Employing a Machine Learning Approach to Detect Combined Internet of Things Attacks Against Two Objective Functions Using a Novel Dataset

John Foley (40296224@live.napier.ac.uk), * Naghmeh Moradpoor (n.moradpoor@napier.ac.uk), and Henry Ochen (ochenyi@hotmail.co.uk)

School of Computing, Edinburgh Napier University, Merchiston Campus, 10 Colinton Road, Edinburgh, EH10 5DT

***Abstract* - One of the important features of Routing Protocol for Low-Power and Lossy Networks (RPL) is Objective Function (OF). OF influences an IoT network in terms of routing strategies and network topology. On the other hand, detecting a combination of attacks against OFs is a cutting-edge technology that will become a necessity as next generation low-power wireless networks continue to be exploited as they grow rapidly. However, current literature lacks study on vulnerability analysis of OFs particularly in terms of combined attacks. Furthermore, machine learning is a promising solution for the global networks of IoT devices in terms of analysing their ever-growing generated data and predicting cyber-attacks against such devices. Therefore, in this paper, we study the vulnerability analysis of two popular OFs of RPL to detect combined attacks against them using machine-learning algorithms through different simulated scenarios. For this, we created a novel IoT dataset based on power and network metrics, which is deployed as part of an RPL IDS/IPS solution to enhance information security. Addressing the captured results, our machine learning approach is successful in detecting combined attacks against two popular OFs of RPL based on the power and network metrics in which MLP and RF algorithms are the most successful classifier deployment for single and ensemble models.**

***Keywords:* IoT Dataset, Objective Functions, Combined IoT Attacks, Machine Learning, Network Metrics, Power Metrics**

I. INTRODUCTION

The Internet of Things (IoT) can be described as the ever-growing global network of smart devices with built-in sensing features and communication interfaces such as Local Area Network (LAN) interfaces, sensors, and Global Positioning devices (GPS). It is expected that by 2022 we will have around 50 billion IoT devices scattered across the globe, a 140 percent increase compared to 2018. Since 1999, when the IoT was conceived, the concept of these smart devices has evolved into a conceptual framework including augmented physical objects, heterogeneous devices and interconnection solutions to share information at scale, across the world (Atzori, Iera and Morabito, 2017). Routing Protocol for Low-Power and Lossy Networks (RPL) is used for IPv6 over Low-Power Wireless Personal Area Networks (6LoWPAN) and IoT networks. RPL link layers operate efficiently using nodes that connect through multi-hop paths to root devices; these devices are responsible for collating and distributing data. A Destination Oriented Directed Acyclic Graph (DODAG) is produced for each root device that accounts for node attributes, link cost, and Objective Function (OF). However, from a security point of view, RPL is a vulnerable protocol given that it does not integrate the security mechanisms needed to avoid intruders from unauthorized access to the data traveling across an IoT network. Due to this fact, RPL is exposed to several types of attack, (Sharma, Mishra and Jain, 2017) provided a concise table of RPL attacks for consideration. RPL nodes utilise OF to identify node of next hop based on power consumption and network metrics (Rehman et al., 2016). Minimum Rank with Hysteresis Objective Function (MRHOF) and Objective Function Zero (OF0) have been defined as two main OFs of IoT devices and RPL protocol by the Internet Engineering Task Force (IETF) work group. Detection and quick response to attacks against MRHOF and OF0 is difficult and the current research lacks study on vulnerability analysis of OFs. Additionally, little investigation has been done on automating the detection and response process particularly for the combined attacks against OFs on IoT networks. However, it is possible that Machine Learning (ML) and data mining can be used for anomaly-based intrusion detection with a focus on identifying attacks based on power consumption and network metrics (Le et al., 2016).

There are four Research Questions (RQs) that will be addressed throughout this paper:

- RQ1. Is there an available IoT dataset that is suitable to meet the research scope in this paper, or is the development of a novel dataset required?
- RQ2. What is the impact of pre-processing (for example normalisation, feature selection, and sampling) on classifier performance to detect combined attacks against MRHOF and OF0?
- RQ3. What is the most successful deployment of ML algorithms to detect a combination of attacks against MRHOF and OF0?
- RQ4. Are ML algorithms more successful in detecting combined attacks against MRHOF or OF0?

In this paper, we use ML to detect a combination of attacks against MRHOF and OF0 based on power consumption and network metric features. Additionally, to conduct our experiments and due to the lack of suitable IoT datasets, we developed a novel dataset which is focused on IoT features and attack parameters including packet delivery ratio and power consumption of nodes in various combined attack scenarios. Detecting a combination of attacks against OFs is a leading-edge technology that will become a necessity as next generation low-power IoT networks continue to be exploited everyday as they grow quickly. For this, we considered combined attacks such as: Rank & Version Attack, Rank & Blackhole Attack, Decreased Path Metric Attack, as well as Rank & Sybil Attack.

The remainder of this paper is organised as follows. In Section II, we review the related work based on our research questions and a critical understanding of available quality sources followed by our research methodology in Sections III. This is trailed by simulated experiments in Section IV, results and analysis in Section V, discussions in Section VI, conclusion, limitations, recommendations in Section VII, and acknowledgement in section VIII as well as references.

II. Literature Review

In this section, relevant academic papers are reviewed and discussed in five groups. IoT Methodologies, MRHOF and OF0 Attacks, IDS Methodologies & Feature Selection, Datasets & ML Classifiers, Pre-processor & Balancing Techniques were identified as the five core topics. Due to page limitation, we have picked two papers from each category to discuss here. However, more related papers from each group, along with gap analysis for our novel approach, can be found in Table 1. Publications, arguments and literature were selected based on practical and simulation experiments, expert opinion, evaluation, analysis and contrasting views. Scholarly literature search engines, libraries and journals were used to identify strengths, weaknesses and gaps in research.

