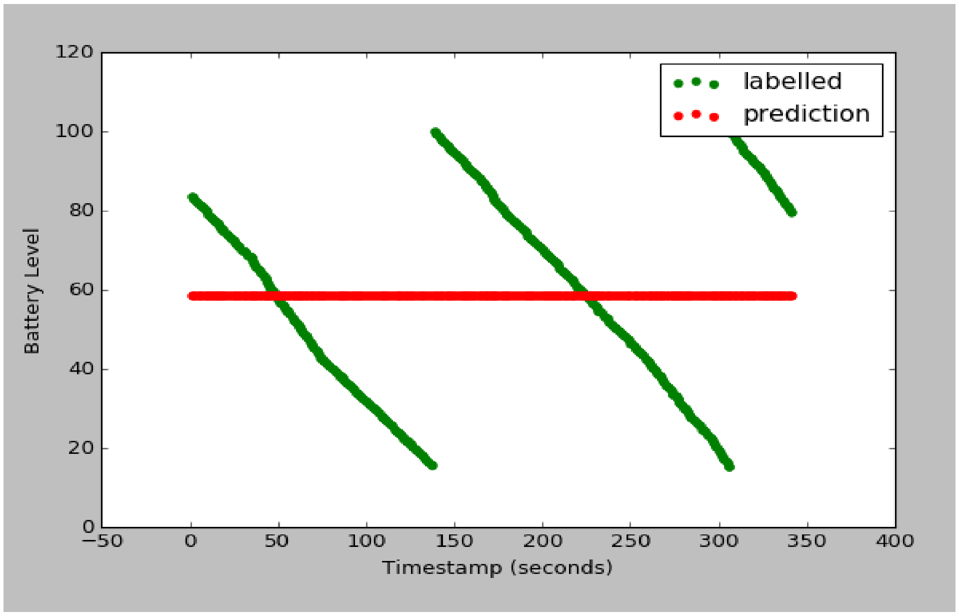


Fig. 74. Regression analysis: expected vs predicted (multi-cycle)

The above result confirms that linear regression is not a preferable method to build a prediction model using the battery power level and the timestamp. However, a well-known technique such as a sliding window for time-series data can be utilized to improve the prediction accuracy.

A number of different regression models are evaluated in the following section to determine whether prediction accuracy can be improved with those regression models. For the aforementioned experiment, the following regression methods are used.

- Linear Regression
- Support Vector Regression with RBF Kernel
- Random Forrest Regression
- Decision Tree Regression

The following result is generated, using a 5000-sample set and the result confirms a lower accuracy, with a model based on linear regression. However, a drastically improved result can be achieved with non-linear regression models such as support vector regression (SV Reg.), random forest regression (RF Reg.) and decision tree regression (DT Reg.).

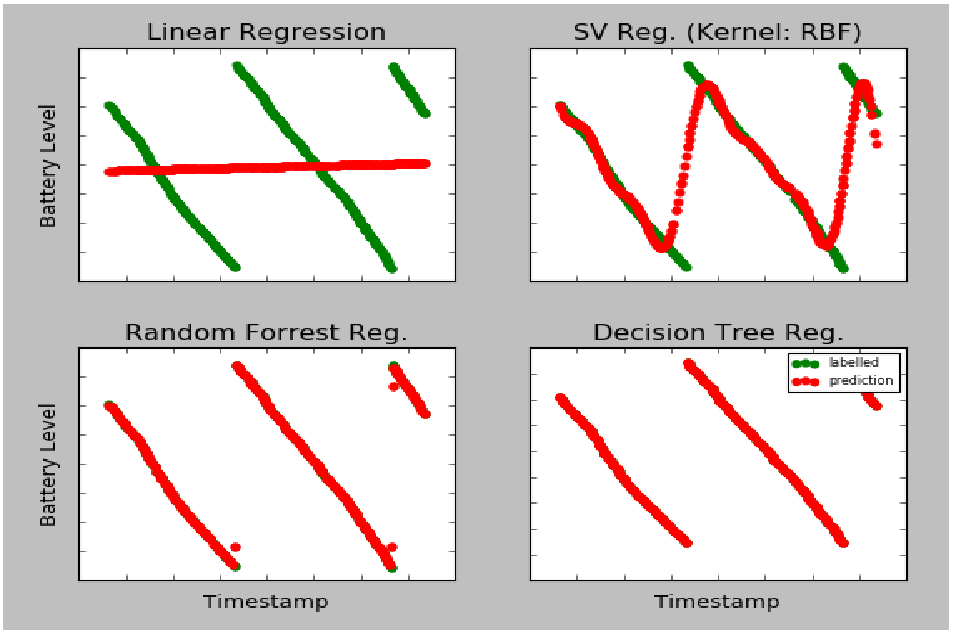


Fig. 75. Regression analysis: linear vs non-linear models (small set)

The above diagram demonstrates highly accurate regression approximations with Random Forest & Decision Tree based regression models. However, such models could lead to over-fitting and models might not be able to predict unseen data with higher accuracy and different machine learning techniques can be used to address the over-fitting issue.

Furthermore, the increment of the sample set doesn’t alter the prediction accuracy significantly. The following diagram is generated using a 10000 sample-set and the four regression models used in the previous test.

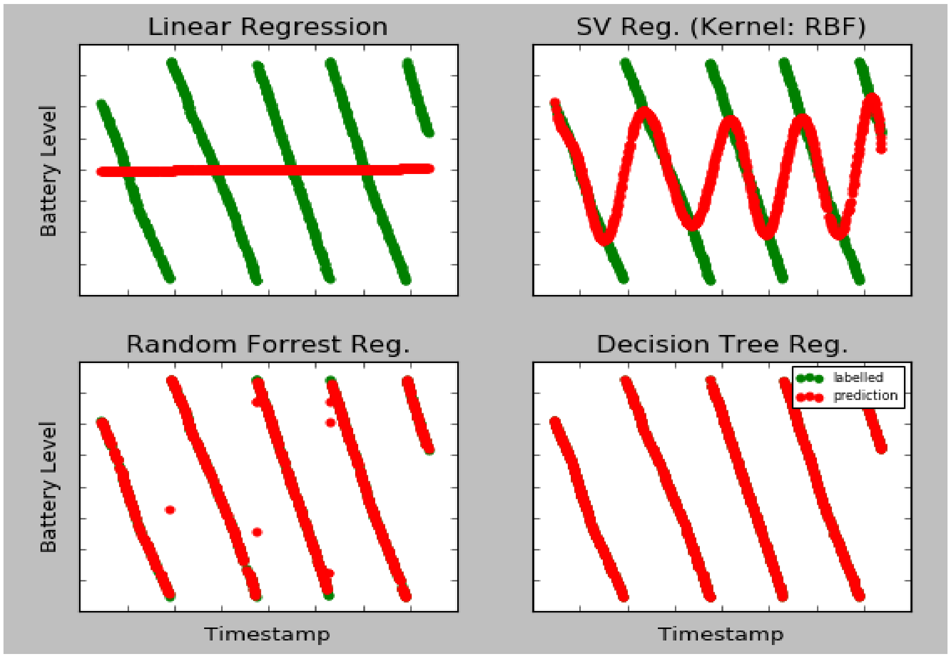


Fig. 76. Regression analysis: linear vs non-linear models (large set)

The above results were based on the assumption that policies and procedures are implemented to replace the batteries when a low battery threshold value is reached. However, in a realistic scenario, a low powered node's battery could be changed at the operator's convenience or in a predetermined schedule.

The following experiment has been done to determine the behaviour of regression accuracy when the battery replacement process is unpredictable. A separate sample set is collected from the simulated environment, operating on IEEE 802.15.4e/TSCH mode for the following portion of the experiment. All the settings including topology and the application-specific parameters are similar to that of previous tests. However, a randomly selected battery level is chosen for the battery replacement process. The following diagram depicts the fundamental differences between two data sets.

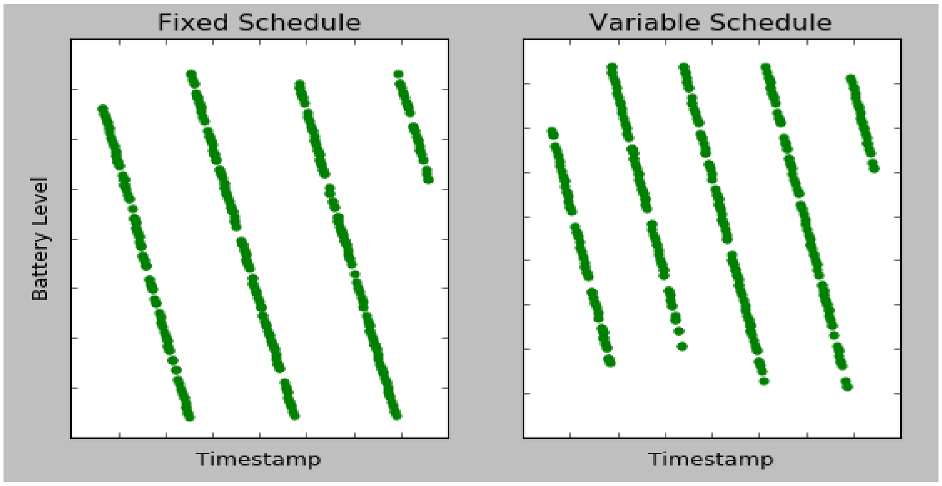


Fig. 77. Comparison between fixed and variable battery replacement process

The above diagram confirms that, within a fixed schedule, the lower value of the battery level is consistent, throughout the dataset. However, with a variable schedule, the lower value of the battery level fluctuates between 0 – 40 percent.

The following diagram depicts the comparison between datasets with fixed low battery level and variable battery level.

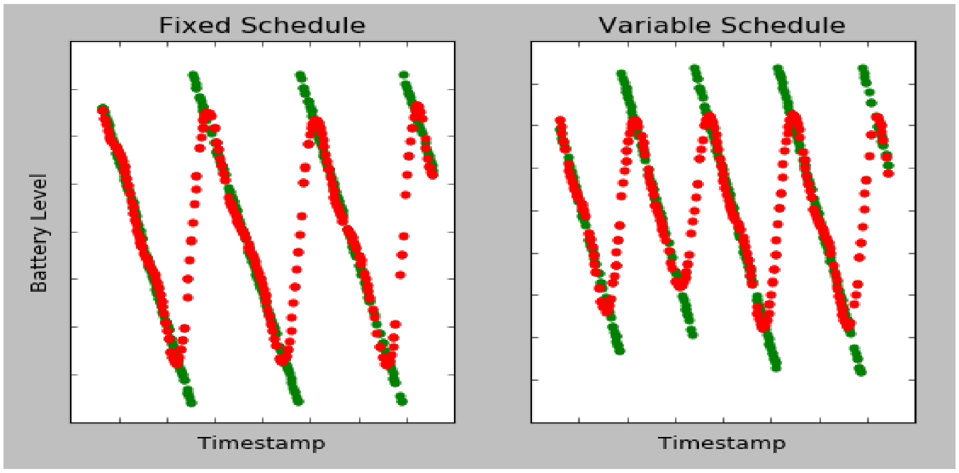


Fig. 78. Regression analysis: fixed and variable battery replacement

The following diagram demonstrates a comparison between the regression accuracy of sample sets, collected from two different populations (fixed low battery threshold and variable threshold). Four different regression methods are tested and the following diagram confirms a smaller gain of accuracy in some regression models with a fixed-low battery level data set.

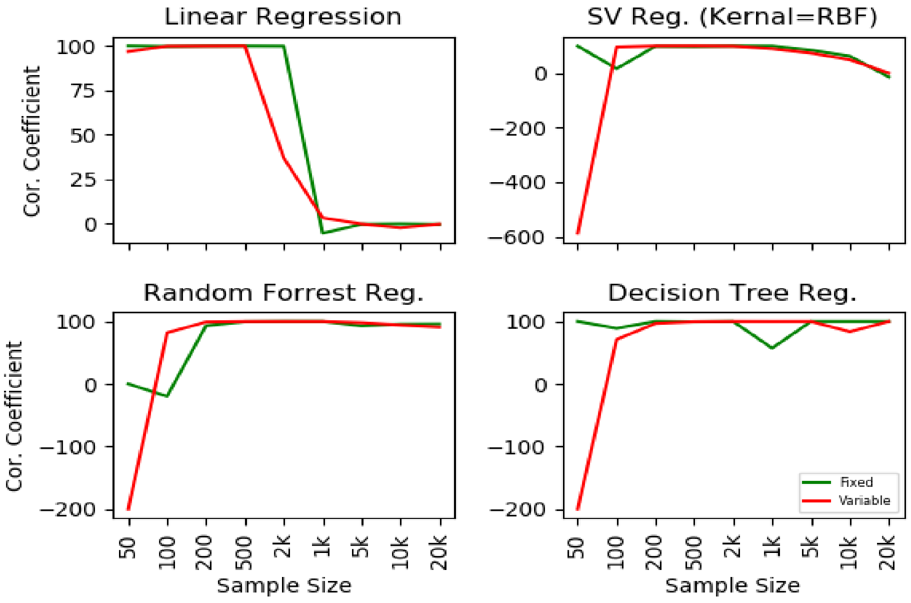


Fig. 79. Correlation coefficient comparison: data set vs [fixed, variable]

Experiments related to the regression analysis of battery usage confirm a potential relationship between battery usage and the operational life cycle of the low powered device. However, regression analysis might not be able to produce a model to determine anomalies of low powered wireless networks. In the following section, several prediction models based on classification algorithms are investigated.

7.3.2 Classification Methods

7.3.2.1 Sample Size

In the first experiment, the relationship between prediction accuracy and the number of samples used in the training process are tested. For this experiment, several datasets ranging from 100 to 50000 samples are used. Prediction accuracy is calculated using a 5000 sample set, extracted from the same population. The following diagram demonstrates the corresponding results.

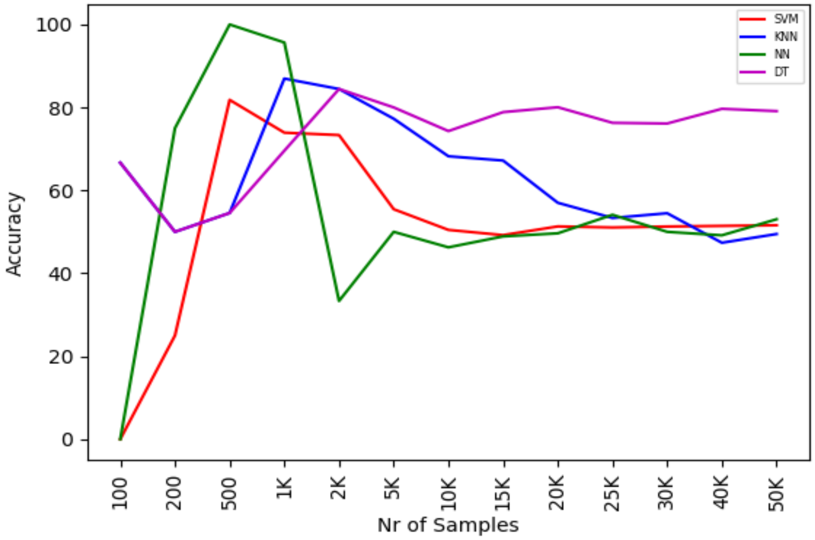


Fig. 80. Training set size vs prediction accuracy

The diagram confirms that only models based on a decision tree (DT) are able to produce usable results. The other three algorithms provide higher accuracy with lower sample rates (less than 1000). This simulation was configured to generate around 1400 samples for a single battery life cycle. This could have contributed to the observed behaviour. Furthermore, the above experiment was conducted using a sample set with fixed low battery level threshold. The following graph was generated using similar parameters, with the use of a dataset at a variable low battery level.

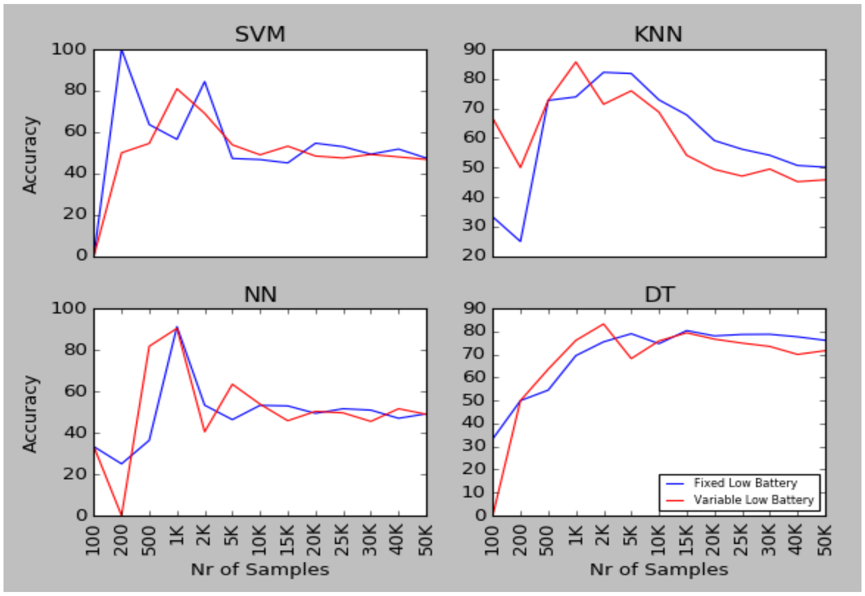


Fig. 81. Prediction accuracy: training set size vs [fixed, variable]

The above diagram confirms no substantial difference in accuracy between a dataset with a fixed battery level and a variable battery level. However, among tested classifiers, only the model based on DT is able to produce higher prediction accuracy.

The following experiment was conducted to investigate whether prediction accuracy can be improved, using ensemble methods, when, the battery level of a low powered, wireless device is used as an input parameter. The below described ensemble machine learning methods are evaluated in the following experiment.

- Bagged Decision Tree (BDT) (Bagging)
- Random Forrest (RF)(Bagging)
- AdaBoost (Boosting)
- Stochastic Gradient Boosting (SGB)(Boosting)

The following diagram is generated using the above-listed ensemble methods for both fixed and variable, low battery level replacement values. The diagram confirms ensemble machine learning methods, based on bagging techniques are able to produce acceptable results. However, models, based on boosting techniques (AdaBoost, stochastic gradient boosting) are not able to improve the prediction accuracy.

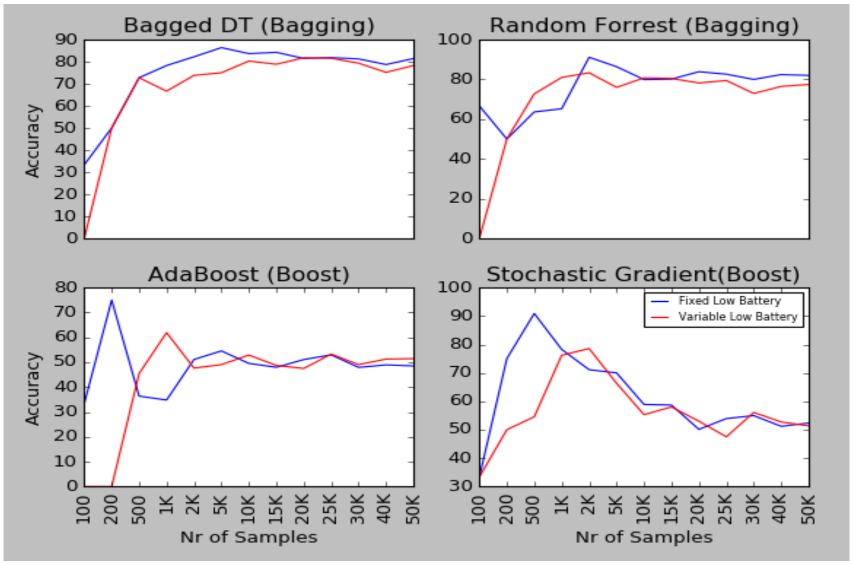


Fig. 82. Prediction accuracy (ensemble): training set size vs [fixed, variable]

The above tests confirm that with three algorithms, the prediction accuracy asymptotes to a certain value. Since the life cycle of a battery contains around 1500 samples, such a result could be expected. However, the behaviour of the Stochastic Gradient may be a result of handling a large amount of repeatable data with boosting technique used in the algorithm.

In the following section, additional tests are conducted to confirm whether models trained with smaller datasets have consistent, higher accuracy. The following results are based on models trained with three different sample sets (500, 1000 and 2000) and several test sets (ranging from 100 to 50000 samples), with different sample sizes, extracted from the same population are used to evaluate the prediction accuracy.