With regards to IoT Methodologies, (Zarpelão et al., 2017) deliver a survey of IoT-IDS and discuss the use of ML, anomaly-based approaches, intrusion detection based on power consumption and analysis of objective function behaviour which are relevant to our research in this paper. The authors discuss power consumption as a parameter that can be used to analyse normal behaviour profile to detect malicious activity, based on mesh-under and route-over schemes. Each node is required to monitor power consumption at specified sampling rates and report deviations from expected values. Deviations from expected values are deemed malicious activity and as such the node is removed from the routing table. The paper expands on the concept of node behaviour focused on power consumption and suggests that packet overhead and memory consumption are adequate metrics that can be used for IoT-IDS. Although the paper provided a broad overview on IoT-IDS, technical depth is limited. Additionally, it is difficult to understand a detailed approach for IoT-IDS using ML from the research alone. Moreover, (Rehman et al., 2016) discuss Rank Attack as an Objective Function (RAOF) vulnerability aimed at RPL protocol. In order for RAOF to be successful, an attacking node corrupts routing metrics so that neighbour nodes' OF favour the attacker as a preferred parent node. The results of their simulation show the impact of the attack when considering a well-positioned attacking node within the RPL network. The paper is interesting since it identified a relationship between OF, power consumption, hop count and routing metrics when considering RAOF. However, it does not provide counter arguments to their approach. For example, a discussion on the relationship between power consumption and hop counts when considering an attack on routing metrics could have been explained in further detail. The paper provided a detailed analysis for OF vulnerabilities across RPL networks and the introduction of an attack that had not been identified in other research.

With regards to MRHOF and OF, (Airehrour et al., 2017) present a trust aware RPL model to detect Selective Forwarding and Blackhole attacks in IoT networks. The trust aware model is compared to MRHOF and OF0 to understand if their proposed solution is successful. Addressing their results, Selective Forwarding attacks against the trust aware model were able to be gradually and significantly reduced. This includes the isolation of malicious nodes. However, MRHOF and OF0 were not able to detect or isolate Selective Forwarding attacks. Additionally, while the trust aware protocol was able to detect and isolate Blackhole attacks through analysis of sent packet sequence and received sequence ID, MRHOF and OF0 were not able to do so. They also did not review or discuss in detail the techniques used by MRHOF and OF0 to detect and isolate malicious nodes. The lack of research in this area is significant to our work in this paper. Moreover, (Airehrour et al., 2018) discuss Secure Trust Protocol (SecTrust-RPL) designed to detect and isolate Rank and Sybil attacks through node trust relationships. Performance of SecTrust-RPL is compared to standard RPL protocol integrating with MRHOF and OF0. MRHOF is identified as a superior RPL protocol over OF0 based on performance metrics when considering network flow and resource. When considering Rank and Sybil attack alone, MRHOF demonstrates higher vulnerability than SecTrust-RPL. Although SecTrust-RPL has been identified as being a more secure protocol over MRHOF and OF0, there was no discussion around IDS that can be used with OF and RPL protocol. This is relevant to our research scope in this paper as RPL and OF may not be required to detect and isolate attacks if ML can be used with IDS. Experiments for MRHOF and OF0 utilising a suitable IDS compared to SecTrust-RPL would have provided a fair evaluation when considering IoT security.

With regards to IDS Methodologies and Feature Selection, (Sheikhan and Bostani, 2017) discuss a security mechanism to detect attacks against IoT networks based on a distributed architecture. Their proposed method is focused on using ML to identify Sinkhole and Selective Forwarding attacks that deviate from normal and abnormal behaviour. Addressing their results, the anomaly detection successfully identifies 80.95% of Sinkhole and Selective Forwarding attacks with a false alarm rate of 5.92%. The misuse-based detection was able to identify 97.88% of Sinkhole and Selective Forwarding attacks with a false alarm rate of 1.96%. Although misuse-based detection was able to identify a higher rate of Sinkhole and Selective Forwarding attacks with a low false alarm rate, this method is only able to detect known attacks. Although, the article highlights the significance of selecting noteworthy behaviour features including packet drop rate, packet receive rate, maximum hop count and average latency, further analysis will be required to select features relevant to OF attacks. (Napiah et al., 2018) discuss Compression Header Analysis Intrusion Detection System (CHA-IDS) coalesced with ML to detect 6LoWPAN and RPL combination attacks: Hello Flood, Wormhole and Sinkhole. CHA-IDS is focused on identifying anomaly and signature-based features for 6LoWPAN intrusion detection through raw data collection and analysis by means of six ML algorithms (MLP, SVM, J48, Naïve Bayes, Logistic, and Random Forest). They provide quantitative data suggesting CHA-IDS performs better than other 6LoWPAN IDS models for combined attack detection. CHA-IDS applies compression header data for 6LoWPAN as a detection feature in contrast to SVELTE and PONGLE that utilise rank and received signal strength indicators. Destination port, context identifier, destination context identifier, next header and pattern identified abnormal routing activities were used with ML algorithms to successfully detect attacks. Addressing their results, J48 was the most successful ML algorithm across a combined dataset while Random Forest ranked second. The strengths of the paper include research on 6LoWPAN and RPL vulnerabilities and flaws in current IDS methods.

With regards to Datasets and ML Classifiers, (Buczak and Guven, 2015) discuss a range of ML and data mining algorithms and classification techniques based on public datasets for intrusion detection. An overview of publicly available datasets is provided comprising: DARPA 1998/1999/2000, KDD 1999, NetFlow, tcpdump, DNS and SSH datasets. It is highlighted that during a research phase it will be important to ensure ML methods are trained using the same dataset to ensure comparison with other research is reliable. However, despite KDD 1999 being the best available labelled dataset, it is limited by attacks that have occurred since the dataset was produced. This may be an issue when considering a reference dataset for use with MRHOF, OF, RPL and IoT. Moreover, (Alam et al., 2016) present a paper on eight data mining algorithms: ANNs, Deep Learning ANNs (DLANNs), C4.5, C5.0, SVM, Naïve Bayes (NB), K-Nearest Neighbours and Linear Discriminant Analysis (LDA) for use with IoT. Their research aim is to understand if conventional data mining algorithms work for IoT datasets and if not, are new algorithms required. For their research, three sensor datasets from University of California Irvine (UCI) data repository were provided. Results conclude DLANN, ANN, C4.5 and C5.0 performed better than LDA, NB, NN and SVM when considering accuracy and elapsed time for IoT datasets. C4.5 and C5.0 were identified to provide high accuracy and processing speed whilst remaining memory efficient. DLANNs and ANNs memory efficiency was poor and computationally expensive although identified as having the highest rates of accuracy. The paper discusses an area of research that was difficult to identify during literature review given that most research into ML and IoT utilise DARPA or KDD datasets. Although the paper includes novel research into an area not commonly explored, the paper would have benefited from a detailed discussion around the three datasets provided by UCI.

With regards to Pre-processor and Balancing Techniques, (Yin and Gai, 2015) discuss the challenges of data mining and ML relating to new and enormous data types introduced to solve complex problems. The paper discusses classification methods, pre-processing, feature selection, and data sampling. The publication explains that there are many classification algorithms available that are mostly based on balanced high-quality datasets. Pre-processing is explained as a common method used to improve the accuracy of a dataset by reducing the number of features selected and by sampling well. Their experimental activity is designed to understand how to achieve and improve pre-processing techniques to deliver high-quality datasets. C4.5 classifier was the only algorithm used during experimentation to eliminate conflicting results across a range of 12 datasets. Results conclude that the accuracy of a classifier is more reliable when feature selection is conducted prior to sampling data. In the event that data is largely imbalanced, experimental results conclude that it is better to under sample data rather than over sample when considering minority class. The paper could have benefited with the inclusion of other pre-processor stages into experimental activity to improve the accuracy of a dataset further.

Critical analysis of the current literature identified the following key areas of interest to address in this paper: power consumption and network related metrics, combination of IoT attacks, MRHOF & OF0 vulnerability analysis, feature selection and the development of a novel dataset based on IoT attacks. For instance, unlike (Zarpelão et al., 2017) and (Napiah et al., 2018), our work in this paper includes the combination of MRHOF and OF0 attacks considering power consumption and network-related metrics as part of a ML-IDS. Furthermore, unlike (Buczak and Guven, 2015) and (Yin and Gai, 2015), a novel IoT dataset has been developed focused on MRHOF and OF0 attacks including pre-processing techniques, feature reduction, sampling and normalisation. Additionally, as far as we are aware, no one else has successfully employed time series ML classifiers alongside a novel IoT dataset, whilst detecting a combination of attacks against multiple objective function (e.g. OF0 and MRHOF) based on

network and power consumption metrics. Table 1 provides a summary of our gap analysis based on the most relevant reviewed quality papers.

Table 1: Summary of Our Gap Analysis

Author	Gap Analysis for Novel Approach
IoT Methodologies	
Zarpelão, B. B. et al (2017)	The paper identified a gap using ML and IDS based on power consumption. **Power consumption metrics will be in the scope of our work.**
Rehman, A. et al (2016)	The paper identified a gap using ML & IDS based on power consumption and hop count for RAOF. **Power consumption and hop count will be in the scope of our work.**
Le, A. et al (2016)	The combination of attacks have not been considered. Power consumption and dropped packet features could be used as a novel approach to anomaly-based detection. **A combination of IoT attacks along with power consumption and dropped packet will be in the scope of our work.**
MRHOF and OF0 Attacks	
Airehrour, D. et al (2017)	Research failed to detect individual attacks against OF0 and MRHOF. **MRHOF and OF0 will be in the scope of our work for each IoT combination attack.**
Airehrour, D. et al (2018)	The paper identified a gap for detecting/isolating a combination of Rank and Sybil attacks within MRHOF and OF0. **A combination of Rank and Sybil attack along with MRHOF & OF0 will be in the scope of our work for each IoT combination attack.**
Mehta, R. and Parma, M. (2018)	The paper identified a gap that possible OF attacks should be detectable. **A combination of IoT attacks along with MRHOF & OF0 will be in the scope of our work for each IoT combination attack.**
IDS Methodologies and Feature Selection	
Sheikhan, M. et al (2017)	Research failed to detect unknown attacks using selected features for misuse-based detection. **Power consumption and dropped packet metrics will be in the scope of our work.**
Mayzaud, A. et al (2016)	Despite authors claiming their research as a feasible solution for anomaly-detection for IoT, there is no evidence of detection for a wide range of attacks beyond DAG. **A combination of IoT attacks will be in the scope of our work.**
Lee, T.H. et al (2014)	OF and MRHOF are not discussed within discussion of detecting malicious activity based on power consumption and network flow. **MRHOF and OF0 for each IoT combination attack along with power consumption, hop count and dropped packet metrics will be in the scope of our work.**
Sousa, N. et al (2017)	The paper discussed OF-FL, CAOF and other relevant OF routing metrics and then excluded them during simulation. **MRHOF and OF0 will be in the scope of our work for each IoT combination attack.**
Napiah, M. N. et al (2018)	The paper discussed reducing features from 77 to 5 removing power consumption to ensure ML algorithms were efficient. **Power consumption metrics along with feature reduction strategy will be in the scope of our work.**
Datasets and ML Classifiers	
Haq, N. et al (2015)	The paper reviews 49 related studies and highlights considerations to be made when developing a ML-IDS. **ML approach, classifier methods, suitable algorithms, datasets and features selection are in the scope of our work.**
Nannan, L. et al (2018)	Research identified a high false alarm rate for anomaly detection. **ML approach, classifier methods, suitable algorithms, datasets and features selection are in the scope of our work.**
Buczak, A. and Guven, E. (2015)	KDD 1999 is limited by attacks that have occurred since the dataset was produced. This includes IoT attacks. **The identification or development of a novel dataset focused on IoT features and attacks is in the scope of our work.**
Alam, F. et al (2016)	The paper identified little research into the use of conventional ML algorithms with IoT datasets. **The identification or development of a novel dataset focused on IoT features and attacks along with ML approach, classifier methods, suitable algorithms, datasets and features selection are in the scope of our work.**
Pre-processor and Balancing Techniques	
Yin, H. and Gai, K. (2015)	The paper reviews 12 datasets and highlights considerations to be made when developing an imbalanced dataset. **Pre-processor techniques, feature extension, sampling as well as train, test and validate dataset are in the scope of our work.**