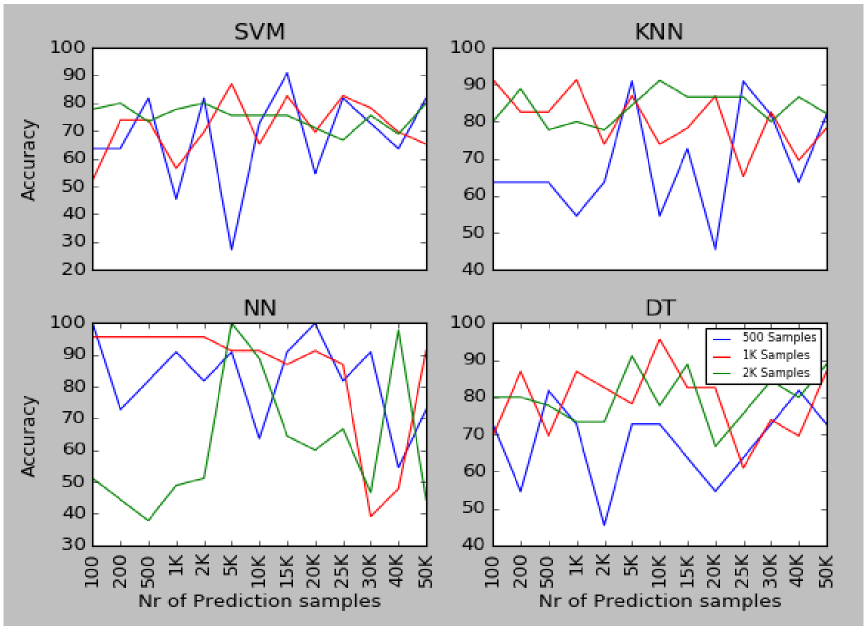


Fig. 83. Training set size vs prediction set size vs prediction accuracy

Compared to previous tests, where the prediction model is trained with a larger sample test, prediction models based on smaller sample sets seem to provide higher accuracy. (The average battery life cycle of devices used in this experiment is around 1400 packets).

The following diagram is generated using algorithms, based on ensemble techniques. The diagram confirms, with the use of a training sample set, size of the battery life cycle would improve prediction accuracy. However, the following diagram further confirms that ensemble techniques are not able to provide drastic improvements in prediction accuracy, compared to prediction accuracy based on default classifiers.

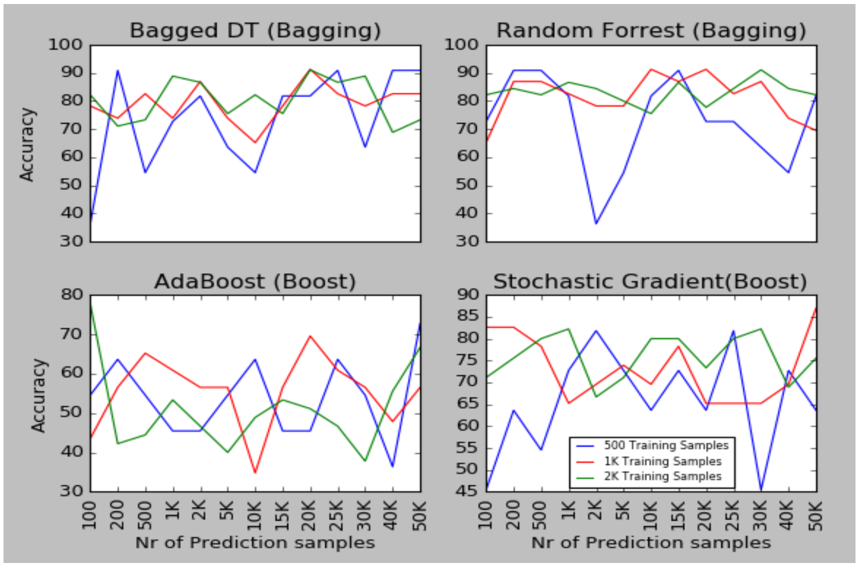


Fig. 84. Training set size vs prediction set size vs prediction accuracy (ensemble)

The previous experiment results are based on models trained with data permutation (rearranging sequence) as a labelling mechanism. To complete this portion of experiments, previous experiments are repeated using a different labelling mechanism, where, random noise is induced to input data to generate an anomaly data set. The following diagram compares the prediction accuracy when two different labelling techniques are used to train prediction models. The anomaly dataset is generated by injecting up to a five percent random noise to the battery level.

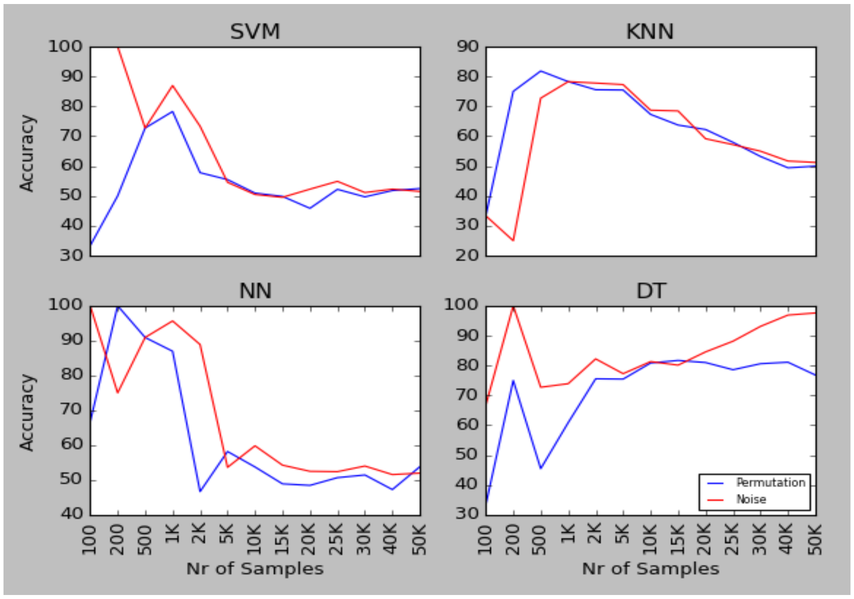


Fig. 85. Prediction accuracy vs [training set size, labelling technique]

The above diagram is generated using several classification methods, trained with different training-set size, and the result confirms no significant difference in prediction accuracy between two labelling mechanisms with all classifiers except SVM when the battery level (not battery usage) is used as an input parameter.

7.3.2.2 Delta Values

In the previous set of experiments, the battery level of a node is used as an input parameter to build prediction models to identify anomalies in low powered wireless networks. In this portion of experiments, the battery level and the timestamp (frame.time_relative), which are considered as continuous data, are transformed to normal distribution models, by computing delta values of battery level (battery usage) and time elapsed (time usage). The following formula is used to compute battery usage and corresponding time values.

$\text{battery\_usage}_t = (\text{battery\_level}_{t-1} - \text{battery\_level}_t)$ (for t Є (time_elapsed))

$\text{time\_usage}_t = \text{time}_t - \text{time}_{t-1}$ (for t Є (time_elapsed))

The following diagram depicts the frequency distribution of battery usage and time usage (Delta) values.

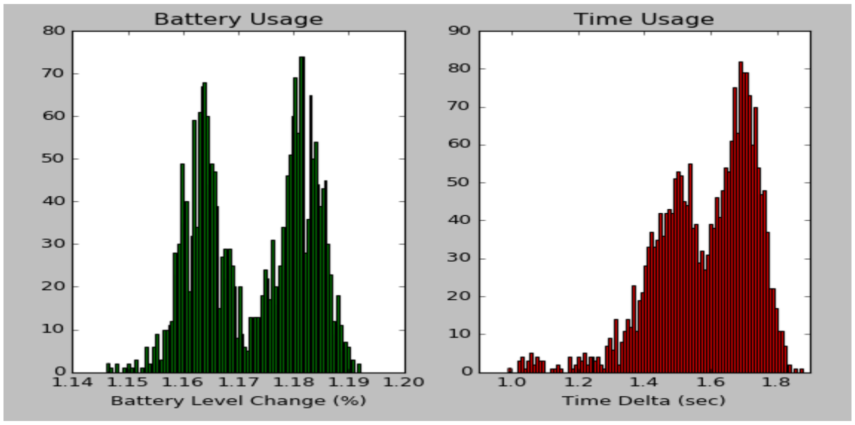


Fig. 86. Frequency distribution battery usage and time usage

The frequency diagram demonstrates an interesting behaviour of battery usage. This could be a result of two primary packet types (enhanced beacons and unicast packets with battery value) are exchanged between the end node and the PAN coordinator. In the following section, a few previously completed experiments are repeated using delta values.

In the following experiment, the relationship between prediction accuracy and the training set size is investigated using battery usage as an input parameter.

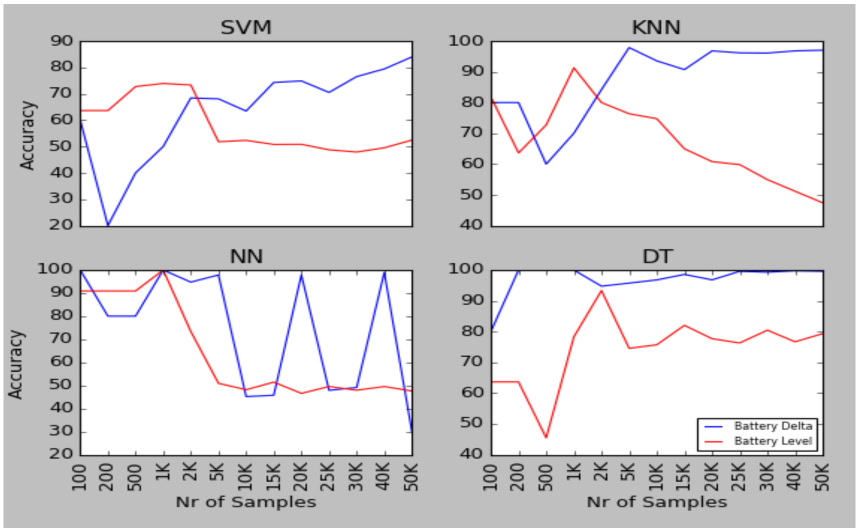


Fig. 87. Prediction accuracy: battery usage vs battery level

The above diagram confirms significantly higher prediction accuracy than the models using the battery level as an input parameter. All classifiers, but NN are able to produce consistent prediction accuracy in models trained with larger training sets and with battery usage as an input variable.

In the following experiment, SVM and NN are replaced by two ensemble methods. The corresponding diagram confirms, models with battery usage as an input parameter consistently performed better than models based on battery level as an input parameter.

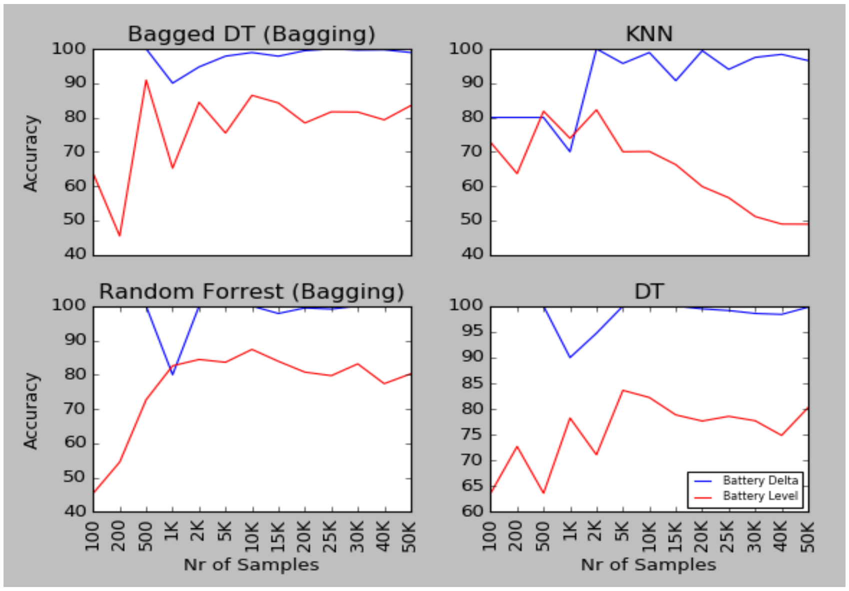


Fig. 88. Prediction accuracy: battery usage vs level (ensemble)

7.3.2.3 Labelling Delimitation (Threshold Boundary)

The objective of this experiment is to determine the relationship between prediction accuracy and the positive/negative delimitation threshold value in binary classification. In the following experiment, the labelling delimitation line, which defines the boundary between normal and anomaly data, is regulated to observe the response from prediction models. For this experiment, a 5000 sample data set is used with the delimitation value ranging from 0 to 15 percent of the battery usage. The delimitation value indicates how much noise (%) is added to battery usage to produce an anomaly (negative) sample set, required for the training process. The following diagram summarizes the findings of this experiment.

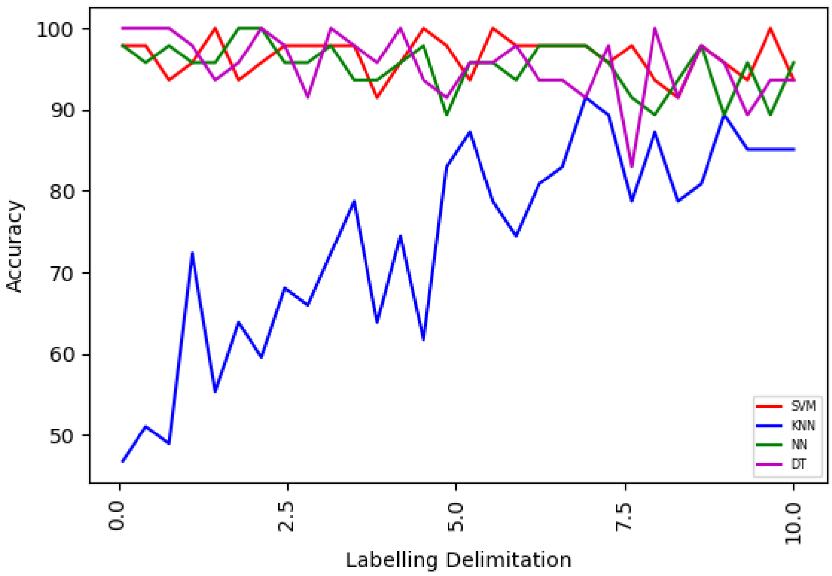


Fig. 89. Relationship: labelling delimitation vs prediction accuracy (normalized data)

The above diagram demonstrates similar behaviour among SVM, NN and DT based classifiers and they are able to obtain higher prediction accuracy with smaller threshold (label boundary) value. However, model based on KNN tend to produce lower prediction accuracy with smaller threshold values and model based on KNN demonstrates a positive relationship between prediction accuracy and the label demarcation threshold value. In essence, by adding marginal noise to battery usage, satisfactory, higher prediction accuracy can be achieved.

7.3.2.4 False Positive/Negative Rates

In the following set of experiments, false positive (FP) and false-negative (FN) rates are evaluated against several factors, including training set size, prediction model aging process, classification algorithm, and delimitation parameters. All experiments in this section utilize the noise injection technique to generate label data required for training and performance analysis processes.

The first experiment investigates the relationship between the labelling boundaries (delimitation) and the false positive (FP)/false negative (FN) rates. 5000 samples are used to train prediction models and 10000 samples, extracted from the same population, are used to evaluate performance, including prediction accuracy and FP/FN rates. The following diagram summarizes the findings of the experiment.

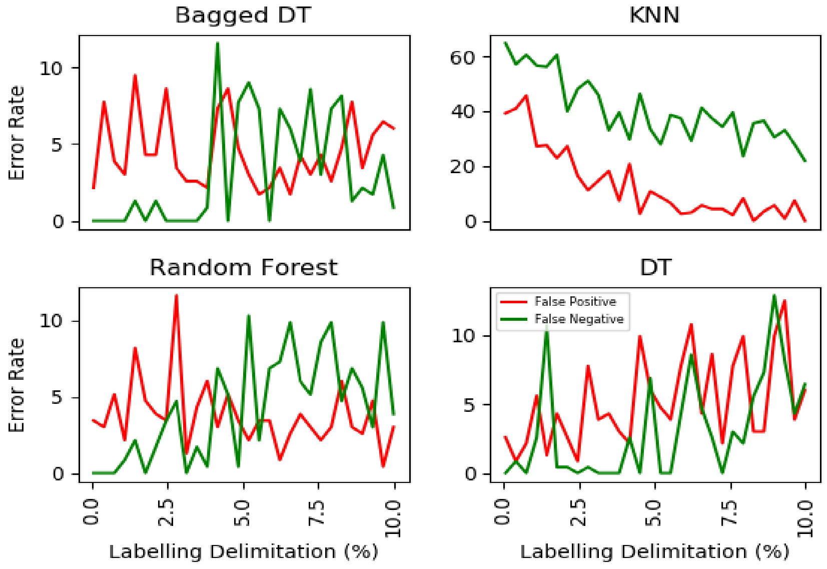


Fig. 90. Relationship: labelling delimitation vs false positive/negative (ensemble)

The above diagram demonstrates less than 10% false positive and negative rates with all but KNN based classifiers with threshold value less than five percent. The above results confirm that carefully selected classification algorithm and parameter settings are able to produce highly effective prediction models.

The objective of the following experiment is to determine the relationship between the number of samples used in the training process and false positive/negative rates. Similar to the previous experiment, 10000 samples are used to evaluate prediction accuracy and FP/FN rates. Prediction models (battery usage as an input parameter) based on both neural networks (NN) and support vector machine (SVM) with default parameters were unable to provide satisfactory FP/FN rates and prediction accuracy. Therefore, those two algorithms are replaced by ensemble methods (bagged DT, random forest) to determine the relationship between training sample size and the performance indicators (prediction accuracy, false positive and false negative rates). The following diagram summarizes the findings of the experiment.

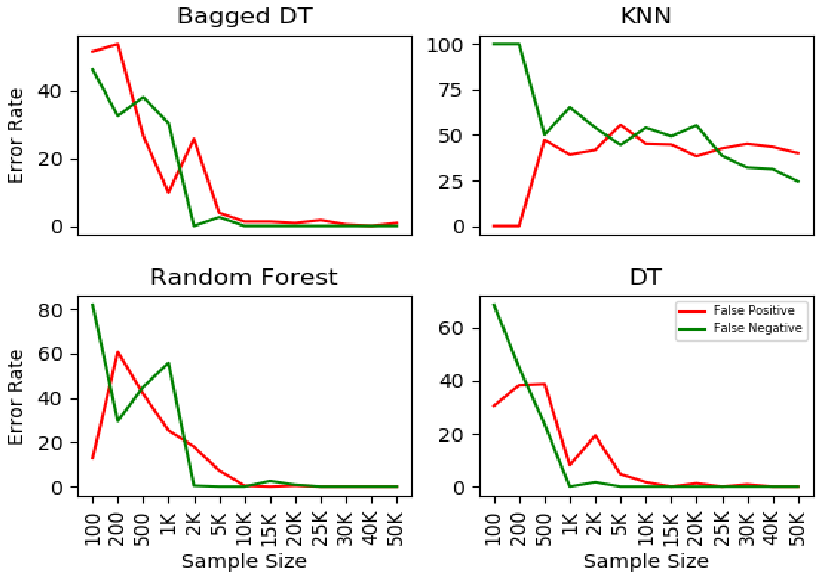


Fig. 91. Training set size vs false positive/negative rates

The above diagram demonstrates significantly high false positive and false negative rates, with smaller training set sizes, for all four classifiers. However, performance (prediction accuracy, false positive/negative rates) is drastically improved with larger training data sets. All classifiers but KNN are able to reduce false positive/negative error rates with prediction models trained with 2000 or more samples.

The final experiment in this section attempts to determine the relationship between false positive/negative rates and the prediction model aging process. Similar to the previous experiment, SVM and NN are replaced by two ensemble models. A 10000-sample data set is used to train the prediction model and several, 2000-record test sample sets are collected in fixed time intervals to evaluate the prediction accuracy and other performance indicators. The following diagram demonstrates the findings.

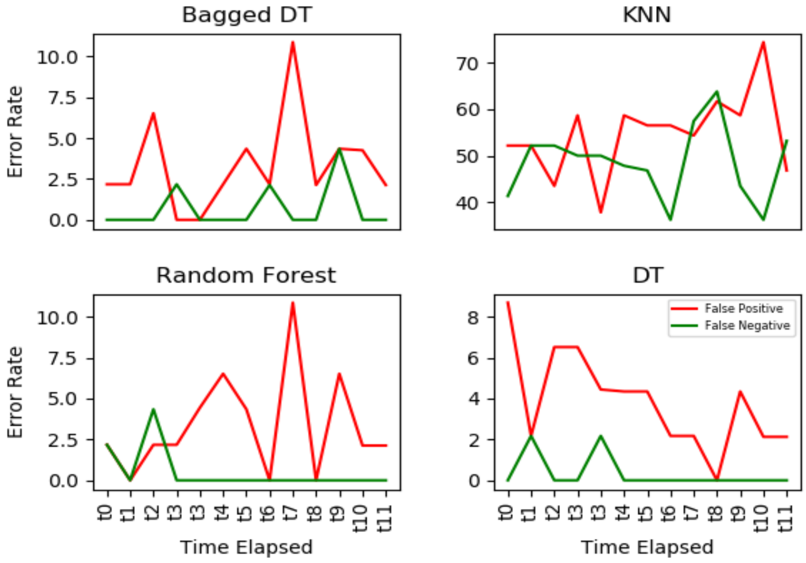


Fig. 92. Time elapsed vs false positive/negative rates

The above diagram confirms low false positive/negative rates with three classifiers. Similar to previous experiment results, a classifier based on K-nearest neighbours (KNN), is unable to produce usable results.