III. Research Methodology

The aim of this paper is to use ML to detect a combination of attacks against OF0 and MRHOF, as two popular OFs for RPL protocol, based on power consumption and network metrics using a novel dataset. These findings will build on previous research in the field to produce a novel research. Therefore, the considerations include:

- Identification and development of a novel dataset focused on IoT features and attacks
- Identifying the most successful deployment of ML algorithms and classifiers
- The impact of pre-processing, normalisation, feature selection and sampling on performance of ML algorithms and classifiers
- Employing and assessing the success rate of detecting a combination of attacks against OF0 and MRHOF

In this paper, eight experiments have been developed and presented based on the remarks identified during gap analysis presented in Table 1. Further detail will be provided during the simulated experiments. Our research methodology follows CRISP-DM (Shafique and Qaiser, 2014) that provides a structured approach for ML-based projects. We employed the six phases of CRISP-DM as follows:

- Business Understanding: (Literature Review, Development of Research Questions, Project Methodology, Gap Analysis, Research Aims)

- Data Understanding: (Exploring Datasets)
- Data Preparation: (Data Pre-processing)
- Modelling: (ML Algorithms and Classifiers)
- Evaluation: (Performance Evaluation)
- Deployment: (This phase is out of the scope of this paper and will be discussed during future recommendations)

IV. Simulated Experiments

This section is designed to outline simulated experiments conducted following CRISP-DM process based on our research methodology. It includes data exploration, pre- processing, ML classifiers, classifier ensemble, and feature selection. The experiments will be run a number of times in a consistent manner to ensure the integrity of results. They are designed as follows.

- ✓ Experiment 1 – Pre-processed Dataset: Considering All Attributes and Metrics
- ✓ Experiment 2 – Normalisation: Considering All Attributes and Metrics
- ✓ Experiment 3 – Normalisation: Considering Network Attributes and Metrics
- ✓ Experiment 4 – Normalisation: Considering Power Attributes and Metrics
- ✓ Experiment 5 – Considering Feature Selection
- ✓ Experiment 6 – Considering Classifier Ensemble
- ✓ Experiment 7 – Considering Detecting Attacks Against MRHOF and OF0
- ✓ Experiment 8 – Considering Detecting Attacks Against MRHOF and OF0 with Balanced Class

1. Exploring Datasets

A range of approaches were considered during dataset exploration including publicly available datasets, privately-owned datasets, implementation of an IoT lab to capture relevant data and IoT simulation.

Publicly available datasets including DARPA 1998/1999/2000 and KDD 1999 were considered since they are used in 71% of ML research experiments (Buczak and Guven, 2015). Despite these datasets being available and well labelled they do not have examples of attacks that have occurred since the datasets were produced. This is presenting an issue when considering attacks against OF0, MRHOF, RPL and IoT. Leading research professionals, universities and corporate organisations were approached and asked to provide raw datasets for research purposes. For security, privacy and the protection of intellectual property each of these organisations were unwilling to share their privately-owned datasets. The implementation of an IoT sensor lab was considered but was unfeasible due to the limited budget and geographic location of available resource.

The identification and development of a novel dataset focused on IoT features and attacks was therefore conducted using simulation. (Alam et al., 2016) acknowledge this approach as difficult and time consuming requiring significant work to collect, label and pre-process an IoT dataset to ensure accuracy of results. A simulation dataset was produced to meet project scope using Contiki and Cooja. The raw dataset was provided as a project resource to understand, analyse and evaluate prior to data pre-processing and modelling ML algorithms to detect attacks against OF0 and MRHOF.

The dataset contained 24 attributes based on network and power metrics. They are all presented and detailed in Table 2. The dataset also includes 418 instances. Attribute and incident metrics were captured in a Contiki and Cooja simulation environment from various sensors during normal activity, and in every 30 seconds whilst under attack. The simulation was configured with eleven network nodes and one malicious node. The malicious node can be seen in Figure 1 labelled as number twelve.

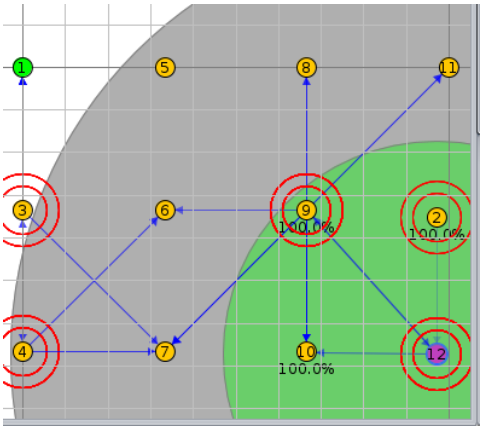

Figure 1: Simulation Environment

Benign activity and four combined malicious attacks were monitored during the simulation scenarios. The malicious attacks include Rank & Version, Rank & Blackhole, Rank & Sybil and Decreased Path Metric against OF0 and MRHOF. Malicious and benign activities can be identified as attribute 23 in our dataset and is the selected class for our experiments. The class is imbalanced. This will be rectified during the data pre-processing phase.

Table 2: Explanation of Dataset Features

Feature	Remarks
Node	Node ID
Received	Packets successfully received
Dups	Duplicate packets received
Lost	Estimated lost packets
Hops	Average number of hops
RtMetric	Average routing metric
ETX	Expected transmission count
Churn	Next hop change count
Beacon Interval	Broadcast by node to sync RPL network
Reboots	Total number of reboots
CPU Power	Average CPU power consumption
LPM Power	Average low power mode consumption
Listen Power	Average listen power
Transmit Power	Average transmission power
Power	Average power consumption
On-time	Power measure time
Listen Duty Cycle	Percentage time listening signal is active
Transmit Duty Cycle	Percentage time transmission signal is active
Avg inter-packet time	Average delay between packets
Min inter-packet time	Minimum delay for the packet to arrive
Max inter-packet time	Maximum delay for the packet to arrive
Simulation time	Average simulation time
Malicious / Benign	Benign activity or type of attack
Objective Function	MRHOF or OF0

2. Data Pre-processing

Data pre-processing phase is designed to prepare the raw dataset for our eight experiments. Pre-processing and data reduction were essential phases of the project. Therefore, sufficient time was spent ensuring a suitably labelled dataset was created to provide high quality results for analysis. Data pre-processing is aimed at reducing the complexity of a dataset so ML models can process data more accurately and faster than a raw dataset. When implementing a data mining process, which is CRISP-DM in this paper, pre-processing often requires more effort and time than the entire data analysis process in excess of 50% total effort (Ramírez-Gallego et al., 2017). The dataset in Table 3 shows a representative example of some of the complex attributes and instances that we captured over the simulation scenarios in this paper. For the data pre-processing phase of this research, we have picked: data cleansing, transformation & feature reduction, normalisation & data analysis, sampling, as well as training, testing and cross-validation stages. They will be applied on our raw dataset as follows.