7.3.2.5 Model Aging Process (Retention Factor)

In this section, the relationship between prediction accuracy and the time elapsed from the training process (also known as prediction model aging process) has been investigated. For the training process, three different data sets (500, 1000 and 2000 samples) are used. For the prediction evaluation process, 12 different data sets (2000 samples each), collected in a fixed time interval are used. The following diagram depicts the outcome of the experiment. Furthermore, the following two experiments are based on battery level (instead of battery usage) as an input parameter. (Experiments related to model aging process for battery usage value can be found after following experiments)

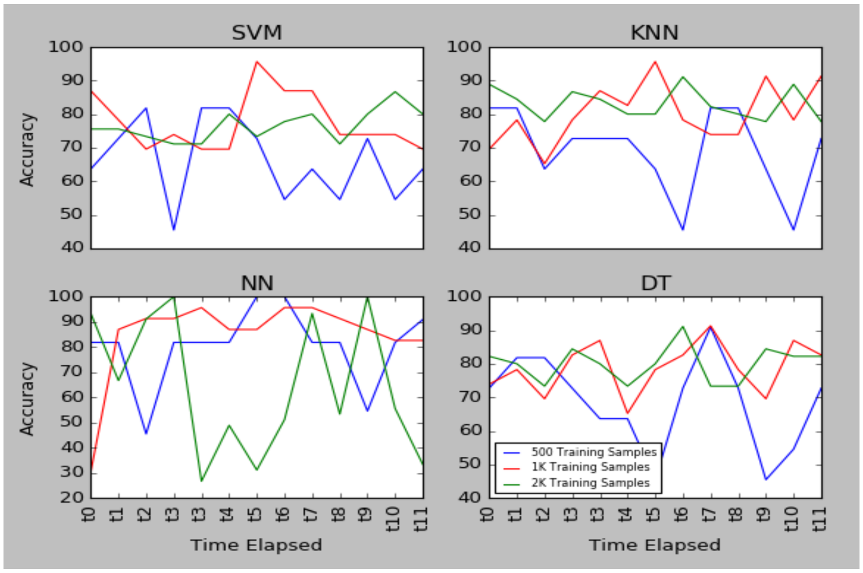


Fig. 93. Model aging process vs training set size (fixed, level)

The above diagram is based on a data set collected from an environment with a fixed lower battery level replacement process. However, as previously mentioned, the battery replacement process can be dependent on various factors, including instructions defined in the battery maintenance policy document associated with a particular business operation. In the following, the previous experiment is repeated with a dataset collected from an environment, where, a variable low battery level is used.

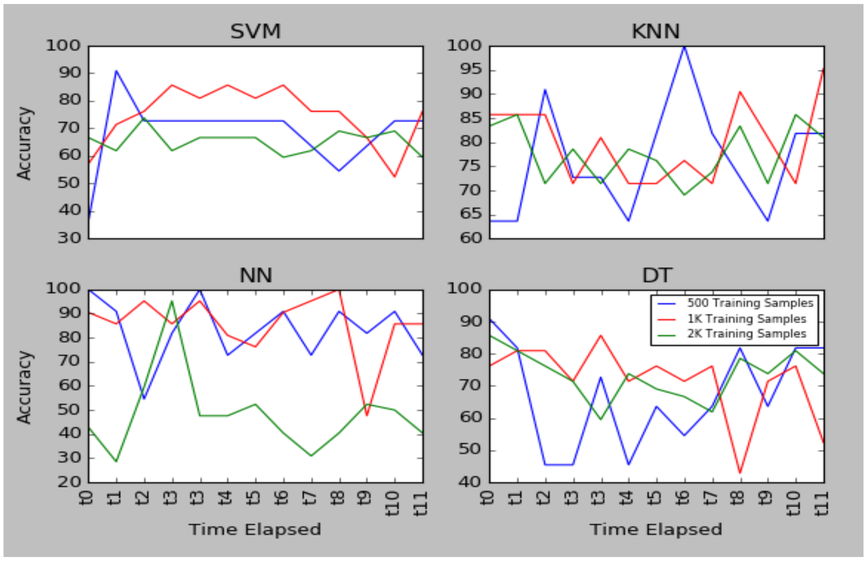


Fig. 94. Model aging process vs training set size (variable, level)

The previous two experiments confirm, a marginally, higher accuracy can be achieved with a data set collected from an environment with a fixed low battery level, maintenance procedure. Hence, strict battery maintenance rules can also improve overall security by improving anomaly detection accuracy.

In the previous two experiments, the battery level (instead of battery usage) is used to examine the prediction model's aging process. In the following experiment, battery usage (battery level change) and time are used to build prediction models. Similar to previous attempts, both SVM and NN are replaced by ensemble models to investigate whether higher accuracy could be achieved using ensemble methods. The corresponding experiment result is depicted in the following diagram.

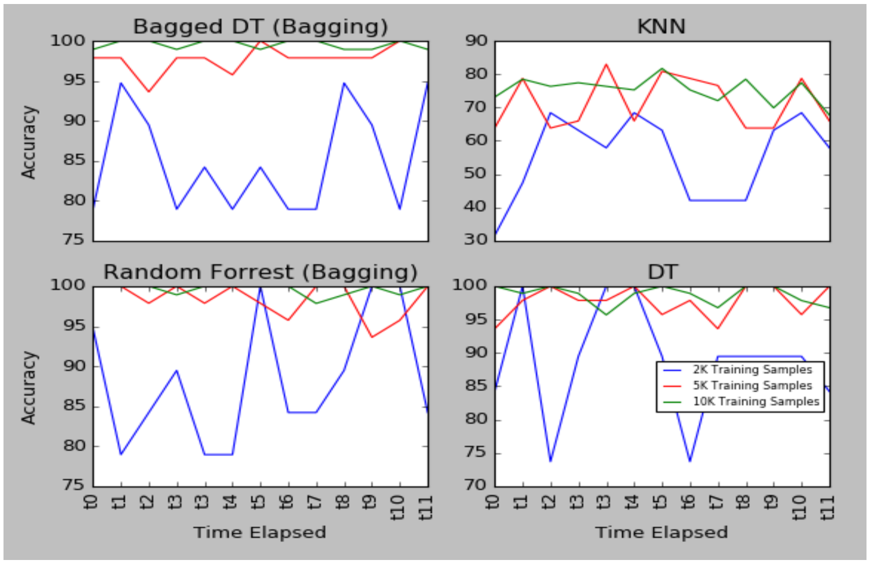


Fig. 95. Model aging process vs training set size (ensemble)

The above diagram confirms that higher accuracy can be achieved by carefully, selecting correct classification algorithms and corresponding parameters. All four classification models are able to provide consistent accuracy over a longer period, with the models trained in larger sample sets.

7.3.2.6 Node Generated Anomalies

One of the challenging elements of this work was to generate adequate anomaly data to train the anomaly prediction models. A few different techniques including noise-injection and sequence-permutation have been used to generate synthetic anomaly data set. Even though such an approach could be used to determine the relationship between threshold values of anomalies and the performance indicators, it is important to evaluate the behaviour of a model with more realistic anomalies. In the following, a number of experiments are conducted using anomaly data generated by the low powered nodes. For the following experiments, anomaly data is generated in individual nodes using random battery level values. The intuition behind such an approach is that a potential attacker must be able to correctly determine the battery status of the victim node to successfully spoof the corresponding node.

In the following experiments, five different data sets are generated using different anomaly ratios (10% to 50%). The following diagram demonstrates the distribution of both normal and anomaly data.

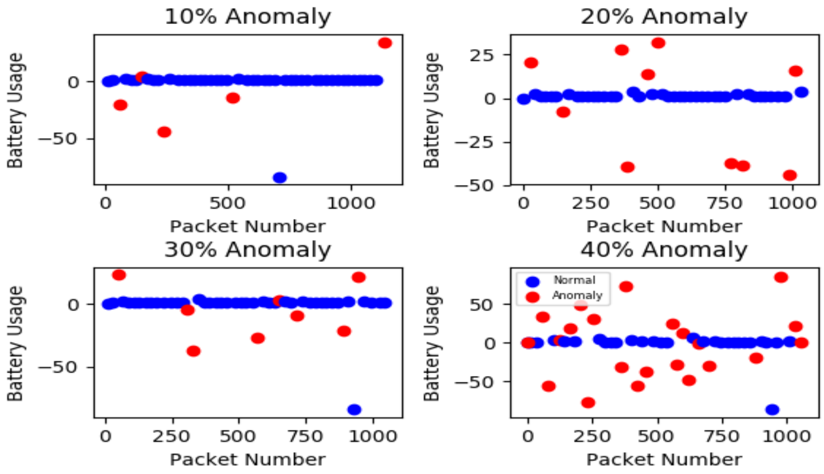


Fig. 96. Data distribution of anomaly & normal data

The above diagram demonstrates a number of extreme battery usage values for normal data. The reason for such a spike is that batteries are replaced according to the operational procedure when the lower threshold value for battery level or the battery-replacement due date is reached.

In the following experiment, the behaviour of number of samples in the training set is evaluated. In this experiment, data set with 20% anomalies are utilized and the following diagram summarizes the results.

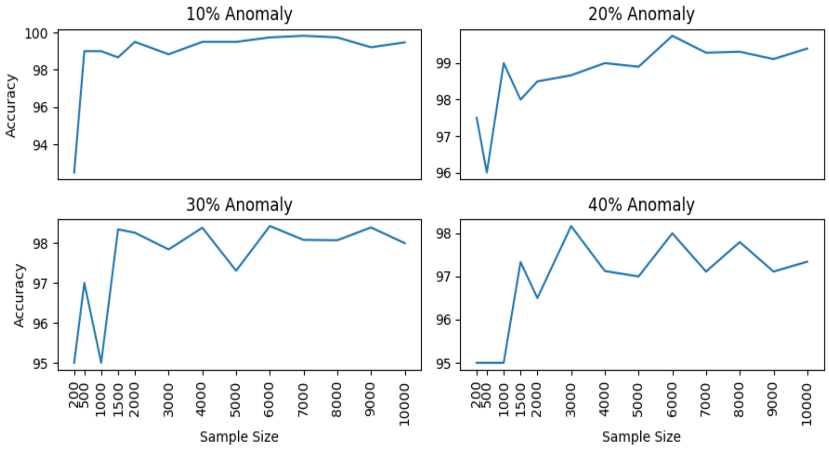


Fig. 97. Prediction accuracy against anomaly ratio and training set size

The above diagram demonstrates similar tendencies for all four anomaly rates, where, the prediction accuracy increases with the larger data set. However, the diagram confirms that over 96% prediction accuracy can be achieved with a significantly smaller training set (less than 200 samples) by carefully selecting the classification algorithm.

In the final experiment related to node-generated anomalies, training set size and normal-anomaly data ratio against the prediction accuracy is compared. The following diagram demonstrates the relationship between training set size, anomaly ratio and the prediction accuracy. A prediction model based on the Decision Tree algorithm is utilized in the following experiment.

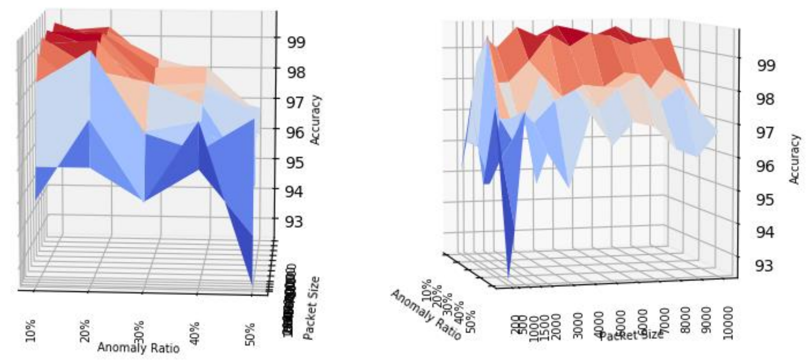


Fig. 98. Relationship between prediction accuracy, anomaly ratio and training set size

The above diagram confirms the previous findings that with the higher anomaly ratio, the prediction accuracy tends to decrease, while the models trained with larger training sets are able to improve the prediction accuracy marginally (A potential conclusion for such a behaviour is discussed under the IEEE 802.15.4e/TSCH characteristics in Experiment Results).

7.3.2.7 Time/CPU/Memory Utilization

In previous experiment sections, resource utilization has been investigated. However, in this section a couple of new machine learning techniques are utilized to improve the prediction accuracy. Therefore, resource utilization of different machine learning methods is compared in the following section. Sample sets listed in the following table are used to train prediction models and a sample set with 5000 samples, is utilized to evaluate the prediction model. The resource utilization for the corresponding test is summarized below.

The following table summarizes the time (seconds) required for each operation.

Size	SVM	KNN	NN	DT	BDT	RF
200	0.800	0.805	1.070	0.796	0.939	0.878
500	0.819	0.819	1.520	0.825	0.970	0.956
1K	0.853	0.860	0.879	0.858	0.994	0.942
2K	0.989	0.953	0.944	0.924	1.061	1.064
5K	1.151	1.155	3.338	1.126	1.297	1.236
10K	1.157	1.489	1.487	1.538	1.641	1.604
15K	1.903	1.825	1.868	1.806	2.001	1.901
20K	2.274	2.201	2.161	2.163	2.387	2.326
25K	2.640	2.574	2.582	2.517	2.698	2.620
30K	3.048	2.881	2.876	2.810	3.042	2.937
40K	3.897	3.552	3.513	3.535	3.798	3.797
50K	4.687	4.228	4.237	4.195	4.482	4.310

Tab. 20. Training set size, classifier vs resource utilization (time)

The following table summarizes the CPU usage (%) for each operation.

Size	SVM	KNN	NN	DT	BDT	RF
200	108.47	108.41	162.47	108.77	108.82	108.46
500	108.75	108.48	359.15	108.85	108.89	108.30
1K	108.76	108.80	298.15	108.85	108.67	108.90
2K	108.81	108.83	233.70	108.87	108.85	108.84
5K	108.86	108.76	338.18	108.86	108.77	108.85
10K	108.81	108.91	329.49	108.85	108.89	108.85
15K	108.79	108.88	327.78	108.46	109.32	108.94
20K	108.80	108.86	243.90	108.81	108.87	108.83
25K	108.50	108.78	291.21	108.85	108.90	108.77
30K	108.84	108.89	307.89	108.85	108.83	108.90
40K	108.82	108.79	326.26	108.79	108.64	108.82
50K	108.90	108.83	308.22	108.88	109.10	108.92

Tab. 21. Training set size, classifier vs CPU utilization

The following table summarizes the memory (MB) usage for each operation.

Size	SVM	KNN	NN	DT	BDT	RF
200	135.99	138.16	139.24	135.58	136.20	136.29
500	135.87	135.58	138.58	136.08	136.20	135.81
1K	136.24	136.11	138.81	135.90	135.93	135.99
2K	136.34	136.07	139.07	136.29	136.52	136.18
5K	137.38	136.70	139.19	136.73	136.86	136.51
10K	137.01	137.16	139.81	136.76	136.53	136.56
15K	138.66	138.55	142.87	138.82	138.84	139.26
20K	139.89	140.10	142.35	140.02	140.38	139.91
25K	141.68	141.58	143.94	141.35	141.37	141.17
30K	141.42	139.72	144.03	139.76	139.68	139.33
40K	145.70	145.75	147.34	145.27	145.29	144.96
50K	141.77	148.21	149.95	148.02	148.64	148.07

Tab. 22. Training set size, classifier vs memory utilization

The above tables demonstrate similar resource utilization among most classifiers. However, the prediction model based on neural networks (NN) demonstrates considerably higher CPU utilization.

7.3.2.8 Result Summary

In this section, the battery level/usage of a low powered node has been examined to determine whether a finite number of potential patterns (behaviours) can be identified. Subsequently, these patterns are used by machine learning methods to build prediction models that can be used with different security control mechanisms, including anomalies detection and data origin authentication, to protect low powered wireless networks. The following list summarizes the most important findings of the experiments completed in this section.

- Regression analysis with battery power usage data:
    - Data set with multiple battery life cycles are not able to produce usable prediction model with linear regression (e.g. 0.02 correlation coefficient obtained with 5000 sample set).
    - None linear regression models can be used to produce drastically improved prediction models (e.g. 85 percent accuracy with support vector regression (SVR) and 97 percent accuracy with random forest regression (RFR) using 5000 sample training set).
    - Larger training sets were unable to increase the prediction accuracy significantly, with non-linear regression models.
    - Battery replacement procedure doesn't affect the prediction accuracy significantly.
- Classification models based on battery power usage:
    - Prediction models, trained with a sample set size of single cycle of battery life, using battery level as an input parameter were are able to produce approximately 80 percent prediction accuracy.
    - Prediction models based on ensemble methods were able to improve the prediction accuracy. For instance ensemble methods utilizing 'Bagging' technique are able to produce over 80 percent prediction accuracy, with different sizes of training sets. However, ensemble methods utilizing a 'boosting' technique are able to produce higher prediction accuracy only with models trained with dataset extracted from a single cycle of battery life.

- Different labelling techniques used to train prediction models were only able to produce a smaller variance. For instance, the noise injection method has marginally higher accuracy compared to the labelling method, based on a sequence permutation technique.
- Prediction models based on battery usage as an input parameter were able to perform consistently better than models using battery level as an input parameter.
- No significant difference of false positive/negative rates in models trained with data collected from environments, with fixed and variable low battery maintenance policies.
- By carefully selecting classification algorithms and training sample sets, over 96 percent prediction accuracy, and less than five percent false positive/negative rates can be achieved.
- Besides NN which has utilized a higher CPU rate, most classification models have consistent resource utilization map, including memory, CPU and processing time.

7.4 Network Layer Characteristics

7.4.1 Classification Methods

7.4.1.1 Sample Size

The objective of this experiment is to understand the relationship between the training sample set size and prediction accuracy. Using network layer characteristics 13 different Network layer attributes of IEEE 802.15.4e are evaluated in this section, with the following combinations of those attributes tested.

Parameter Set 1	frame.len, wpan.frame_length
Parameter Set 2	ipv6.plen, wpan.frame_length
Parameter Set 3	wpan.src64, wpan.frame_length, frame.time_delta
Parameter Set 4	ipv6.src,wpan.src64, wpan.frame_length, frame.time_delta

Tab. 23. Input feature sets (combinations) used in experiments

An arbitrary mechanism is utilized to assign input features to individual parameter set groups. However, statistical methods such as PCA or cross-validation could be used to identify optimal parameter sets to improve the prediction accuracy. However, manual tweaking of hyper-parameters or use of various model optimization techniques are not investigated in this study. A number of different sample sets, ranging from 100 to 50000, are used in the training and testing process. Similar to previous tests, four classification models (SVM, KNN, NN and DT) are examined, using the four aforementioned parameter sets. The following diagram depicts the relationship between the prediction accuracy, number of samples, classification algorithm and the input parameters.

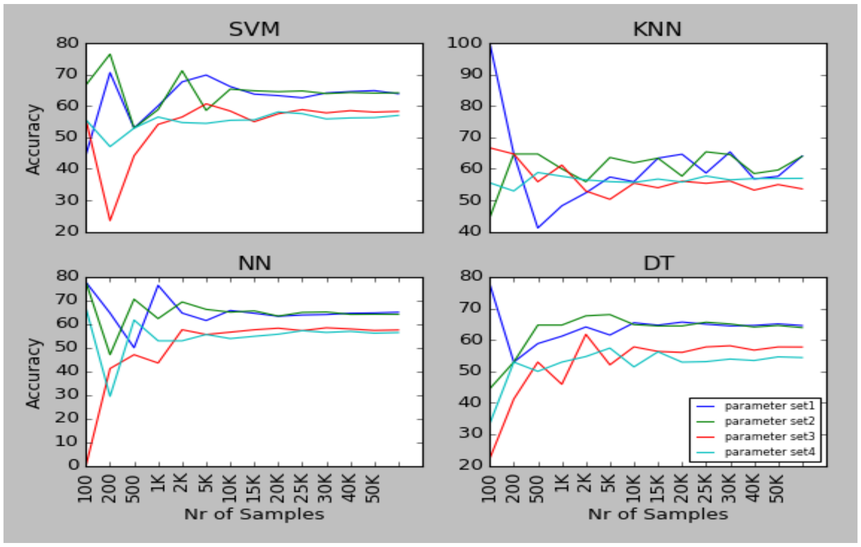


Fig. 99. Relationship: training set size, feature set vs prediction accuracy

The preliminary result indicates, regardless of the number of samples used in the training process, for all four classifiers, prediction accuracy doesn't typically exceed the 70 percent mark. Nevertheless, such a low accuracy may not be adequate for an anomaly detection mechanism for a low powered wireless network. The labelling technique used in this experiment is based on the sequence permutation of a single attribute and the result indicates that the variance created by permutation of a single attribute may not be sufficient enough to generate two distinct classes. Instead of a permutation, noise can be added to normal data in an effort to generate a negative or an anomaly data set for the training process (the relationship between threshold value of noise and the prediction accuracy is tested in Noise Threshold/Labelling Delimitation subsection). The following result is a repetition of the previous test, substituting sequence permutation with a noise injection technique.

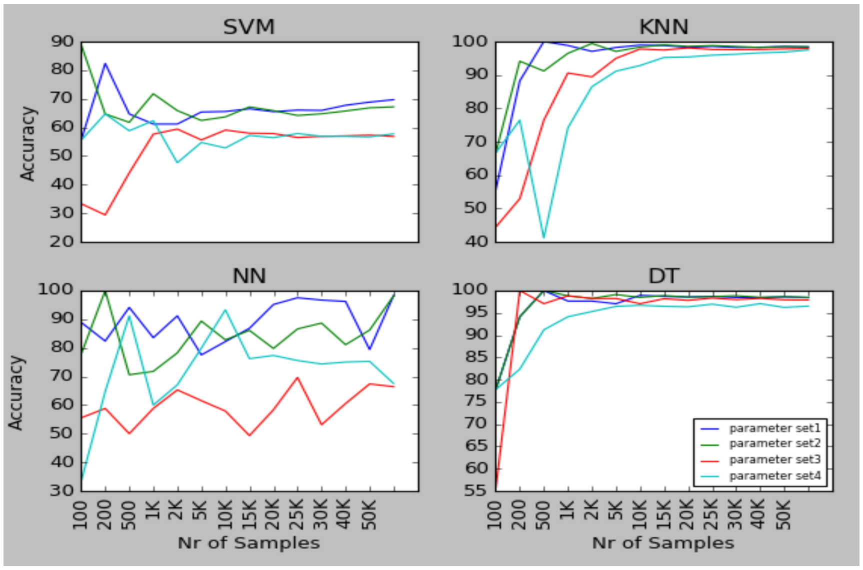


Fig. 100. Relationship: training set size, feature set vs prediction accuracy (noise injected)

The diagram confirms that classification algorithms react differently with the use of noise injection technique. For instance, compared to the sequence permutation labelling method, with the use of noise injection technique, both KNN and DT are able to improve the prediction accuracy significantly. In essence, it is possible to achieve higher accuracy by carefully selecting a classification model, the training set size and noise threshold values.

Binary classification models are based on two possible outputs and an anomaly detection system can be a binary classification model, where a particular packet is classified as normal or as an anomaly. However, if a particular prediction model is used to classify packet/flow to more than two classes, a different labelling mechanism is required. For instance, if the objective of a prediction model is to determine the authenticity of data origin, multi-label classification models are required, since a possible packet/flow could originate from any node belonging to the corresponding network. To satisfy a multi-label (non-binary) classification requirement, different labelling technique must be employed. If a packet flow can be described using a set of input parameters, such as ipv6.src, wpan.src64, wpan.frame_length, and frame.time_delta, the corresponding value of a single parameter is dependent on the values of rest of the parameters. The following formula summarizes the above statement.

Flow ≈ [parameter_1, parameter_2, …, parameter_n]

parameter_k ≈ [parameter_1, …, parameter_k-1, parameter_k+1, … ,parameter_n] where k Є (1 ... n)

In the following experiment, a parameter which belongs to each parameter set is removed from the input parameter list and used as a label to train the prediction model. The objective of this approach is to understand the relationship between the number of training samples and the prediction accuracy when an input parameter is used as an output label. The following parameters are used as output (label) for each group.

	Label
Parameter Set 1	frame.len
Parameter Set 2	ipv6.plen
Parameter Set 3	wpan.src64
Parameter Set 4	ipv6.src

Tab. 24. Output (label) list for Parameter Sets

The following diagram describes the findings of the experiment.

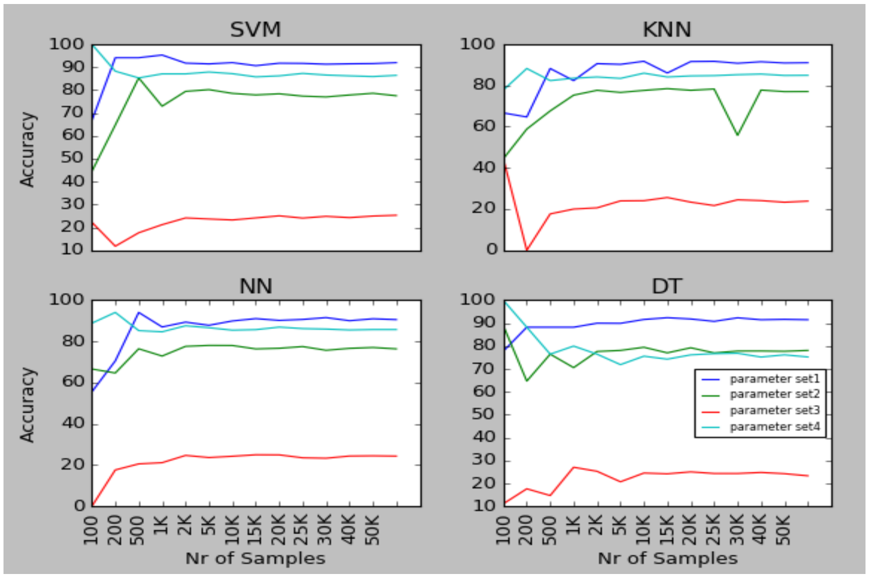


Fig. 101. Relationship: training set size, feature set vs prediction accuracy (multi-label)

The result indicates a significant improvement in the accuracy of multi-label (non-binary) classification models. Furthermore, the diagram also confirms the importance of carefully selecting the input parameters. For instance, according to the above diagram, parameter set 1 predicts higher accuracy, while parameter set 3 provides dismal prediction accuracy. Different techniques such as principal component analysis (PCA) can be used in machine learning to identify the most influential input parameters.

In the following experiment, a subset of samples, sent from a single node is used to train the prediction model. The following diagram summarizes the resulting comparison between prediction models trained using data belonging to all nodes and data belonging to a single node. For this experiment, attributes ipv6.plen, wpan.src64, frame.len, wpan.frame_length, and frame.time_delta are used in the input feature list and a filtered dataset is obtained by using attribute wpan.src64 as a filter.

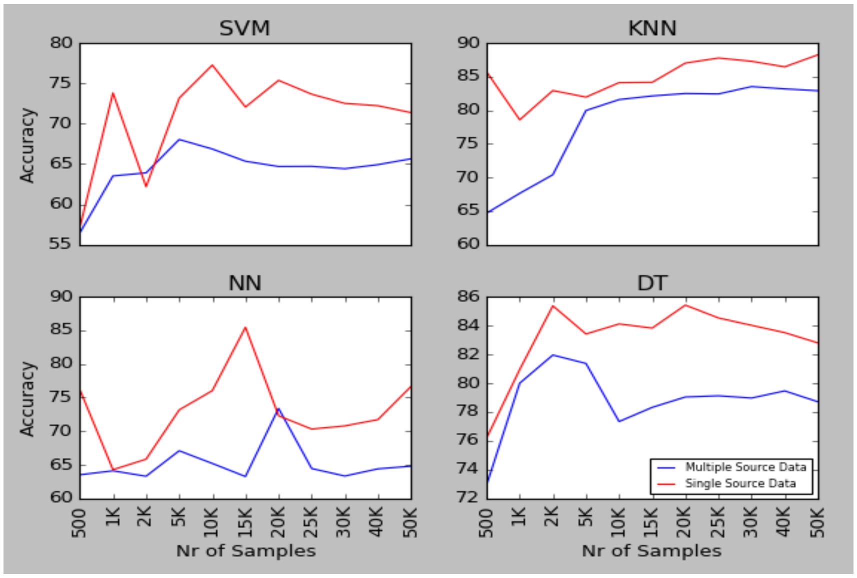


Fig. 102. Training size vs prediction accuracy (single vs multi-source)

The result indicates that considerably higher prediction accuracy can be obtained using multiple prediction models. In essence, if a particular network comprises N nodes, it is possible to achieve higher prediction accuracy by creating a separate classification model for each individual node. Such prediction models can be used with different security controls including anomaly detections and data origin authentication.

7.4.1.2 Noise Threshold/Labelling Delimitation

The labelling delimitation process defines the boundary between normal and anomaly data. By increasing the variance between normal and anomaly data, prediction accuracy could be improved. However, larger delimitation values could lead to higher false positives and negatives. Therefore, the demarcation between normal and anomaly data is dependent on the security objectives of a particular environment. Some attributes may generate anomalies only with a marginal deviation. However, other attributes could allow a higher fluctuation of data still within an acceptable range. The following experiment is done to determine the behaviour of the prediction accuracy and the labelling boundary. The experiment is performed with a 5000 training set, using a binary classification model. The boundary-value of the labelling process is examined for a single parameter, ranging from 0 to 15 percent. The following diagram depicts the result of the experiment.

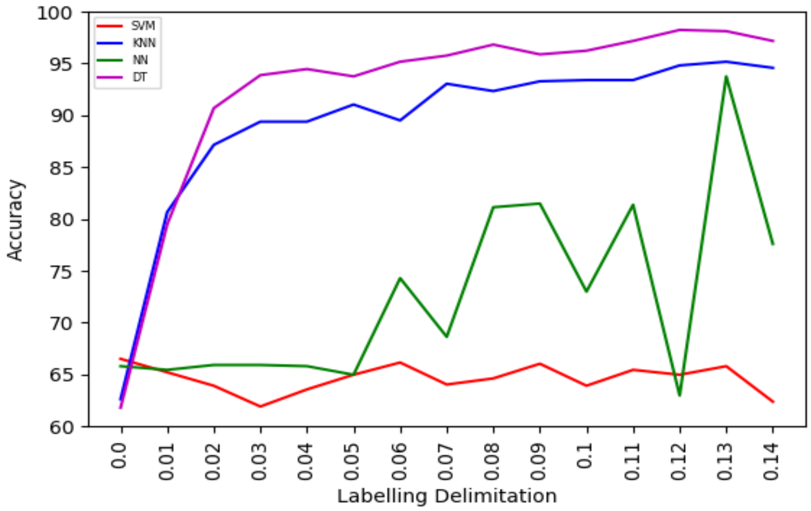


Fig. 103. Labelling delimitation vs prediction accuracy

The above diagram confirms that different classification algorithms respond differently to the labelling delimitation adjustment. However, all classifiers, but SVM, demonstrate an increase of accuracy with a larger boundary value.

7.4.1.3 Model Aging Process (Retention Factor)

One of the important characteristics of a well-designed prediction model is a generalization. In essence, a well-generalized model should generate comparable predictions for unseen data. The following experiment is performed to understand the behaviour of prediction accuracy with unseen (future generated) data. The following experiment uses a prediction model trained with 5000 samples. The evaluation process is carried out using multiple 2000 data samples collected at a fixed time interval. The following diagram summarizes the findings of the above experiment.

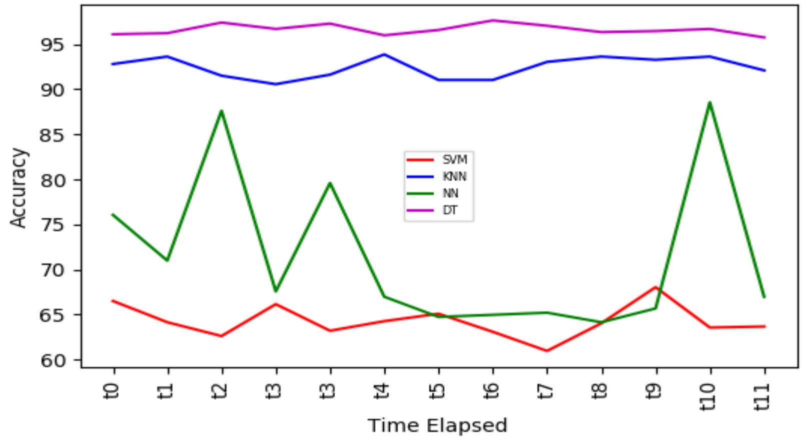


Fig. 104. Model aging process

The diagram demonstrates mixed prediction accuracy, with four default classification models used in this work. Prediction models based on SVM and NN algorithms demonstrate low accuracy rates.

7.4.1.4 Binary/Multi-label

The objective of the following experiment is to compare the relationship between training sample size, performance indicators (prediction accuracy, FP and FN) and the classification model type (binary, multi-label). The following diagram depicts the outcome of the experiment.

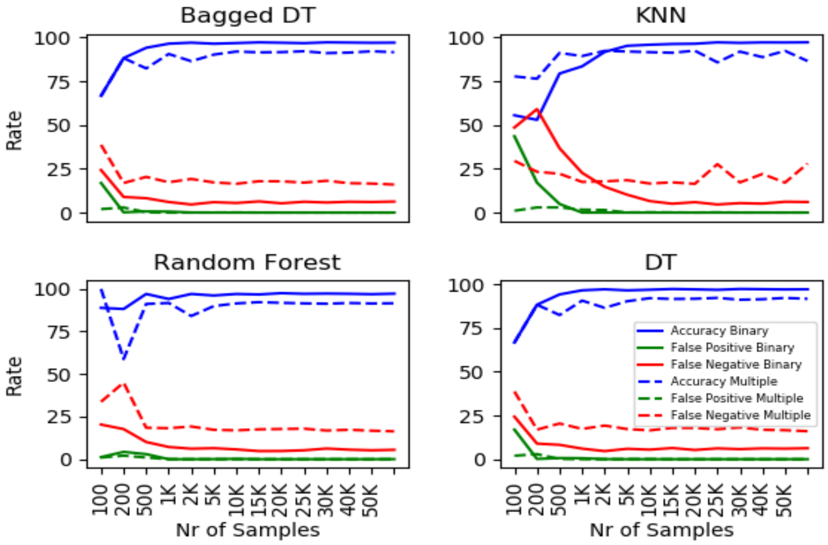


Fig. 105. Binary vs multi-label classification model comparison

The above result indicates that binary models are able to produce marginally higher prediction accuracy and lower FP and FN rates. Furthermore, individual graphs confirm that training set size doesn't influence the performance (prediction accuracy, false positive/false negative rates), significantly.

The following experiment compares the binary and multi-class prediction model's performance for an unseen dataset. 5000-sample training set and several sample-sets (sizes of 2000 records), collected in fixed time intervals, are used to evaluate the prediction accuracy and FP/FN rates. The following diagram depicts the experimental result.

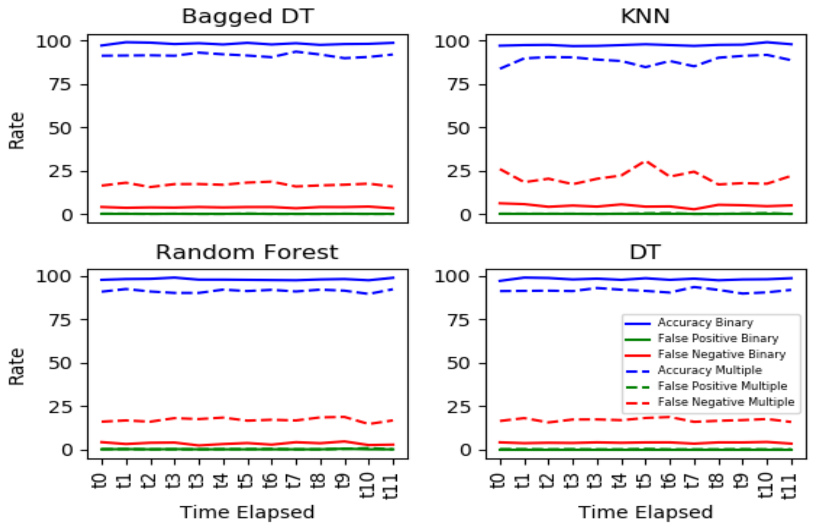


Fig. 106. Binary vs multi-label model aging process

According to the above result, binary classification models demonstrate a significant performance gain compared to the multi-label classification models.

7.4.1.5 Single Node/Multiple Nodes

In the following experiments, data is clustered, based on the source identity to classify into multiple classes. The objectives of the following experiments are to compare the overall prediction accuracy and the prediction accuracy of sub-models.

In the first experiment, performance aging process is compared for the single cluster and multi-cluster prediction models. The following diagram is generated using two different prediction models. The first model is based on an unfiltered data set where data originating from all nodes are used in the training process. However, in the second test, data originating from a single node is used in the training process. Similar to the previous experiment, a

5000 sample set is used in the training process and multiple data sets, containing 2000 samples, collected at a fixed time interval, are used to evaluate the prediction accuracy of the model. The findings are summarized in the following diagram.

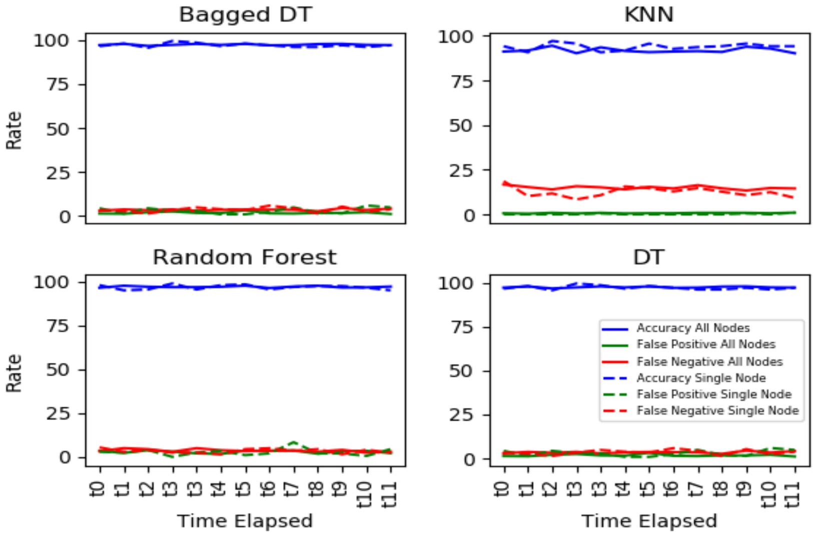


Fig. 107. Performance comparison: single node vs multi node data

The above diagram confirms that there is no significant difference between the overall prediction accuracy and the clustered prediction accuracy.

In the following, the above experiment is repeated. However in this experiment, instead of performance aging process, the influence of training set size on clustered and non-clustered prediction models are evaluated. Several sample sets are used to train prediction models and to compare the performance indicators (prediction accuracy, false positive and false negative rates) for both, the main model and the clustered-models. The following diagram demonstrates the findings of the experiment.

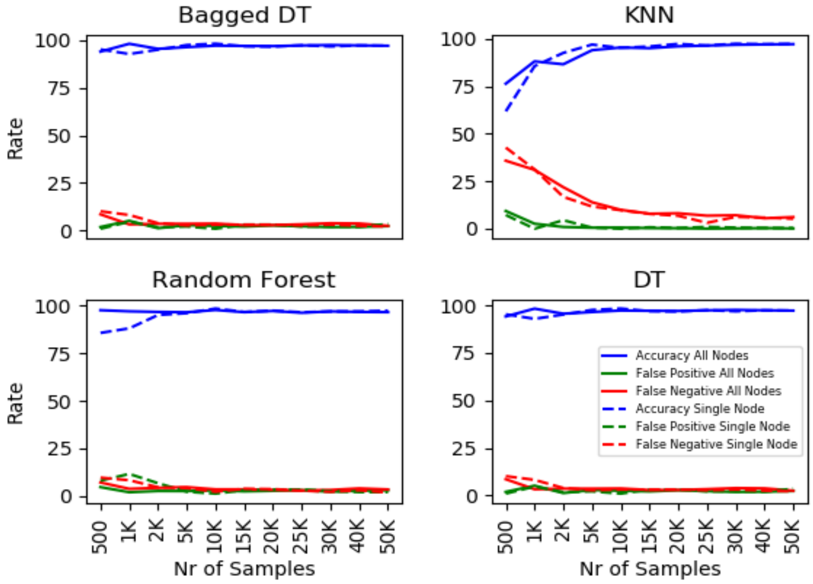


Fig. 108. Single vs multi-node performance against training set size

Both of the previous experiments confirm that clustered prediction models don't improve the performance, significantly. The number of nodes in the network and the lack of distinctive differences of behaviour among nodes could contribute to the above result. Although clusterization doesn't improve the performance with the data set and the feature vector used in the above experiments, clusterization is a known technique to improve performance in machine learning models.

7.4.1.6 False Positive/Negative Rates

The first experiment of this section (false positive/false negative), investigates the relationship between the training set size and the FP/FN rates. For this experiment, several sample sets ranging from 100 – 50000 samples, extracted from a single distribution is used. Each sample set is equally divided to generate a balanced label count and the permutation technique is used to generate a negative sample set, as required for the supervised learning process. Four default classification algorithms (SVM, KNN, NN and DT) are used to evaluate the false positive and false negative rates. For this experiment, attributes ipv6.plen, wpan.src64, frame.len, wpan.frame_length, and frame.time_delta are used in the input feature list. The following diagram describes the outcome of the experiment.