2.1. Data Cleaning, Transformation & Feature Reduction

The raw dataset seen in Table 3 was reviewed and issues were identified such as missing, incomplete and inconsistent values. Additionally, the irrelevant data and errors were identified. To reduce the complexity of the 10,032 entries within the raw dataset, data cleaning and transformation was conducted to represent all data in a standard numeric form. Table 4 describes the steps taken for each feature.

Features that were of no benefit to the ML model, nor did they contain relevant data, were removed from the dataset. Similarly, instances were reviewed and the entries containing no predictive power were removed reducing total instances from 418 to 338. Entries that contained null values or errors were replaced with the mean value for that entry. At this stage, the initial pre-processed dataset was completed and was used later in the project post normalisation and sampling. It was important to use this dataset for initial assessment and comparison against the final pre-processed dataset to understand the effect that normalisation has on overall performance.

Table 3: Raw Dataset

Node	Received	Dups	Lost	Hops	RtMetric	ETX	Churn	Beacon Interval	Reboots	CPU Power	On-time	Listen Duty Cycle	Max inter-packet time	Simulation t	Malicious / Benign	Objective Function
1.1	0	0	0	0.000	0.000	0.000	0	0	0	0.000	0	0.000	0	40	m-rankandversion	mrhof
2.2	60	0	1	3.017	1040.183	48.771	1	1645266	0	0.3859	7756254	0.8725	138000	40	m-rankandversion	mrhof
3.3	62	0	1	1.000	512.000	16.000	0	1643451	0	0.4022	790122	0.8768	113000	40	m-rankandversion	mrhof
4.4	59	0	0	2.000	773.051	32.000	0	1723711	0	0.3543	717160	0.7488	113000	40	m-rankandversion	mrhof
5.5	63	0	0	1.143	573.127	19.048	4	1605000	0	0.4258	787136	0.9806	118000	40	m-rankandversion	mrhof
6.6	61	0	2	1.098	537.180	17.574	2	1667459	0	0.4715	797966	1.0172	204000	40	m-rankandversion	mrhof
7.7	59	0	1	2.000	770.237	32.000	1	1668305	0	0.4091	722396	0.9103	116000	40	m-rankandversion	mrhof
8.8	59	0	0	2.000	771.576	32.059	0	1724813	0	0.3958	723483	0.7992	116000	40	m-rankandversion	mrhof
9.9	58	0	0	2.000	770.603	32.000	0	1728568	0	0.4448	713151	0.8766	117000	40	m-rankandversion	mrhof
10.1	59	0	0	2.000	774.186	32.000	0	1690525	0	0.3749	663158	0.7730	83000	40	m-rankandversion	mrhof
11.1	59	0	0	3.000	1036.542	48.167	0	1701491	0	0.3573	736572	0.7743	115000	40	m-rankandversion	mrhof
1.1	0	0	0	0.000	0.000	0.000	0	0	0	0.0000	0	0.0000	0	10	b-rankandblackhole	OF0
2.2	23	0	0	3.000	1024.000	48.000	0	1167913	0	0.3794	282283	0.7976	104000	10	b-rankandblackhole	OF0
3.3	23	0	0	1.000	512.000	16.000	0	1099521	0	0.3870	277460	0.7298	111000	10	b-rankandblackhole	OF0

Table 4: Data Pre-processing Steps

Feature	Remarks
Node	Rounded to 0 decimal places
Received	No change
Dups	Feature removed
Lost	No change
Hops	Rounded to 0 decimal places
RtMetric	Rounded to 0 decimal places
ETX	Rounded to 0 decimal places
Churn	No change
Beacon Interval	No change
Reboots	Feature removed
CPU Power	Converted from W to fW (1×10 power−15)
LPM Power	Converted from W to fW (1×10 power −15)
Listen Power	Converted from W to fW (1×10 power −15)
Transmit Power	Converted from W to fW (1×10 power −15)
Power	Converted from W to fW (1×10 power −15)
On-time	No change
Listen Duty Cycle	No change
Transmit Duty Cycle	No change
Avg inter-packet time	No change
Min inter-packet time	No change
Max inter-packet time	No change
Simulation time	Feature removed
Malicious / Benign	All benign and malicious activities were grouped and values converted to numeric. 1 = Rank and Version Attack 2 = Benign Activity 3 = Rank and Blackhole Attack 4 = Decreased Path Metric Attack 5 = Rank and Sybil Attack
Objective Function	All values were converted to numeric. 1 = MRHOF, 2 = OF0

2.2. Normalisation and Data Analysis

Normalisation is a scaling technique that is used to provide a new range of data from an existing range. Min-Max normalisation can be used to fit data from one range into a predefined boundary in another. Due to the dataset containing complex numbers, statistical analysis was conducting. Additionally, standard deviation was used to categorise data in a nominal format to achieve the aim of predefined boundaries (Linn et al., 2016). The average value of each attribute was taken, and the standard deviation was calculated. A boundary range between 1 – 14 was determined based on the diagram presented in Figure 2. For example, numbers 7 & 8 represent the most normal behaviour, or behaviour closest to mean, and numbers 1 & 14 represent the most abnormal behaviour or behaviour furthest from mean, Figure 2. Despite feature reduction condensing entries to 7,098, an automated system was developed to produce a new range of data from the existing range based on standard deviation and boundary selection to ensure data normalisation was accurate and timely when dealing with large datasets.

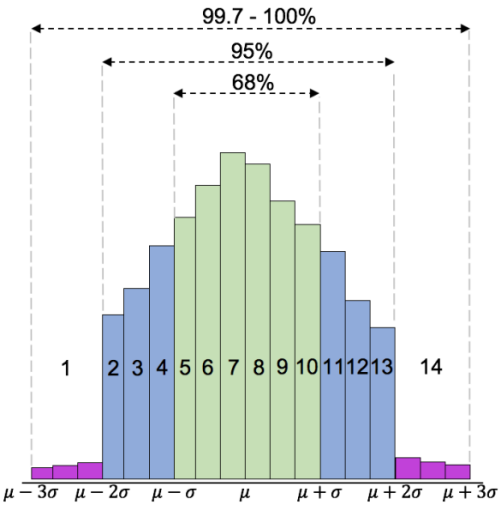

Figure 2: Normalisation Boundary Range Based on Standard Deviation

The dataset in Table 5 shows a representative example of the attributes and instances that have undergone data cleansing, transformation, feature reduction and normalisation phases of data pre-processing stage. As can be seen, all of the values are numeric and fall between the boundary range of 1 -14. The dataset is now in a format that is suitable to be loaded into WEKA [33] and converted to .arrf file prior to sampling and creating training, testing, and cross-validation datasets. WEKA is a popular and powerful tool for data mining and machine learning.

Table 5: Final Pre-processed Dataset

Node	Received	Lost	Hops	RtMetric	ETX	Churn	Beacon Interval	CPU Power	LPM power	Listen Power	Transmit Power	Power	On-time	Listen Duty	Transmit Duty	Avg Inter	Min Inter	Max Inter	OF	Mal/Ben
2	8	1	3	12	13	1	9	7	8	7	7	7	9	7	7	7	8	7	1	1
3	9	1	1	4	4	0	9	7	7	7	7	7	9	7	7	7	6	6	1	1
4	8	0	2	8	8	0	10	6	9	6	6	7	8	6	7	7	8	6	1	1
5	9	0	1	5	5	4	9	8	7	8	8	8	9	8	8	7	8	7	1	1
6	9	2	1	4	4	2	9	9	5	8	7	8	9	8	7	7	7	6	1	1
7	8	1	2	8	8	1	9	7	7	7	7	7	8	7	7	7	6	6	1	1
8	8	0	2	8	8	0	10	7	8	7	6	7	8	7	7	7	6	6	1	1
9	8	0	2	8	8	0	10	9	6	7	7	7	8	7	7	7	6	6	1	1
10	8	0	2	8	8	0	9	7	8	7	6	7	8	7	7	7	6	6	1	1
11	8	0	3	12	12	0	9	6	9	7	6	7	8	7	7	7	7	6	1	1
2	4	0	3	12	12	0	5	6	9	7	6	7	4	7	7	6	8	6	2	2
3	4	0	1	4	4	0	6	7	8	6	5	6	4	6	6	6	5	6	2	2
4	4	0	2	8	8	0	5	6	9	7	6	7	4	6	7	6	8	6	2	2
5	4	0	1	4	4	0	5	7	8	7	5	7	4	7	6	7	8	6	2	2
6	4	0	1	4	4	0	5	8	6	7	5	7	4	7	6	7	9	6	2	2
7	4	0	2	8	8	0	5	7	8	7	6	7	4	7	7	6	9	6	2	2
8	4	0	2	8	8	0	5	7	8	7	7	7	4	7	7	7	6	6	2	2
9	4	0	2	8	8	0	5	9	6	8	8	8	4	8	8	7	8	6	2	2
10	4	0	2	8	8	0	5	6	8	7	6	7	4	7	7	7	7	6	2	2
11	4	0	3	12	12	0	5	6	9	7	6	7	4	7	7	7	7	6	2	2

2.3. Sampling

Prior to conducting sampling for the final pre-processed dataset, some housekeeping was performed within WEKA. Package Manager was used to load numeric to nominal format, randomisation and SMOTE filters. The .csv dataset was loaded into WEKA, converted to .arrf file and the numeric to nominal filter applied to ensure all entries were stored in nominal format. A sampling strategy was considered. It was identified that there was a class imbalance for malicious and benign class (64, 140, 65, 33 and 36 malicious/benign events distributed across the dataset), Figure 3. (Haixiang et al., 2017) acknowledge that ML algorithms are typically sensitive to detecting majority class and not minority. Therefore, a balanced dataset was used. Class imbalance can be rectified by oversampling the minority class or by under sampling the majority class depending on which event is to be identified. Over-sampling can reduce performance for large datasets and introduces the possibility of overfitting if data is not randomised. Under-sampling introduces the possibility of removing important data when a minority class is particularly low, reducing the amount of data available for training, testing and cross-validation. In this paper, both over and under sampling techniques will be evaluated to ensure benign activity and IoT attacks will have the best opportunity of being detected. We used SMOTE filter to over sample events 1, 3, 4 and 5 (Rank & Version, Rank & Blackhole, Decreased Path Metric and Rank & Sybil attacks) so each event had approximately 135 instances, Figure 4. Each time SMOTE filter sampled it placed new instances at the bottom of the dataset introducing a possible overfitting. Therefore, at the end of the filtering process data was randomised with a separate filter. Oversampling with SMOTE increased total instances from 338 to 674. We also used Spread Sub Sample filter to under sample events 1, 2, 3 and 5 (Rank & Version, benign events, Rank & Blackhole and Rank & Sybil attacks) so that each event had 33 instances, Figure 5. To adjust filter settings, random seed was set to 1 to ensure that each sample was randomised. Under sampling decreased total instances from 338 – 165.