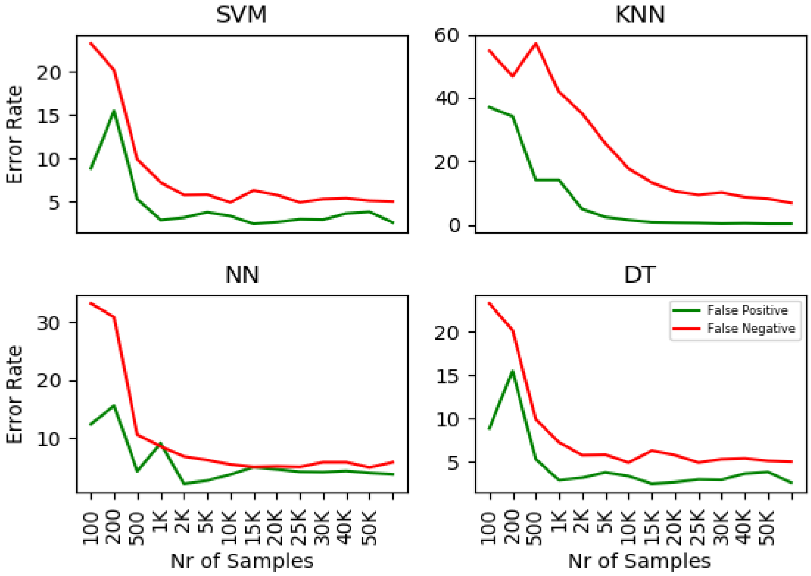


Fig. 109. Relationship: training set size vs false positive/negative rates

The above diagram demonstrates a similar trend among all four classifiers. All four classifiers are able to reduce both false positive and negative rates with models trained with larger data sets.

In the following experiment, the behaviour of false positives/negative rates against unseen future data is evaluated. (Data is considered to be unseen when such data is not used in the training process). The following experiment is completed using a 5000 sample set to train the prediction model. To determine the FP/FN rates, 2000 prediction samples collected at fixed time intervals, from the same population are used. The following diagram summarizes the findings of the above experiment.

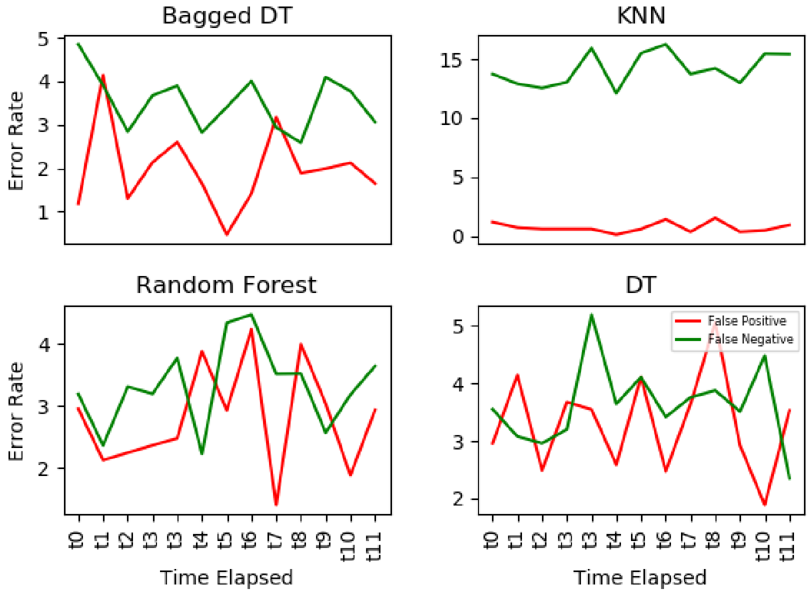


Fig. 110. Time elapsed vs false positive/negative rates

The diagram confirms that all four classifiers are able to hold comparable rates for false positives and false negatives over a longer period of time.

7.4.1.7 Time/CPU/Memory Utilization

Classification algorithms, data extraction and computational methods used in this section are similar to the methods used under the LR-WPAN characteristics section. Therefore, performance-related indictors should produce similar results.

7.4.1.8 Result Summary

The primary objective of this portion of experiments is to determine whether the network layer attributes, extracted from low powered wireless network data, operating in IEEE 802.15.4e/TSCH mode, can be used to predict behaviour of the corresponding network. Several factors including training set size, classification algorithm, noise impact and model aging process were evaluated to determine their influence on prediction accuracy, false positive and false negative rates. Furthermore, a number of different techniques, including ensemble machine learning methods were tested to improve the performance. The summary of the findings is listed below.

- Characteristics of Network layer attributes:
    - Node identification attributes such as source IP (ipv6.src), destination IP (ipv6.dst), source MAC (wpan.src64), destination MAC (wpan.dst64) were successfully able to cluster the network into a finite number of groups equal to available nodes in the corresponding network.
    - Frame.number, frame.time_relative and wpan.seq_no demonstrated a linear relationship with the time (time-series data).
    - Although zep.channel_id is used by IEEE 802.15.4e networks operating in TSCH mode to assign a channel for data exchange between two nodes, descriptive statistics confirmed a random channel allocation mechanism, utilized in the simulation.
    - A strong positive correlation (Pearson r value: 0.99) could be found between frame.number and the frame.time_relative. This confirms that TSCH based

networks operate in a synchronized (time) scheduling mechanism.

  - wpan.seq_no is used by the MAC layer as a flow control mechanism to control the data sequence. IEEE 802.15.4e MAC header utilizes 8-bit value to store wpan.seq_no. However, wpan.seq_no was able to generate a positive linear relationship with time (Pearson r value 0.94) with smaller data sets. Yet, different techniques, such as sliding time window, can be used to transform wpan.seq data to use in regression analysis.
  - Data flow diagrams confirmed recurring patterns in packet length (wpan.frame_len), source IP address (ipv6.src) and source MAC address (wpan.src64).
  - Several input parameter lists were able to divide the data set into multiple dynamic classes successfully, using meanShift clustering algorithm.
- Classification models:
  - With the use of sequence permutation as a labelling technique, the increment of training sample set size was unable to improve the prediction accuracy.
  - None of the default classification algorithms (SVM, KNN, NN and DT) were able to produce higher prediction accuracy rates (over 70 percent) with models trained using different sample sets (ranging from 100 to 50000 samples) and sequence permutation as a labelling technique.
  - Noise-injection labelling technique was able to improve the prediction accuracy, considerably.
  - With carefully selected input parameter list and classification algorithm, multi-label prediction models

trained with 1000 or more samples were able to produce over 90 percent prediction accuracy.

- Implementation of an individual prediction model for each cluster was able to improve the prediction accuracy significantly.
- By implementing cluster-based prediction models and using default classification algorithms, regardless of the training sample set size, over a ten percent improvement of prediction accuracy could be achieved.
- Carefully selected demarcation (labelling boundary) could increase the prediction accuracy with some classification models. However, some classification algorithms such as SVM responded negatively for the increment of the delimitation line.
- Ensemble methods were able to improve the prediction accuracy considerably. For instance, bagged DT and RF classification models were able to produce over 95 percent prediction accuracy with a small adjustment of labelling delimitation value.
- Retention Factor of prediction models trained with carefully selected network layer attributes was satisfactory; during the experiment, the time elapsed from the training process didn’t significantly impact the prediction accuracy.
- Compared to multi-label classification models, binary classification models were able to provide higher prediction accuracy, lower false positive and false negative rates for different sample sets and unseen data.
- Prediction models labelled with permutation technique seems to produce higher false-negative rates (around 30 percent with models based on SVM, NN and DT).

However, prediction models utilizing the noise-injection labelling technique were able to reduce the false positive and false negative rates to an acceptable level (less than eight percent).

7.5 Aggregated Model

In this work, four, distinctive, feature groups (IEEE 802.15.4e/TSCH, physical layer (wireless), low powered and network) based on the properties of low powered wireless networks and associated protocols were identified, and the features belonging to each group were evaluated separately. The prediction models based on each individual group performed, differently. Furthermore, group-specific bias, such as seasonal effects on wireless properties, could have influenced the performance of some prediction models. Such an influence can be neutralised through bundling multiple, prediction models, constructed by using attributes of different feature groups. Yet, it is important for each prediction model to be trained with data from a single distribution. However, as previously noted, data used in this work has come from several distributions and the feasibility of bundling multiple prediction models was limited.

In the following, a number of attributes, representing different feature groups (IEEE 802.15.4e/TSCH, physical, low powered and network), belonging to a single distribution are used to construct a model based on bundling multiple prediction models. Four attributes, namely, ASN (IEEE 802.15.4e/.TSCH), timestamp (physical), battery usage (low powered) and source address (network) are evaluated.

In the first experiment, five separate classification models are generated using five different classification algorithms (SVM, KNN, NN, DT and RF). All classification models are based on a binary classification and a single 2000-sample data set is used to train all prediction models. The aggregated classification model is based on a simple voting majority of the five classification models. For instance, if three or more prediction models identify a certain input as an anomaly, the aggregated classification model labelled data as an anomaly. In this experiment, the prediction accuracy of the aggregated classification model and the conventional prediction model (single classifier), based on neural networks is compared. (All aggregated models evaluated in this thesis are based on hard-voting mechanism where each model has equal say on the outcome) All models are based on similar parameters including the training set size. The experiment is repeated using a randomly selected, 2000-sample data set and the following diagram demonstrates the corresponding result for 100 rounds of the experiment.

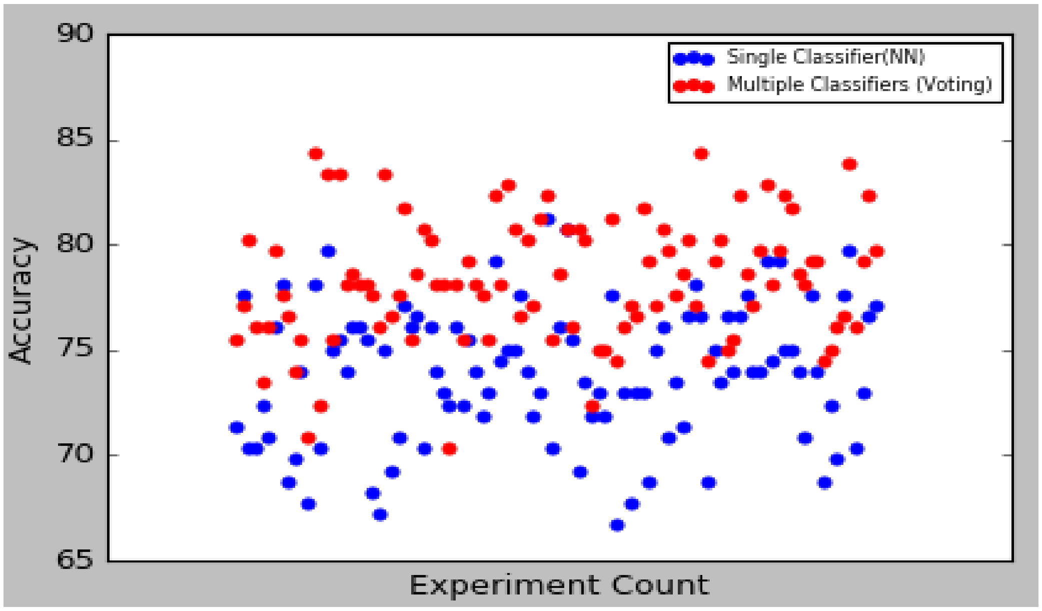


Fig. 111. Prediction accuracy comparison: single classifier vs aggregated classifier

The above diagram demonstrates consistent, higher prediction accuracy for the aggregated classifier model, based on the voting technique. Statistical data in the above experiment further confirms that a 4.2 percent improvement of prediction accuracy is achieved by aggregating multiple, prediction models.

In the following experiment, false positive and negative rates are compared in the aforementioned experiment. All experimental parameters, including training/testing data sets and classification algorithms, were kept intact and the following diagram depicts the corresponding results.

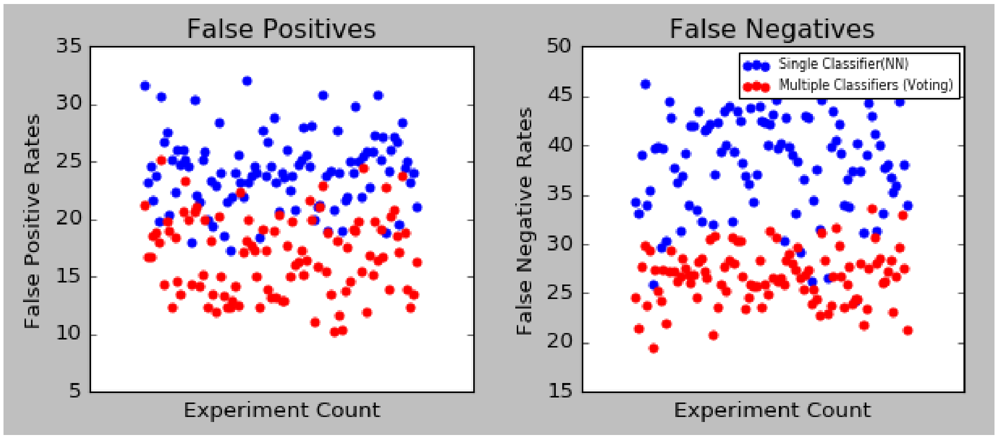


Fig. 112. False-positive/negative rates for single and aggregated classifiers

The above diagram confirms the false positive and negative rates can be significantly, reduced by using aggregated classification models based on the simple, voting technique. The above experiment has produced a 6.32 percent reduction in the false-positive rates and a 12 percent reduction in the false-negative rates, within the aggregated classification model.

In the following experiment, the prediction accuracy of models, based on a subset of features, belonging to individual feature groups is compared. Three, separate, prediction models were created using attributes belonging to each individual, feature group. The experiment is performed using a neural network based classification model, trained with a 2000-sample data set. A 2000-sample test set is used to evaluate the prediction accuracy. While the “All Features” classification model is based on all available attributes, the “Subset” classification model is based on a selected set of attributes belonging to each individual, feature group (IEEE 802.15.4e/TSCH, low powered, network). The experimental result is based on 100 repeatable tests using randomly extracted 2000-sample training and test data sets. The corresponding result is depicted in the following diagram.

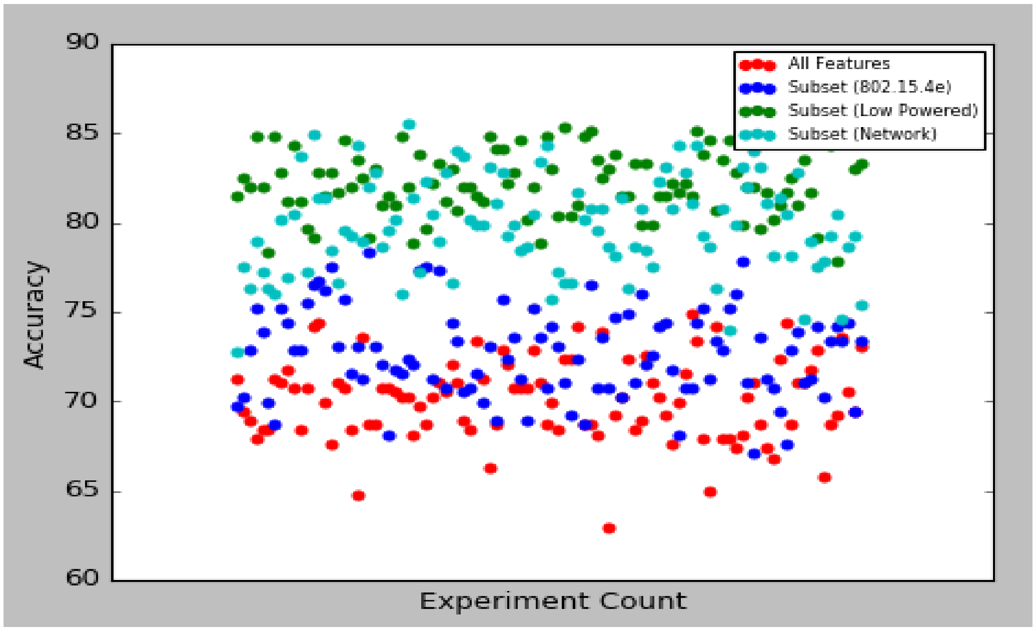


Fig. 113. Prediction accuracy of models based on features from individual experimental groups

The above diagram demonstrates a consistent accuracy in each prediction model. However, some prediction models were able to produce significantly, higher accuracy compared with a general model trained in all input features. In the above experiment, the classification models based on a subset of features belonging to an individual feature group is able to improve the prediction accuracy by 10.43 percent.

In the final experiment of this thesis, a comparison between a classification model based on all input features (base classifier) and a voted classification model based on an aggregation of multiple classification models is examined. Similar to the previous experiment, three classification models are constructed, utilizing the attributes of individual feature-groups. As previously explained, a voting mechanism based on a simple majority is used to determine the final classification group (output class). The following diagram describes the prediction process for the vote-based classification model.

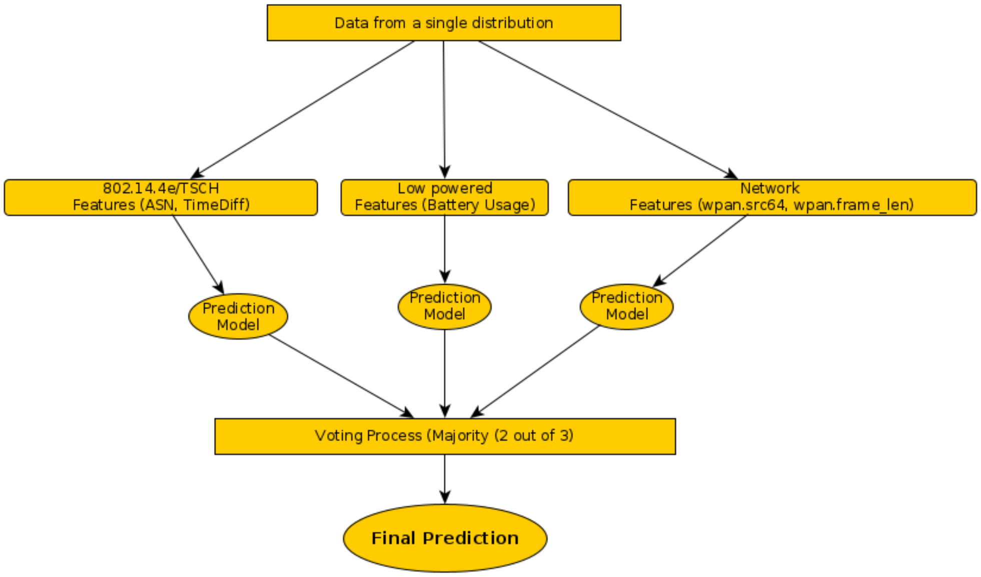


Fig. 114. The vote-based aggregated classification model

As described in the above diagram, label group which has minimum of two votes is selected as the final prediction.

A 2000-sample, data set is used to train each model and the experiment is repeated 100 times, using randomly selected 2000-sample test set, to obtain a more generalized result. The following diagram demonstrates the finding of the experiment.

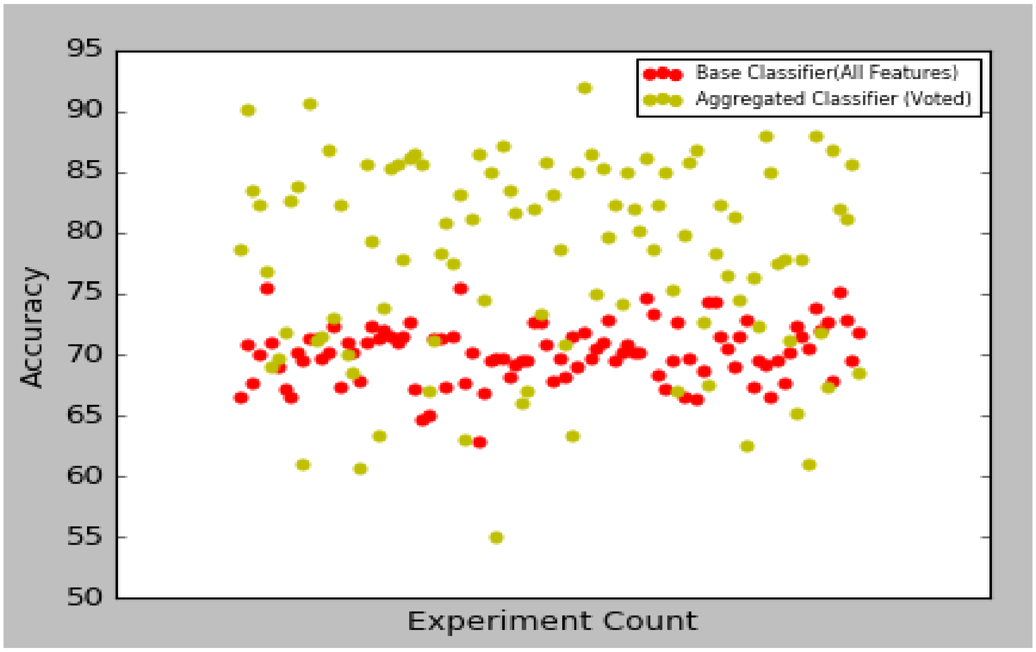


Fig. 115. Comparison between basic classifier (all features) vs aggregated classifier (voted)

The above diagram demonstrates a higher variance in prediction accuracy with the aggregated classifier. However, statistical analysis of the experiment indicate that there is an average of a 7.55 percent improvement in prediction accuracy; this is achieved by bundling multiple, sub-models, constructed using selected attributes from individual feature groups (IEEE 802.15.4e/TSCH, physical layer (wireless), low powered and network).

Chapter 8. Summary and Discussion

The objective of this thesis was to investigate the usability of low powered wireless characteristics and protocol-specific attributes to identify traffic anomalies in low powered environments, using machine learning methods. The work was organized into several sections to dissect critical elements related to the research. A comprehensive discussion about threat vectors, available prevention tactics and potential models to mitigate certain attacks, were proposed in previous chapters. In the following, a summary of critical elements discovered in this work and possible direction to further enhance the research is discussed.

Security

In the background research, a comprehensive investigation regarding the security of LoWSNs was conducted. In the following, some of the critical elements, uncovered with respect to the security of LoWSNs are outlined.

- Security objectives of low powered wireless data are implementation-specific. Several attributes including the nature of data, business objectives, jurisdictional laws and the financial constraints significantly, influence security objectives.
- Security of LoWSNs has a complex landscape and multiple security elements must work cohesively, to provide complete protection to the low powered data.
- Low powered data can be threatened on several fronts, and corresponding controls can be classified into three classes namely, physical, administrative and technical, based on the nature of the threat. Each of these classes comprises of a complex attack vector that can be used by adversaries.
- Heterogeneity of low powered nodes prevents implementation of node-based security controls.
- Low powered nodes may have a higher risk of exploitation due to their service-oriented nature.

- Low powered networks, operating in open network protocols, could be prone to known network attacks.
- The adaptability of expensive cryptographic techniques in low powered environments is highly unlikely.
- The open nature of the wireless medium drastically increases the attack surface of low powered wireless networks.
- Lack of physical layer protection in most low powered wireless networks introduces a new set of attacks, including disassociation, de-authentication and flooding (beacon, request, acknowledgments) and threatens the availability of low powered data.
- Protocol specific design flaws contribute to numerous new attacks, including back-off timer violation, clear channel assessment violation, acknowledgment manipulation, CAP violation and slot-time abuse attacks.

Adversaries may take advantage of the above-mentioned vulnerabilities and this might create a threat to low powered wireless data. A number of experiments were conducted to address those concerns and several recommendations were made. These recommendations included protocol enhancement and security control adaptations such as intrusion detection, access controls, encryption and authentication.

The Approach & Discoveries

The quest of this research was to design an anomaly detection mechanism, protecting data in low powered wireless environments. Although, a number of researchers have proposed models to detect anomalies in low powered wireless networks, using various techniques (discussed in related work), according to information found in the public domain, properties of low powered wireless networks and attributes of IEEE 802.15.4e/TSCH were never utilized to identify normal baseline operational parameters in LoWSNs using machine learning methods.

For this comprehensive investigation, experiments were divided into the following four groups: These were based on similarities in features.

- IEEE 802.15.4e/TSCH properties
- Physical layer (Wireless) properties
- Low powered properties
- Network layer properties

Several factors, including the number of training samples, number of nodes, data variance, classification algorithms and the model aging process, were examined against performance indicators. This included the prediction accuracy and false positive/negative rates. Each of these attributes was thoroughly investigated and the results were discussed in the Experiment Results section.

8.1 Contribution

There are a number of interesting works related to data protection of low powered networks, discussed in the Related Work section. However, most of the proposed work can be classified into two primary groups, based on the deployment scenarios. They are:

1. Models based on implementing a middleware solution to low powered nodes
2. Models based on modifying communication protocols

However, the software and hardware architecture of low powered nodes are highly diverse and resources are restricted, thus designing universal software to operate in diverse software/hardware architectures and resource constrained environments would be a challenge. Furthermore, implementing a middleware to be used in resource-constrained environments including memory, storage and processing power is also a challenge. Furthermore, a majority of low powered networks operate in open standards such as IEEE 802.15.4 and all amendments and modifications to communication protocols are controlled by IEEE and implementing changes to a communication protocol would be a complex and an exhausting process.

Unfortunately, models based on both the above groups are unable to provide a usable solution unless low powered nodes or existing communication protocols are modified. Consequently, those models are unable to provide any protection to over 20 billion low powered devices already deployed, operating in heterogeneous software and hardware. However, the model proposed in this work is able to protect the already deployed, billions of low powered devices, by detecting traffic based anomalies. The model proposed in this work can be implemented without modifying communication protocols, operating systems or the addition of any software to low powered nodes.

Contrary to work of Xiao et al. [102], this work is based on a number of attributes belonging to four primary groups (IEEE 802.15.4e/TSCH physical layer (wireless), low powered characteristics and network layer). More than ten attributes of low powered networks, including timestamps, absolute slot number, received signal strength indicator, (RSSI), signal to noise ratio (SNR), link quality indicator (LQI), link distance, battery level, source/destination address, packet payload size and service types are investigated in this work. Furthermore, this work is further enhanced by examining performance indicators, including prediction accuracy and false positive/negative rates for various properties, such as machine learning algorithms, training sample sizes, number of nodes, retention factors, noise influences, seasonal effects, network clusters and model aggregation effect. More than 80 independent experiments are performed in this work to provide a detailed overview of anomaly detection models on low powered attributes. Furthermore, data collected from an operational network with over 100 nodes, during a six month period was used to achieve more realistic results.

8.2 Limitations

Even though great effort was put into this work, a number of limitations were found. In the following, those limitations are discussed.

- **Simulated data**

  A simulation was utilized to generate sample data used in some of the experiments. Furthermore, the simulation software version used in this work (OpenWSN) had its own limitations which included the maximum number of nodes (restricted to five) and the capability of assigning link quality parameters. Node limitation caused by a software glitch in the routing protocol (RPL) implementation of OpenWSN resulted in simulation software, which was unable to generate a sample set based on a larger number of low powered nodes. Furthermore, the OpenWSN version used in this work was unable to adjust the link quality parameters for the wireless links. As a consequence, the simulation software was unable to provide noisy, unstable, wireless environments to generate a more realistic sample set. Furthermore, since data from two different distributions were utilized in this work, the results did not demonstrate the true effect of an aggregated detection model utilizing all low powered attributes. (The aggregated models used in this thesis are based on the data collected from a single distribution and physical (wireless) layer attributes are not utilized in developing the aggregated models. However, a potential aggregated-sequence model is proposed as a future work and such a model requires data collected from a single distribution representing all four groups (protocol specific, wireless layer, low powered and network layer attributes))

- **Anomaly data**

  Machine learning based anomaly detection models rely on both anomalous and normal data to build prediction models so that the accuracy and the generalization factor are higher. However, both simulated and live data were unable to produce an adequate amount of anomalous data. Therefore, other techniques such as noise injection and sequence permutation technique were also used to generate anomalous data. Although most prediction models evaluated in this thesis were able to withstand to higher noise threshold values positively, such methods may not be able to accurately represent the complete map of anomalous behaviour of a given network. To address these concerns, node-based anomalies are generated by modifying the software of simulated nodes. However, still, such an approach might not represent the complete range of potential, anomalous behaviour of a given network.

- **Mobility**

  Data used in this work have been collected from wireless networks with static topologies. Each node is stationed in a relatively, fixed location and a negligible, distance variance between nodes has occurred by the accuracy threshold of the distance measuring mechanism. However, in real-world applications, things such as wearable sensors are highly mobile and the active location of such a device may be highly unpredictable. However, some low powered wireless devices may produce a deterministic motion, which can be learned by using different techniques such as machine learning. In both scenarios, the complexity of the LoWSN increases and composite mechanisms may be required to predict the behaviour of such a complex architecture. It is a challenge to implement this in a simulated environment and wireless devices, operating in IEEE 802.15.4e/TSCH mode, are required for a comprehensive analysis to identify the influence of mobility on the prediction accuracy.

- **Realistic Influence (noise)**

  The wireless communication, utilizing open protocols and unlicensed frequencies are substantially impacted by several external factors including interference, obstruction, multipath fading, network topology and the behaviour of peer nodes. However, the influence of each of these factors is environment specific and finding a generalized model to approximate the external influence on each wireless node may be a complex process. Furthermore, the external influence on individual nodes may be time-specific and the artificial induction of noise may not provide a realistic effect. As the mobility of nodes, external influence mapping processes would be implementation-specific.

- **Seasonal Effects**

  Seasonal influences are a major factor when the LoWSNs are deployed in outdoor environments. It is critical for the prediction model to be trained with data representing the most diverse dataset. However, non-simulated data, used in this work, was collected during a six months period and it might not have produced the complete map of data distribution. Although previous experiments confirmed a satisfactory resistance, against random noise used with simulated data, it may not have been able to represent time-sensitive and prolonged fluctuations, accurately. The following diagram is generated, using 180000 samples, collected during a six month period; this compared the fluctuation of signal strength attribute (RSSI) during the testing period. The continuous lines describe the min, max and mean values for all data sets collected during the test period. However, the dotted lines indicate min, max and mean values for a particular week.

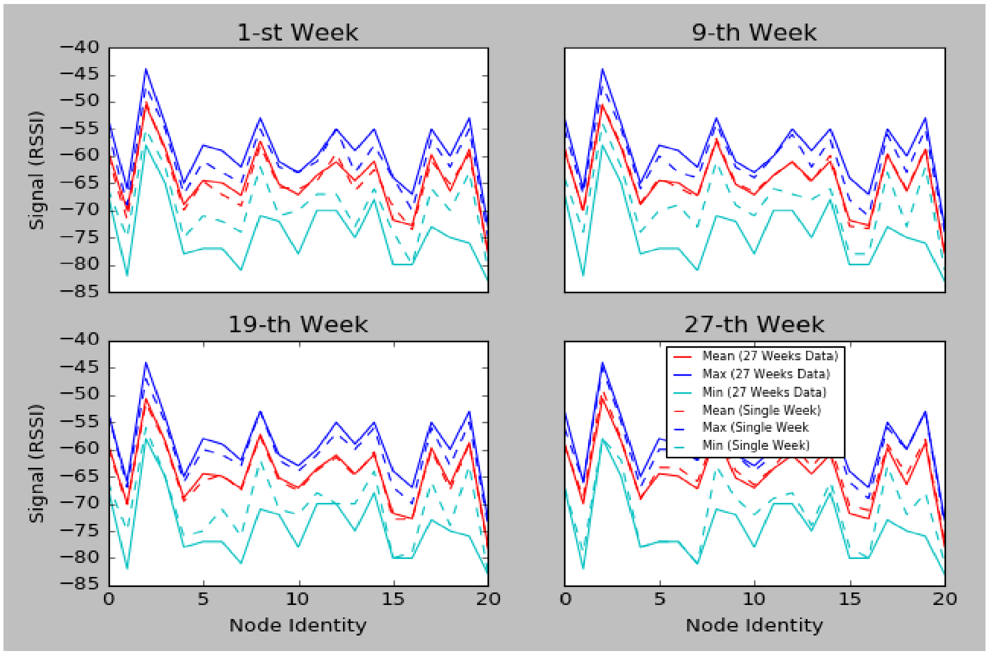


Fig. 116. RSSI variance comparison (weekly)

The above diagram demonstrates the fluctuation of signal strength during the test period. However, statistical data indicate a negligible variance of the average mean (red line) between the weekly data and the total data set (six months).

The following diagram describes the variance of statistical properties (min, max, mean and standard deviation) of data collected from a 20-node, wireless segment during a six month period.

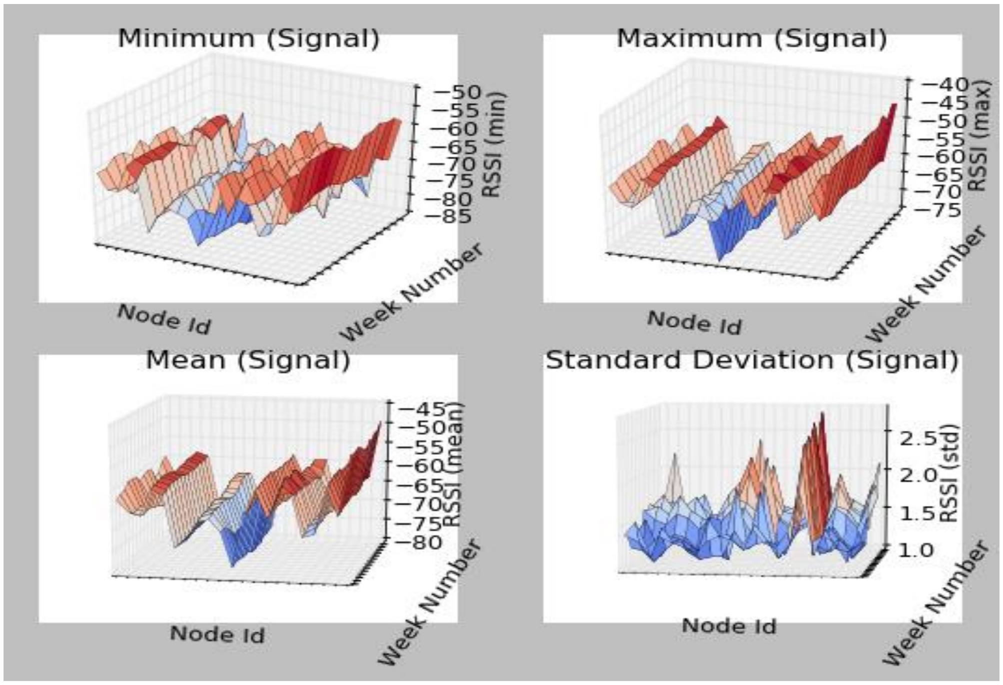


Fig. 117. Descriptive statistics: RSSI vs week

According to the above diagram, the mean and the standard deviation values for each node confirm a marginal variance of signal strength (wireless property) during the test period (six months). However, data were collected during the months of May and October and the variance caused by some seasonal effects, such as severe winter weather, may not be properly represented in the data set.

As previously mentioned, data collected from several wireless sections with diverse environmental characteristics were used in experiments where the input features were the physical layer and wireless attributes. The following diagram demonstrates the variance of the input feature (signal strength/RSSI) used in those experiments. The following result is generated, using 180000 samples, collected from four separate wireless segments during a six months period.

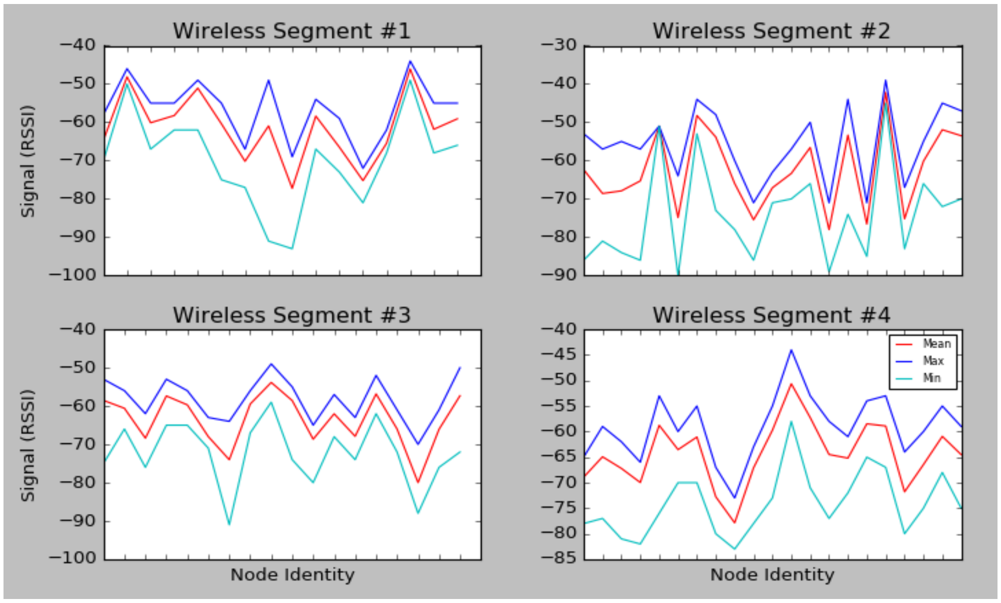


Fig. 118. Comparison of RSSI variance in different segments

The above diagram doesn’t demonstrate a distinctive variance of the RSSI between segments. However, in each segment, some nodes maintained a lower variance during the test period and different factors including interference, physical location and installation type could contribute to this lower variance. The following diagram depicts the variance (standard deviation) of signal strength (RSSI) of each node in four separate wireless segments during the six months, testing period.

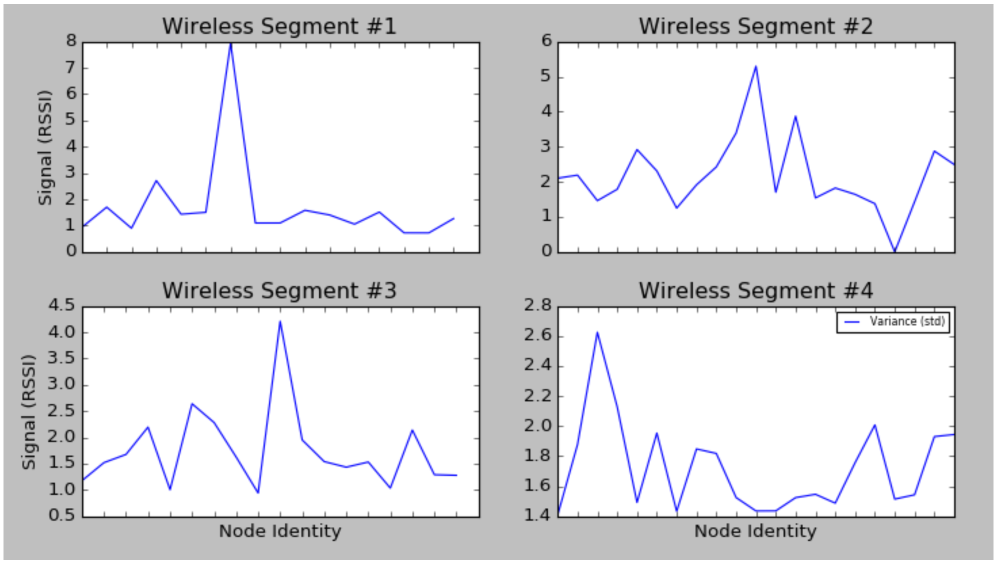


Fig. 119. The standard deviation of RSSI in different wireless segments

The above diagram demonstrates a higher RSSI variance for certain nodes. However, a significant contrast of signal fluctuation between different segments is not visible.

8.3 Comparison between experimental groups

In this section, the results including prediction accuracy and false positive/negative rates, obtained in different experimental groups, are compared. For the comparison, two factors, including training set size and the classification algorithm are used. The following notation is used in the comparison reports.

Group 1	IEEE 802.15.4e/TSCH properties
Group 2	Wireless (physical layer) properties
Group 3	Low powered properties
Group 4	Network properties
SVM	Support Vector Machines
KNN	K-Nearest Neighbours
NN	Neural Networks

DT	Decision Tree
RF	Random Forest (ensemble)

Tab. 25. The notation used in the following discussion

- **Training set size (classification algorithm: DT)**

	Group 1	Group 2	Group 3	Group 4
100	60.0	40.0	50.0	76.47
200	82.75	77.5	100.0	67.64
500	90.27	78.0	80.0	80.0
1k	86.98	80.0	100.0	74.11
2k	92.80	81.5	94.7	76.92
5k	96.29	82.6	97.85	77.35
10k	97.25	82.75	97.32	78.04
20k	97.42	84.27	99.46	77.04

Tab. 26. Prediction accuracy comparison between experimental groups and training set size

The following diagram provides a different perspective of the above data.

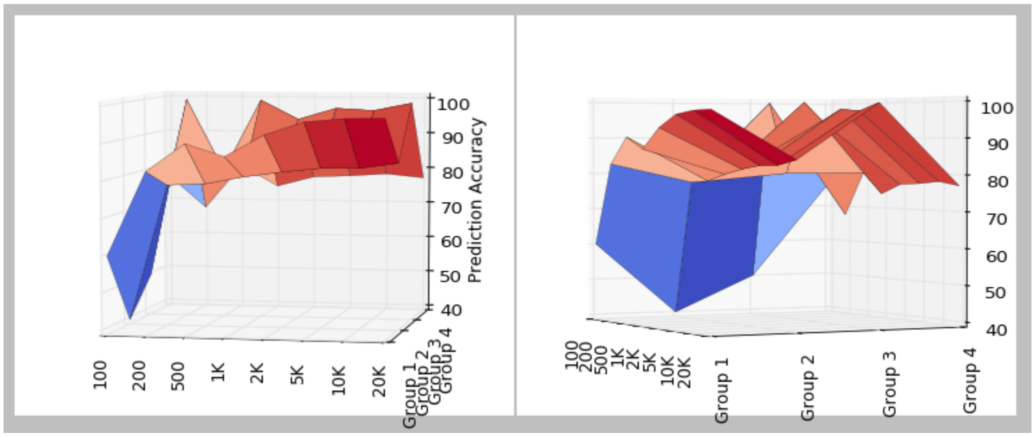


Fig. 120. Prediction accuracy comparison between experimental groups and training set size

The above diagram confirms that, regardless of the experimental group, models trained with larger data sets are able to produce higher accuracy. Furthermore, the prediction accuracy of models, based on experimental group 1 (IEEE 802.15.4e/TSCH attributes) and group 3 (low powered attributes), were substantially, higher than the accuracy of the other experimental groups.