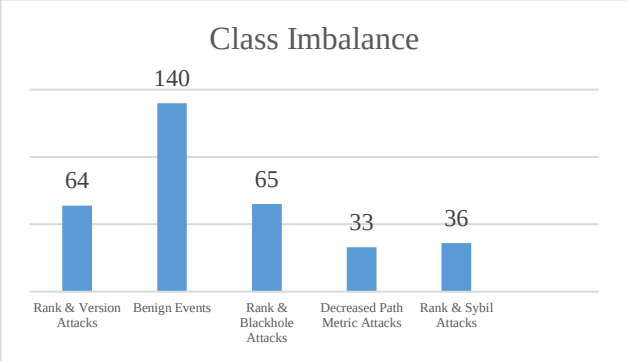

Figure 3: Class Imbalance Malicious/Benign Class

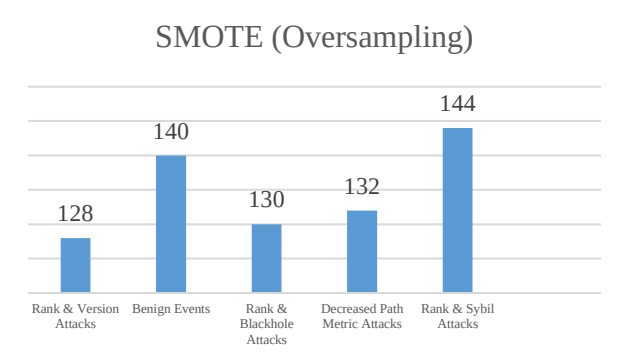

Figure 4: SMOTE Oversample Malicious/Benign Class

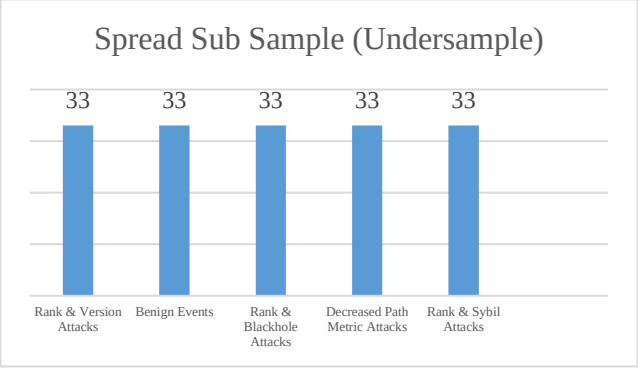

Figure 5: Spread Sub Sample (Undersample) Malicious/Benign Class

2.4. Training, Testing and Cross-validation

To produce a classifier, data is processed through a ML algorithm. Once a classifier is produced data is processed through the classifier generating results for evaluation. It is important that data processed through the ML algorithm and classifier are not from the same dataset. This project is resourced with one simulation dataset; therefore, the dataset must be split into training, testing and cross-validation sets. Training data contains 70% of total instances with the remaining 30% split equally between testing and cross-validation sets. WEKA does not have direct functionality to do this so the resample filter was exploited to achieve the aim for both SMOTE and Spread Sub Sample datasets.

3. Machine Learning Algorithms and Classifiers

WEKA was used to train, test and cross-validate our selected five classifiers: Naïve Bayes (NB), Support Vector Machines (SVM), Multilayer Perceptron (MLP), Random Forest (RF) and ZeroR classifiers. ZeroR was used to determine a performance baseline.

NB was the first model to be built in WEKA and was completed using default settings to take the product of probabilities providing a forecast ratio to identify likely outcomes. NB's default options were not altered since the dataset was in nominal format. Therefore, "useKernelEstimator" and "useSupervisedDiscretization" were not required to be changed.

When developing SVM parameters, it is important to understand that the model is designed to separate classifiers using a boundary. When using WEKA's LibSVM classifier, setting a suitable boundary allowed the generalisation of a training dataset to be more accurate. This was achieved by optimising parameters setting cost to "C" and kernel type to "gamma", relating to X and Y axis respectively.

Default MLP settings were applied during the initial classifier model and subsequently tuned to enhance results. A critical parameter that was evaluated was hidden layers. Hidden layer parameter within WEKA can be used to train data on attributes, classes or combinations of attributes and classes. This parameter can also be adjusted to determine the number of layers within the MLP model. The MLP model used for experimentation consisted of three layers trained on both attributes and classifiers. Increasing training time from 500 – 2000 epochs improved results allowing backpropagation and a multi-layer approach more time to train each MLP layer.

When developing the RF algorithm consideration was made to the depth of the tree and number of features to be randomly selected. The default setting was applied, the depth of tree set, and number of features set to 0 (unlimited depth). When processing large datasets these values can be set to reduce the depth of the tree and number of features to enhance the performance of the classifier. Since the dataset is relatively small the RF classifier was able to process data with unlimited depth and features providing best results overall.

4. Classifier Ensemble

Classifier ensemble techniques were used to gain highly accurate ML classifications through the combination of multiple ML algorithms. The parameters of AdaBoost, Bagging and Stacking were researched, and experiments conducted with ML algorithms but were unable to improve results beyond single classifiers. Voting was used successfully. Additionally, classifier and combination rule parameters were amended. However, optimum results were obtained setting classifier to "MPL and RF" with combination rule set to "Average of Probabilities". For all classifier models, excluding NB since the option is not available, seed was set to zero ensuring random number generation was not conducted so comparison of results would be consistent.

5. Feature Selection

For the purpose of practical investigation feature selection was split into the following areas: all dataset metrics, network metrics and power metrics. Each of these areas were evaluated during experiments 2 – 4 to understand which features provided greatest detection. Feature selection was developed further during experiment 5 removing attacks that were not detectable based on power consumption metrics.

This section outlined the implementation of the CRISP-DM process based on the steps described during research methodology. The application of data exploration, pre-processing, ML classifiers, classifier ensemble, and feature selection have been described in detail and results will be presented in the next section.

V. Results & Analysis

There were eight simulated experiments conducted to capture results for discussion. The experiments were designed to understand how ML can be used to detect a combination of attacks against OF0 and MRHOF based on power consumption and network metrics. The experiments were also designed to understand the impact normalisation, sampling, feature selection and classifier ensemble techniques have on results. The classifier ZeroR was used to determine a performance baseline as a reference to consider when comparing NB, SVM, MLP and RF algorithms. The baseline confirmed malicious and benign classifier prediction at 20.59% which is reasonable since there is a single benign behaviour and four attacks. Each experiment becomes more complex during investigation to deliver results and understand whether the aims of this study have been achieved.