- **Classification algorithm (training set size = 10,000)**

	Group 1	Group 2	Group 3	Group 4
SVM	82.41	83.05	74.19	85.28
KNN	96.08	83.60	95.69	85.69
NN	89.97	83.25	97.84	85.46
DT	97.25	82.75	100.0	78.04
RF (ensemble)	97.45	84.35	97.85	77.33

Tab. 27. Prediction accuracy vs [classification algorithm, experimental group]

The following diagram summarizes the graphical point of view of the above findings.

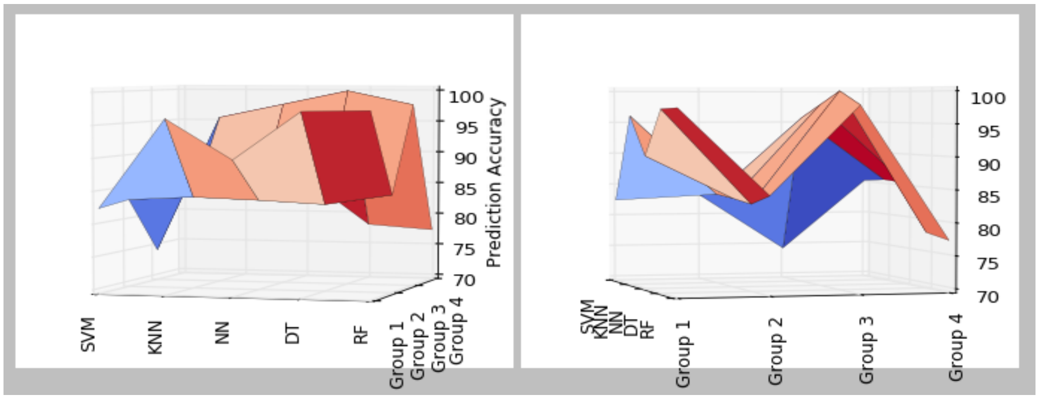


Fig. 121. Prediction accuracy vs [classification algorithm, experimental group]

The above diagram confirms models based on the DT algorithm were able to produce higher accuracy, regardless of the input feature vector used in experiments. However, similar to the previous conclusion, experiment group 1 and group 3 were able to produce substantially, higher accuracy.

8.4 Use Cases (Ultra-Low Powered Sensor Networks)

Unprecedented growth in low powered devices has been witnessed recently, and over eight billion, active, low powered devices were deployed as of 2017; this accounts for over a 30 percent growth compared with the previous year [112]. As previously noted, industry experts estimate there will be over 20 billion, active, low powered devices by 2020. Applications in several fronts have contributed to this enormous growth. Some of these are:

- Asset management
- Process automation
- Health care
- Security
- Predictive analytics
- Data Intelligence and strategic planning
- Product-based to service-based transformation

Each of these operational contexts are characterized by different objectives and associated data attributes that require diverse security demand. In the following, several applications of ultra-low powered wireless networks are discussed, briefly.

- **Agriculture Industry**

  Cattle farmers perform regular check up on cows to maintain a healthy diet and such a procedure could be expensive and tedious, especially if a particular farm consists of thousands of animals. If an ultra-low powered sensor device is implanted into animal's body, sensor data can be collected regularly for prolong periods (years) without replacing the battery.

- **Health Care**

  Elite athletes such as Olympic sprinters are going through a complex and rigorous preparation process. Although regular screening and adequate diet are common among athletes, real-time monitoring of critical physiological properties such as muscle movements, blood pressure or a reaction to a particular diet may provide an advantage over regular screening processes. Implantable or swallow-able ultra-low powered devices with appropriate sensors would be able to monitor the human body for longer periods and such a mechanism can be used on any patient who requires continuous health screening.

- **Emergency Response**

  Use of cutting edge technologies is critical in emergency response to prevent catastrophes and fatalities. Various ultra-low powered sensor devices can be used as part of rapid deployment process during an emergency including, to locate humans and animals, detect water and air contaminants, identify environmental properties such as temperature in a fire-affected area or to measure the wind and water flow speed to estimate the affected perimeter.

- **Manufacturing**

  Various sensors are used to measure different physical properties of mechanical instruments such as heavy machinery. However, in most cases those sensors are attached to the outer cell of the corresponding instrument. Complex multiple-layer instruments can take advantage of ultra-low powered wireless sensors to measure physical properties of embedded components, otherwise with no direct access.

- **Animal Tracking**

  Endangered animals living in inaccessible areas such as deep jungles and Arctic regions die due to various reasons, including lack of food or water. However, if it is possible to collect regular updates about their health as well as environmental attributes, proactive steps can be taken to prevent fatalities. Ultra-low powered wireless sensors operating in IEEE 802.15.4e/TSCH mode can be deployed to build a mesh network to connect ultra-low powered devices implanted in animals as well as low powered wireless devices installed in the surrounding environment.

- **Food Industry**

  Various mobile applications (Apps) are available to get the detailed information including nutrition facts such as a number of calories, sugar, protein and fibre level of a particular food product by using associated bar-code or manufacturer part number. The corresponding application connects to the manufacturer’s database or to a third party location to collect the necessary information. However, such information may not include the real-time status of a particular product. For instance, vegetables, dairy products, canned food can be easily contaminated or inconsumable due to various reasons. If a grocery store is able to provide a sample of a particular product by using ultra-low powered embedded technology, the consumer may be able to get real-time data about a particular food item.

8.5 Deployment scenario (topologies)

Several considerations must be taken into account when implementing a security mechanism based on the findings of this work. The data collection procedure could be dependent on several factors including security and environmental restrictions. Since IEEE 802.15.4e doesn’t outline the deployment requirements and procedures, the design requirements should be implementation, specific. Different IEEE 802.15.4e modes have different topological restrictions and low powered wireless networks, operating in TSCH mode, can support both mesh and star topology. However, transmission restrictions of low powered nodes often take advantage of partially, meshed topology to exchange data. Consequently, a number of deployment options are available including the use of a PAN coordinator with unrestricted resources to manage the network. The majority of non-management data is accessed by the PAN coordinator and it can be easily captured using a PAN coordinator-system specific tool. Several deployment options are available for the capturing mechanism and the following diagram depicts a few potential choices:

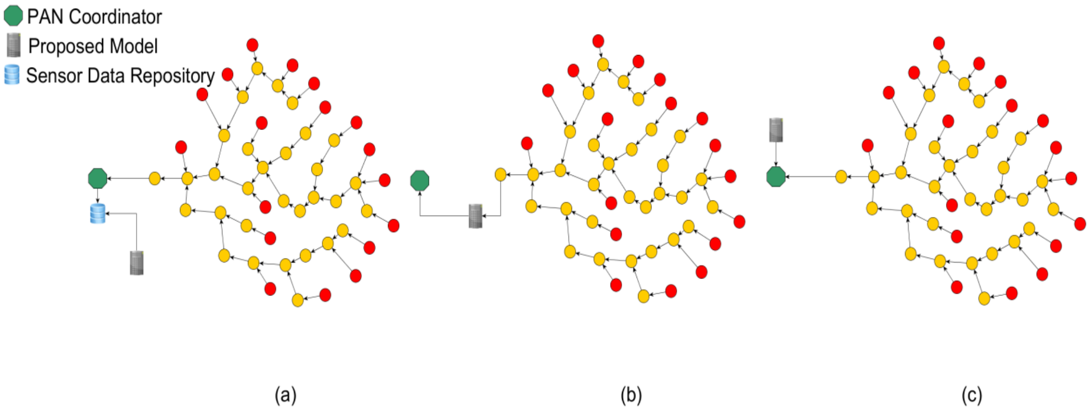


Fig. 122. Potential data capturing scenarios

Three possible data collection scenarios are demonstrated in the above diagram. In the first option (a), data is collected from a central data repository. The data repository could be integrated into the PAN coordinator or operate as a separate entity, such as an external hard-drive, backup drive, cloud storage or SAN (Storage Area Network). In most cases, repositories are configured to store, only, sensor-generated data. However, it could require a modified, data collection, mechanism to capture network control and management data such as enhanced beacons (EB).

In the second scenario (b), wireless data is collected from a capturing mechanism directly attached to the corresponding low powered wireless network. The capturing device should include a wireless interface card, supporting IEEE 802.15.4e/TSCH mode, and should be able to intercept the communication between nodes. The capturing mechanism must be positioned, in such a way, as to allow the most remote, low powered nodes to be listened to, passively. If a single capturing device is unable to capture adequate data, multiple capturing devices may be required.

With the third option (c), the capturing mechanism and the data storage facility is included in the PAN coordinator. If the PAN coordinator is operating in a proprietary system, the application programming interface (API) or another mechanism should be provided to capture wireless data. However, if the PAN coordinator is operating in an open system, a number of tools are available to capture wireless data directly, from a wireless interface card, operating in IEEE 802.15.4e/TSCH, promiscuous mode.

Chapter 9. Conclusion

In this work, low powered wireless network characteristics and supporting protocols were extensively, investigated to identify strong features, which can be used to build a model defining the operational behaviour of low powered wireless networks. Both simulated and real data, collected from wireless networks operating in open standards, was utilized with machine learning methods to learn the behaviour of low powered networks. Several common factors, such as the number of nodes, training set size, classification algorithms, model aging process and the noise impact were examined to identify the impact on performance including prediction accuracy and false positive/negative rates. A comprehensive analysis identified a number of potential prediction models to detect traffic anomalies in low powered environments. As previously noted, due to the restricted accessibility of wireless modules, operating in IEEE 802.15.4e/TSCH mode, the simulated environment was used to collect data for representing 802.15.4e/TSCH properties, including ASN (Absolute Slot Number). Since wireless attributes in the physical layer demonstrate similar characteristics in different open, wireless protocols, data collected from an operational wireless network, with a few hundred nodes, were utilized in experiments related to input features based on physical layer (wireless) attributes.

Future Work

Model aggregation is briefly discussed and the experimental result confirmed a significant gain in performance including prediction accuracy and false positive/negative rates. However, model bundling (aggregation) techniques rely on single-distribution data sets. As previously mentioned, due to restrictions in hardware availability, data extracted from multiple distributions were used in experiments and the full potential of the model aggregation process was not properly investigated. A potential aggregation mechanism could be used in future work and it is described using the following diagram.

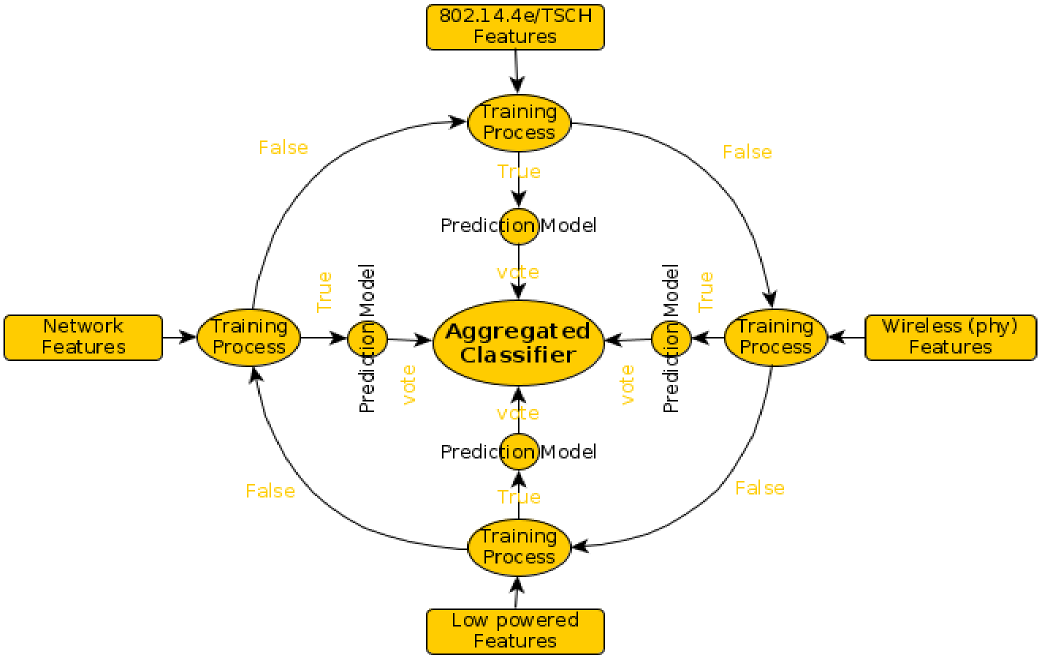


Fig. 123. Peer dependent aggregated prediction model

In the above proposal, each model is trained using a subset of features belonging to the individual, feature-group (IEEE 802.15.4e/TSCH, physical (wireless), low powered and network). However, each unsuccessful prediction could be re-trained with the next prediction model, until, finding a model with a successful prediction (next prediction model selection could be sequential or random). Consequently, all training data need to be successfully predicted by at-least a single model and an aggregated classification model could be obtained by using a simple, voting mechanism as discussed previously. To evaluate such a model, data extracted from a single distribution is compulsory and modification to the existing machine learning libraries may be required.

The above model can be further enhanced by identifying the influence of each feature-group in the classification process. In essence, instead of using a simple voting technique to determine the output class for the corresponding input data, the prediction of each model can be used as input data to train a new prediction model. The following diagram is used to demonstrate the potential model.

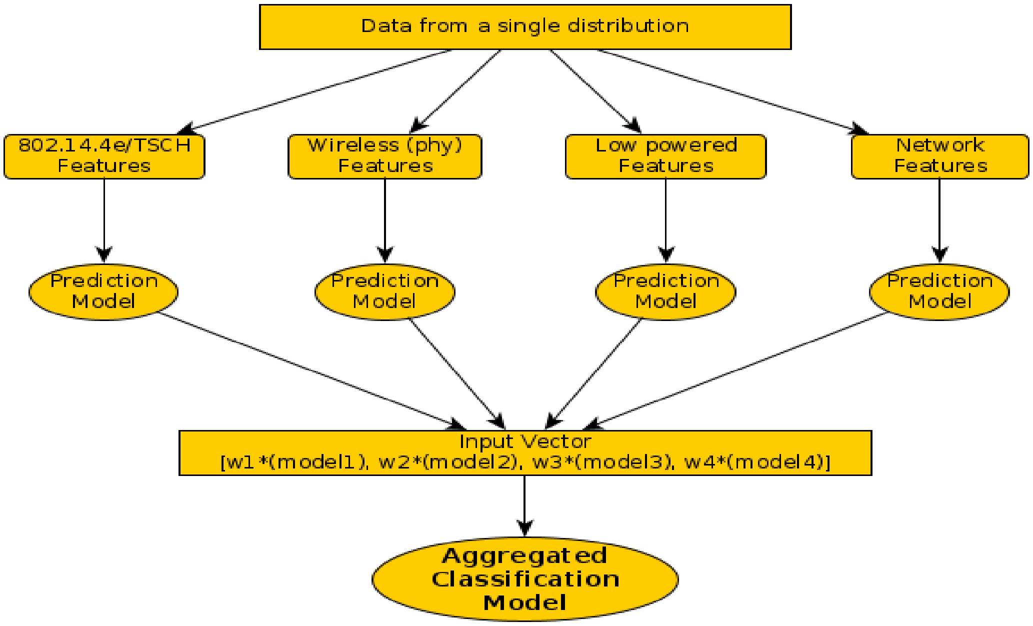


Fig. 124. A framework to enhance performance through a re-training mechanism

In the aforementioned model, four, separate prediction models (primary) are trained using attributes belonging to each feature-group. Subsequently, the secondary prediction model is trained using the outcome of four primary prediction models as an input. The advantage of this approach over a simple voting mechanism is that certain primary prediction models can enforce a higher influence on final prediction. For instance, if features belonging to a physical layer (wireless) feature-group have a stronger influence on the final prediction, a vote based mechanism is unable to interpret the influence; since the simple voting mechanism is based on the majority vote count. However, by using the outcome of primary models as the input vector for the secondary model, more complex relationships can be found.

Final words

The growth of low powered wireless devices including the Internet of Things (IoT) has been unprecedented in recent years and experts estimate, by 2020, 25 percent of identified attacks on enterprise environments will be related to low powered devices [113]. Furthermore, experts predict, by 2022, 50 percent of the IoT security budget will be allocated to fix existing faults in low powered devices [111]. The magnitude of potential security concerns in low powered devices is not fully realized by stakeholders and this is mainly due to a lack of regulatory control. However, with the predicted 20 billion, low powered devices, deployed in heterogeneous systems by 2020, security enforcement on these individual devices may not be feasible. Centralized, security mechanisms such as anomaly detection and perimeter security controls may be the most favourable approach to secure existing, low powered infrastructures. However, anomaly detection mechanisms can be used as a supplementary tool with other security controls such as access control, source node authentication and network health monitoring systems.

Experiments performed in this work confirmed over 95 percent anomaly detection accuracy can be achieved by carefully selecting input features and classification models. By evaluating machine learning techniques with low powered input features, this work demonstrated a further improvement of performance can be achieved by segregating input features based on common characteristics and aggregating multiple prediction models. Finally, a framework for a classification model, based on a sequential (or random) peer dependent learning mechanism to further improve the performance is proposed. However, to evaluate such a model, the data from a single distribution may be required.

Even though, the objective of this work is to compare the performance of different classification models against various factors, it is important to summarize some of the critical findings of this thesis. With four basic classification algorithms used in this thesis, models based on Decision Tree were able to perform significantly better with protocol specific and low powered attributes. However, with Physical layer attributes, NN based models were able to perform marginally better with different size of training sets. With the Network layer attributes, both KNN and DT based models have produced similar results. However, in overall, models based on DT performed better than other three basic classifiers and DT based models also consumed marginally low resources during the training process. As far as input features are concerned, performance of models based on low powered attributes and protocol specific attributes outperformed the rest of the categories. However, models based on battery usage as input parameter were able to classify traffic anomalies significantly better than rest of the input features evaluated in this thesis. In conclusion, a model based on Decision Tree using battery usage as input feature were able to outperform other models in detecting traffic anomalies in low powered wireless sensor networks while keeping the resource utilization low.

Chapter 10. Reference

[1] R. Daidone, G. Dini and M. Tiloca, On experimentally evaluating the impact of security on IEEE 802.15.4 networks, Proceedings of the 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS 2011), pp. 20-25, 2011

[2] D. De Guglielmo, A. Seghetti, G. Anastasi, and M. Conti, A performance analysis of the network formation process in IEEE 802.15.4e TSCH wireless sensor/actuator networks, in Proceedings of 2014 IEEE Symposium on Computers and Communication (ISCC), pp. 23-26, 2014

[3] M. Lemmon, Q. Ling, Y. Sun, Overload management in sensor actuator networks used for spatially distributed control systems. In Proceedings of the 1st international conference on Embedded networked sensor systems, ACM, 2003

[4] J. Xu, G. Yang, Z. Chen, Q. Wang, A survey on the privacy-preserving data aggregation in wireless sensor networks, China Communications, vol. 12/5, pp. 162 – 180, 2015

[5] A. Ahmad, A. Riedl, W. Naramore, N. Chou, M. Alley, Comparative study of security in IEEE 802.11-2007 and IEEE 802.15.4-2006 for patient monitoring environments, 2010 Seventh International Conference on Information Technology: New Generations, pp. 914 - 918, 2010

[6] A. F. Skarmeta, J. L. Hernández-Ramos, M. V. Moreno, A decentralized approach for security and privacy challenges in the Internet of things, In Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-IoT), pp. 67 - 72, 2014

[7] Y. Xiao, C. Bandela, Y. Pan, Vulnerabilities and security enhancements for the IEEE 802.11 WLANs, GLOBECOM '05. IEEE Global Telecommunications Conference, vol. 3, pp. 5, 2005

[8] IEEE std. 802.15.4, Part. 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs), IEEE standard for Information Technology, 2006

[9] M. J. Covington, R. Carskadden, Threat implications of the Internet of things, 2013 5th International Conference on Cyber Conflict (CYCON 2013), pp. 1 – 12, 2013

[10] IEEE std. 802.15.4e, Part. 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC sublayer, IEEE standard for Information Technology, 2012

[11] M. Henze, L. Hermerschmidt, D. Kerpen, R. Häußling, B. Rumpe, K. Wehrle, User-driven privacy enforcement for cloud-based services in the Internet of things, 2014 International Conference on Future Internet of Things and Cloud, pp. 191 - 196, 2014

[12] A. Gaware, S. B. Dhonde A survey on security attacks in wireless sensor networks, 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 536 – 539, 2016

[13] J. Zhou, A. E. Xhafa, R. Vedantham, R. Nuzzaci, A. Kandhalu, X. Lu, Comparison of IEEE 802.15.4e MAC features 2014 IEEE World Forum on Internet of Things (WF-IoT), pp. 203 – 207, 2014

[14] E. Vogli, G. Ribezzo, L. A. Grieco, G. Boggia, Fast join and synchronization schema in the IEEE 802.15.4e MAC, 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), pp. 85 – 90, 2015

[15] O. Song, J. Kim, An efficient design of security accelerator for IEEE 802.15.4 wireless sensor networks, 2010 7th IEEE Consumer Communications and Networking Conference, pp. 1 – 5, 2010

[16] S. M. Sajjad, M. Yousaf, Security analysis of IEEE 802.15.4 MAC in the context of Internet of things, 2014 Conference on Information Assurance and Cyber Security (CIACS), pp. 9 – 14, 2014

[17] M. R. Palattella, N. Accettura, M. Dohler, L. A. Grieco, G. Boggia, Traffic aware scheduling algorithm for reliable low-power multi-hop IEEE 802.15.4e networks, 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC), pp. 327 – 332, 2012

[18] K. Pister and L. Doherty, TSMP: Time Synchronized Mesh Protocol (TSMP), in Proceedings of International Symposium of Distributed Sensor Networks (DSN), Florida, USA, Nov. 2008

[19] HART communication protocol and foundation, Retrieved from https://fieldcommgroup.org/technologies/hart (Dec-08-2019)

[20] C. Shih, A. E. Xhafa, J. Zhou, Practical frequency hopping sequence design for interference avoidance in 802.15.4e TSCH networks, 2015 IEEE International Conference on Communications (ICC), pp. 6494 – 6499, 2015

[21] IEEE std. 802.11-2012 - IEEE standard for information technology--Telecommunications and information exchange between systems local and metropolitan area networks--specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2012

[22] C. Kolias, G. Kambourakis, A. Stavrou, S. Gritzalis, Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset, IEEE Communications Surveys & Tutorials, vol. 18/1, pp. 184 – 208, 2016

[23] Y. Xiao, C. Bandela, Y. Pan, Vulnerabilities and security enhancements for the IEEE 802.11 WLANs, GLOBECOM '05. IEEE Global Telecommunications Conference 2005, vol. 3, pp. 5, 2005

[24] C. He, J. C. Mitchell, Security analysis and improvements for IEEE 802.11i, in proceedings of 12th Annual NDSS Symposium, Stanford, CA, USA, pp. 90–110, 2005

[25] L. Wang, B. Srinivasan, Analysis and improvements over DoS attacks against IEEE 802.11i standard, 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing, vol. 2, pp. 109 – 113, 2010

[26] K. Pelechrinis, M. Iliofotou, S. V. Krishnamurthy, Denial of service attacks in wireless networks: The case of jammers, IEEE Communications Surveys & Tutorials, vol. 13/2, pp. 245 – 257, 2011

[27] S. Fluhrer, I. Mantin, and A. Shamir, Weaknesses in the key scheduling algorithm of RC4, in Proceeding of Selected Areas Cryptography, pp. 1–24. 2001

[28] A. Bittau, Additional weak IV classes for the FMS attack, Department of Computer Science, University College London, London, U.K., 2003

[29] H. Berghel, J. Uecker, WiFi attack vectors, communication. ACM, volume 48/8, pp. 21–28, 2005

[30] R. Chaabouni, Break WEP faster with statistical analysis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Tech. Rep., 2006

[31] E. Tews, R. P. Weinmann, A. Pyshkin, Breaking 104 bit WEP in less than 60 seconds, in proceeding of 8th International Conference of Information Security Application, pp. 188–202, 2007

[32] Y. Liu, Z. Jin, and Y. Wang, Survey on security scheme and attacking methods of wpa/wpa2, in Proceeding of 6th International Conference of WiCOM, pp. 1–4, 2010

[33] K. Bicakci, B. Tavli, Denial-of-service attacks and countermeasures in IEEE 802.11 wireless networks, Computer Standards & Interfaces, vol. 31/5, pp. 931–941, 2009

[34] F. Ferreri, M. Bernaschi, L. Valcamonici, Access points vulnerabilities to DoS attacks in 802.11 networks, 2004 IEEE Wireless Communications and Networking Conference, vol. 1, pp. 634 – 638, 2004

[35] R. Beyah, A. Venkataraman, Rogue-access-point detection: challenges, solutions, future directions, IEEE Security Privacy, vol. 9/5, pp. 56–61, 2011

[36] L. Wang, B. Srinivasan, Analysis and improvements over DoS attacks against IEEE 802.11i standard, in proceedings of 2nd International Conference of NSWCTC, vol. 2, pp. 109–113, 2010

[37] A. Tsakountakis, G. Kambourakis, S. Gritzalis, Towards effective wireless intrusion detection in IEEE 802.11i, in proceedings of 3rd International Workshop SECPerU, pp. 37–42, 2007

[38] C. He, J. C. Mitchell, Analysis of the 802.11i 4-Way handshake, in proceedings of the 3rd ACM workshop on Wireless security, Philadelphia, PA, USA, pp. 43 – 50, 2004

[39] B. Aslam, M. H. Islam, S. A. Khan, Pseudo randomized sequence number based solution to 802.11 disassociation denial of service attack, in Proceedings of the First Mobile Computing and Wireless Communication International Conference, Amman, pp. 215 – 220, 2006

[40] Z. Afzal, J. Rossebø, B. Talha, M. Chowdhury, A wireless intrusion detection system for 802.11 networks, 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 828 – 834, 2016

[41] R. d. Carmo, M. Hollick, DogoIDS: A mobile and active intrusion detection system for IEEE 802.11s wireless mesh networks. In Proceedings of the 2nd ACM workshop on Hot topics on wireless network security and privacy, pp. 13–18, 2013

[42] F. Hugelshofer, P. Smith, D. Hutchison, N. J. P. Race, Openlids: A lightweight intrusion detection system for wireless mesh networks. In Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, pp. 309–320, 2009

[43] G. Vigna, S. Gwalani, K. Srinivasan, E. M. Belding-Royer, R. A. Kemmerer, An intrusion detection tool for aodv-based ad hoc wireless networks. In Proceedings of the 20th Computer Security Applications Conference, pp. 16–27, 2004

[44] U. Premaratne, U. Premarathne, K. Samarasinghe, Intrusion detection for IEEE 802.11 based industrial automation using possibilistic anomaly detection, 2010 Seventh International Conference on Wireless and Optical Communications Networks - (WOCN), pp. 1 – 5, 2010

[45] H. Alipour, Y. B. Al-Nashif, P. Satam, S. Hariri, Wireless anomaly detection based on IEEE 802.11 behavior analysis, IEEE Transactions on Information Forensics and Security, vol. 10/10, pp. 2158 – 2170, 2015

[46] M. Agarwal, S. Biswas, S. Nandi, Detection of de-authentication denial of service attack in 802.11 networks, 2013 Annual IEEE India Conference (INDICON), pp. 1 – 6, 2013

[47] M. Agarwal, S. Biswas, S. Nandi, Detection of de-authentication DoS attacks in Wi-Fi Networks: A machine learning approach, 2015 IEEE International Conference on Systems, Man, and Cybernetics, pp. 246 – 251, 2015

[48] C. Panos, P. Kotzias, C. Xenakis, I. Stavrakakis, Securing the 802.11 MAC in MANETs: A specification-based intrusion detection engine, 2012 9th Annual Conference on Wireless On-Demand Network Systems and Services (WONS), pp. 16 – 22, 2012

[49] R. Gass, J. Scott, and C. Diot, Measurements of in-motion 802.11 networking, in Proceedings of 7th IEEE WMCSA, pp. 69–74. 2005

[50] J. V. Kumar, A. Jain, P. N. Barwal, Wireless sensor networks: security issues, challenges and solutions, International Journal of Information and Computation Technology (IJICT), vol. 4/8, pp. 859–868, 2014

[51] Yasmin M. Amin, Amr T. Abdel-Hamid, Classification and analysis of IEEE 802.15.4 PHY layer attacks, 2016 International Conference on Selected Topics in Mobile & Wireless Networking (MoWNeT), pp. 1 – 8, 2016

[52] P. Jokar, H. Nicanfar, V. C. M. Leung, Specification-based intrusion detection for home area networks in smart grids, IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 208-213, 2011

[53] C. P. O'Flynn, Message denial and alteration on IEEE 802.15.4 lowpower radio networks, 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp. 1-5, 2011

[54] R. Sokullu, I. Korkmaz, O. Dagdeviren, A. Mitseva, N. R. Prasad, An investigation on IEEE 802.15.4 MAC layer attacks. In Proceedings of the 10th International Symposium on Wireless Personal Multimedia Communications (WPMC) 2007 (pp. 1019-1023)

[55] A. D. Wood, J. A. Stankovic, G. Zhou, DEEJAM: Defeating energy-efficient jamming in IEEE 802.15.4-based wireless networks, 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), pp. 60-69, 2007

[56] M. Wilhelm, J. B. Schmitt, V. Lenders, Practical message manipulation attacks in IEEE 802.15.4 wireless networks

[57] R. Sokullu, O. Dagdeviren, and I. Korkmaz, On the IEEE 802.15.4 MAC layer attacks: GTS attack, IEEE Second International Conference on Sensor Technologies and Applications (SENSORCOMM '08), pp. 673-678, 2008

[58] V. B. Misic, J. Fung, and J. Misic, MAC layer security of 802.15.4-compliant networks, IEEE International Conference on Mobile Adhoc and Sensor Systems, 2005

[59] S. S. Jung, M. Valero, A. Bourgeois, R. Beyah, Attacking beacon-enabled 802.15.4 networks, security and privacy in communication networks, lecture notes of

the institute for computer sciences, social informatics and telecommunications engineering, vol. 50, pp. 253-271, 2010

[60] H. Shuijing, Big data analytics: key technologies and challenges, 2016 International Conference on Robots & Intelligent System (ICRIS), pp. 141 – 145, 2016

[61] R. Memisevic, Deep learning: Architectures, algorithms, applications, 2015 IEEE Hot Chips 27 Symposium (HCS), pp. 1 – 12, 2015

[62] R. Deebalakshmi, V. L. Jyothi, A survey of classification algorithms for network traffic, 2016 Second International Conference on Science Technology Engineering and Management (ICONSTEM), pp. 151 – 156, 2016

[63] T. Thuy, T. Nguyen, G. Armitage, A survey of techniques for Internet traffic classification using machine learning, IEEE Communications Surveys & Tutorials, volume 10/4, pp. 56 – 76, 2008

[64] H. S. Hippert, C. E. Pedreira, R. C. Souza, Neural networks for short term load forecasting: A review and evaluation, IEEE Transaction on Power Systems, vol. 16, pp. 44-55, 2001

[65] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing, vol. 14/3, pp. 199-222, 2004

[66] A. Munther, A. Alalousi, S. Nizam, R. Othman, M. Anbar, Network traffic classification: A comparative study of two common decision tree methods: C4.5 and Random Forest, 2014 2nd International Conference on Electronic Design (ICED), pp. 210 – 214, 2014

[67] M. K. Alsmadi, K. B. Omar, S. A. Noah, I. Almarashdah, Performance comparison of multi-layer perceptron (back propagation, delta rule and perceptron) algorithms in neural networks in 2009 IEEE International Advance Computing Conference (IACC 2009), pp. 296-299, 2009

[68] S. Mahfouz, F. Mourad-Chehade, P. Honeine, J. Farah, H. Snoussi, Non-parametric and semi-parametric RSSI/distance modeling for target tracking in wireless sensor networks, IEEE Sensors Journal, vol. 16/7, pp. 2115 – 2126, 2016

[69] M. L. Das, Two-factor user authentication in wireless sensor networks, IEEE Transactions. Wireless Communication 2009, vol. 8, pp. 1086-1090, 2009

[70] M. K. Khan, and K. Alghathbar, Cryptanalysis and security improvements of 'Two-factor user authentication in wireless sensor networks, Sensors 2010, vol. 10/3, pp. 2450-2459, 2010

[71] B. Vaidya, D. Makrakis, H. T. Mouftah, Improved two factor user authentication in wireless sensor networks, Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on, pp. 600-606, 2010

[72] L. H. Freitas, K. A. Bispo, N.S. Rosa, P.R.F. Cunha, SM-Sens: Security middleware for wireless sensor networks, Proceedings of the Information Infrastructure Symposium, 2009.

[73] R. Daidone, G. Dini, M. Tiloca, STaR: A reconfigurable and transparent middleware for WSNs security, Proceedings of the 2nd International Conference on Sensor Networks (SENSORNETS 2013), 2013

[74] G. Piro, G. Boggia, L. A. Grieco, A standard compliant security framework for IEEE 802.15.4 networks, 2014 IEEE World Forum on Internet of Things (WF-IoT), pp. 27 – 30, 2014

[75] F. X. Standaert, G. Rouvroy, J. J. Quisquater, J. D. Legat, A methodology to implement block ciphers in reconfigurable hardware and its application to fast and compact AES RIJNDAEL, International Symposium on Field-Programmable Gate Arrays (FPGA 2003), Monterey, CA, 2003

[76] P. Chodowiec, P. Khuon, K. Gaj, Fast implementations of secret-key block ciphers using mixed inner- and outer-round pipelining, International Symposium on Field-Programmable Gate Arrays (FPGA 2001), Monterey, CA, 2001

[77] I. Verbauwhede, P. Schaumont, H. Kuo, Design and performance testing of a 2.29-GB/s rijndael processor, IEEE Journal of Solid-State Circuits, vol. 38/3, 2003

[78] T. F. Lin, C. P. Su, C. T. Huang, C. W. Wu, A high-throughput low-cost AES cipher chip, IEEE Asia-Pacific Conference on ASIC, 2002

[79] A. K. Lutz, J. Treichler, F. K. G¨urkaynak, H. Kaeslin, G. Basler, A. Erni, S. Reichmuth, P. Rommens, S. Oetiker, W. Fichtner, 2Gbit/s hardware realizations of RIJNDAEL and SERPENT: A comparative analysis, cryptographic hardware and embedded systems (CHES 2002), San Francisco Bay, CA, 2002

[80] P. Hamalainen, M. Hannikainen, T. D. Hamalainen, Efficient hardware implementation of security processing for IEEE 802.15.4 wireless networks 48th Midwest Symposium on Circuits and Systems, pp. 484 – 487, 2005

[81] T. Hao, Y. Jia, X. Tian, Research on the forecast model of security situation for information system based on Internet of things, 2013 International Conference on Anti-Counterfeiting, Security and Identification (ASID), pp. 1 – 5, 2013

[82] L. Marin, A. Jara, and A.F. Skarmeta, Shifting primes: Optimizing elliptic curve cryptography for smart things. In Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on, pp. 793–798. 2012.

[83] H. Zhang, T. Zhang, Short paper: A peer to peer security protocol for the Internet of things: secure communication for the sensible things platform, 2015 18th International Conference on Intelligence in Next Generation Networks, pp. 154 – 156, 2015

[84] C. Liu, Y. Zhang, H. Zhang, A novel approach to IoT security based on immunology, 2013 Ninth International Conference on Computational Intelligence and Security, pp. 771 – 775, 2013

[85] M. Xie, S. Han, B. Tian, S. Parvin, Anomaly detection in wireless sensor networks: A survey, Journal of Network and Computer Applications, Volume 34/4, pp. 1302-1325, 2011

[86] S. Rajasegarar, C. Leckie, M. Palaniswami, Anomaly detection in wireless sensor networks, IEEE Wireless Communications, vol. 15/4, pp. 34 – 40, 2008

[87] O. Can, O. K. Sahingoz, A survey of intrusion detection systems in wireless sensor networks, 2015 6th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO), pp. 1 – 6, 2015

[88] D. S. Shukla, A. C. Pandey, A. Kulhari, Outlier detection: A survey on techniques of WSNs involving event and error based outliers, 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH), pp. 113 – 116, 2014

[89] Y. Zhang, N. Meratnia, P. Havinga, Outlier detection techniques for wireless sensor networks: A survey, IEEE Communications Survey and Tutorials, vol. 12, No.2, 2010

[90] T. Palpanas, D. Papadopoulos, V. Kalogeraki, D. Gunopulos, Distributed deviation detection in sensor networks. SIGMOD Record 2003, pp. 77–82, 2003

[91] M. Tiwari, K. V. Arya, R. Choudhari, K. S. Choudhary, Designing intrusion detection to detect black hole and selective forwarding attack in WSN based on local information, 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, pp. 824 – 828, 2009

[92] S. Rajasegarar, C. Leckie, M. Palaniswami, J. C. Bezdek, Distributed anomaly detection in wireless sensor networks, 2006 10th IEEE Singapore International Conference on Communication Systems, pp. 1 – 5, 2006

[93] H. Wang, Z. Yuan, C. Wang, Intrusion detection for wireless sensor networks based on multi-agent and refined clustering, 2009 WRI International Conference on Communications and Mobile Computing, vol. 3, pp. 450 – 454, 2009

[94] A. Agah, S. K. Das, K. Basu, A non-cooperative game approach for intrusion detection in sensor networks, IEEE 60th Vehicular Technology Conference, 20. VTC2004-Fall., vol. 4, pp. 2902 – 2906, 2004

[95] A. P. R. da Silva, M. H. T. Martins, B. P. S. Rocha, A. A. F. Loureiro, L. B. Ruiz, H. C. Wong, Decentralized intrusion detection in wireless sensor networks, Proceeding Q2SWinet '05 Proceedings of the 1st ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks, pp. 16-23, 2005

[96] B. Yu, B. Xiao, Detecting selective forwarding attacks in wireless sensor networks, Proceedings 20th IEEE International Parallel & Distributed Processing Symposium, pp.8, 2006

[97] J. Ho, D. Liu, M. Wright, S. K. Das, Distributed detection of replicas with deployment knowledge in wireless sensor networks, 2009 IEEE International Conference on Pervasive Computing and Communications, pp. 1 – 6, 2009

[98] I. Onat, A. Miri, An intrusion detection system for wireless sensor networks, (WiMob'2005), IEEE International Conference on Wireless and Mobile Computing, Networking And Communications, vol. 3, pp. 253 - 259, 2005

[99] D. Curiac, O. Banias, F. Dragan, C. Volosencu, O. Dranga, Malicious node detection in wireless sensor networks using an auto regression technique, Networking and Services Third International Conference on, pp. 83 – 83, 2007

[100] D. Dallas, C. Leckie, K. Ramamohanarao, Hop-count monitoring: Detecting sinkhole attacks in wireless sensor networks, 15th IEEE International Conference on Networks, pp. 176 – 181, 2007

[101] A. Balakrishnan, P. Rino, A novel anomaly detection algorithm for WSN, 2015 Fifth International Conference on Advances in Computing and Communications (ICACC), pp. 118 – 121, 2015

[102] Z. Xiao, C. Liu, C. Chen, An anomaly detection scheme based on machine learning for WSN, 2009 First International Conference on Information Science and Engineering, pp. 3959 – 3962, 2009

[103] A. Abid, A. Kachouri, A. Mahfoudhi, Anomaly detection through outlier and neighborhood data in wireless sensor networks, 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 26 – 30, 2016

[104] The Personal Information Protection and Electronic Documents Act (PIPEDA), office of the privacy commissioner of Canada, Retrieved from https://www.priv.gc.ca/en/ (Dec 08 2019)

[105] Health Insurance Portability and Accountability Act (HIPAA) of 1996, public law 104-191, 104th Congress, Retrieved from https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996 (Dec 08 2019)

[106] European union directive on the protection of personal data, Retrieved from https://ec.europa.eu/info/index_en (Dec 08 2019)

[107] Electronic Communications Privacy Act of 1986 (ECPA), 18 U.S.C. § 2510-22. Retrieved from https://it.ojp.gov/privacyliberty/authorities/statutes/1285 (Dec 08 2019)

[108] A. Ramesh, A. Suruliandi, Performance analysis of encryption algorithms for information security, 2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT), pp. 840 – 844, 2013

[109] J. Postel, Internet Protocol (IP), STD 5, RFC 791, September 1981, Retrieved from http://www.rfc-editor.org/rfc/rfc791.txt (Dec 08 2019)

[110] J. Postel, Transmission Control Protocol (TCP), STD 7, RFC 793, DOI 10.17487/RFC0793, September 1981, Retrieved from https://www.rfc-editor.org/rfc/rfc793.txt (Dec 08 2019)

[111] Gartner top strategic predictions for 2018 and beyond, Retrieved from https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/ (Nov 29 2017)

[112] Gartner says 8.4 billion connected "things" will be in use in 2017, up 31 percent from 2016, Retrieved from https://www.gartner.com/newsroom/id/3598917/ (Dec 08 2019)

[113] Gartner says worldwide IoT security spending to reach $348 million in 2016, Retrieved from https://www.gartner.com/en/newsroom/press-releases/2016-04-25-gartner-says-worldwide-iot-security-spending-to-reach-348-million-in-2016 (Dec 08 2019)