✓ Experiment 1 – Pre-processed Dataset: All Attributes and Metrics

The aim of *Experiment 1* is to create classifiers for each ML algorithm based on a pre-processed dataset considering all attributes and metrics. This means considering both power and network parameters from our novel dataset. Results are captured in Table 6 using 10-fold cross-validation techniques for both SMOTE and Spread Sub Sample balancing methods.

The overall aim of *Experiment 1* is to:

- Compare each classifier against ZeroR
- Compare results from balancing techniques
- Carry results forward to experiment 2 for comparison against a normalised dataset

Correctly classified instances alone can lead to inaccurate results when evaluating ML algorithms and classifiers. However, for *Experiment 1,* this was considered a suitable evaluation tool for comparison against ZeroR classifier at a measurement of 20.59%.

As can be seen in Table 6, for SMOTE and Spread Sub Sample balancing techniques, each classifier improved performance on ZeroR with the exception of Sub Sample – SVM. Additionally, SMOTE outperformed Sub Sampling for this experiment. These results will be considered against normalised pre-processed datasets for all attributes and metrics in *Experiment 2*.

Table 6: Experiment 1 Results

	Correctly Classified Instances (%)	**Balancing Technique Average (%)**
ZeroR	20.59	20.59
SMOTE – NB	86.14	**73.76**
SMOTE – SVM	34.65	
SMOTE – MLP	**91.09**	
SMOTE – RF	83.17	
Sub Sample – NB	**52.00**	34.00
Sub Sample – SVM	12.00	
Sub Sample – MLP	24.00	
Sub Sample – RF	48.00	

✓ Experiment 2 – Normalisation: All Attributes and Metrics

The aim of *Experiment 2* is to create classifiers for each ML algorithm based on a normalised pre-processed dataset for all attributes and metrics. This means considering both power and network parameters from our novel dataset. Results are captured in Table 7 using 10-fold cross validation techniques for both SMOTE and Spread Sub Sample balancing techniques.

The overall aim of *Experiment 2* is to:

- Compare each classifier against *Experiment 1* results to understand the impact of normalisation
- Compare balancing techniques on the overall performance of classifiers
- Identify highly efficient algorithms for detecting specific attacks
- Carry results forward to *Experiment 3 and 4* to compare against network and power metrics

Correctly classified instances will be used initially to compare the value normalisation plays within ML. We considered Root Mean Square Error (RSME), Mean Absolute Percentage Error (MAPE), Receiver Operating Characteristics (ROC), correctly classified instances and confusion matrix as evaluation metrics for *Experiments 2 – 6.*

As can be seen in Table 7, for SMOTE and Spread Sub Sample, balancing each classifier post pre-processing and normalisation improved performance on *Experiment 1* with the exception of SMOTE – NB. Additionally, in *Experiment 2*, SMOTE outperformed Sub Sampling and SMOTE – MLP outperforms all other classifiers.

We have also observed the confusion matrix and accuracy by class which resulted in the recognition of highly efficient algorithms for detecting specific attacks for each classifier.

The following highly efficient algorithms were identified: SMOTE – NB for Decreased Path Metric attacks with a ROC of 0.99, SMOTE – MLP for Rank & Blackhole attacks with a ROC of 1.00 (Rank & Version and Decreased Path Metric attacks scored ROC 0.99) and SMOTE – RF for detecting all attacks with a ROC in excess of 0.99 for each.

These results will be considered against network and power metrics in *Experiment 3 & 4* with a focus on ROC average, to determine performance of a classifier. Overall, in *Experiment 2*, MLP and RF algorithms performed best when balancing and normalisation had been oversampled, Table 7.

Table 7: Experiment 2 Results

	RSME	MAPE (%)	ROC Average	Correctly Classified Instances (%)	Balancing Technique Average (%)
ZeroR	N/A	N/A	N/A	20.59	20.59
SMOTE – NB	0.23	24.22	0.98	84.31	**87.99**
SMOTE – SVM	0.27	22.01	0.89	82.35	
SMOTE – MLP	**0.16**	**10.15**	**0.99**	**93.14**	
SMOTE – RF	0.16	20.48	0.99	92.16	
Sub Sample – NB	0.32	41.90	0.88	68.00	65.00
Sub Sample – SVM	0.51	79.83	0.64	36.00	
Sub Sample – MLP	0.29	31.20	0.91	76.00	
Sub Sample – RF	0.28	55.86	0.94	80.00	

✓ **Experiment 3 – Normalisation: Network Attributes and Metrics**

The aim of *Experiment 3* is to create classifiers for each ML algorithm based on a normalised pre-processed dataset for network attributes and metrics. Results are captured in Table 8 using 10-fold cross validation techniques for both SMOTE and Spread Sub Sample balancing techniques.

The overall aim of *Experiment 3* is to:

- Compare each classifier against *Experiment 2* results to understand the significance of network attributes and metrics on a classifiers performance
- Compare balancing techniques
- Identify highly efficient algorithms for detecting specific attacks
- Carry results forward to *Experiment 4* to compare against power metrics

As seen in Table 8, reducing the dataset to only include network metrics improved performance for all classifiers during *Experiment 3* with the exception of SMOTE – RF. Additionally, SMOTE outperformed Spread Sub Sampling technique. Moreover, highly efficient algorithms for detecting specific attacks using confusion matrix were: SMOTE – NB, SMOTE - MLP and SMOTE - RF which detected each attack with a ROC in excess of 0.99. SMOTE – MLP successfully identified Rank and Blackhole attacks 100% of the time with no errors. These results will be considered against power metrics in *Experiment 4*.

Table 8: Experiment 3 Results

	RSME	MAPE (%)	ROC Average	Correctly Classified Instances	Balancing Technique Average (%)
SMOTE – NB	0.19	19.13	0.99	86.28	88.68
SMOTE – SVM	0.25	19.67	0.90	84.29	
SMOTE – MLP	**0.15**	**11.15**	**0.99**	**93.07**	
SMOTE – RF	0.18	24.58	0.99	91.09	
Sub Sample – NB	0.30	40.63	0.92	76.00	71.00
Sub Sample – SVM	0.47	69.85	0.69	44.00	
Sub Sample – MLP	0.28	29.25	0.94	76.00	
Sub Sample – RF	0.25	51.16	0.96	88.00	

✓ **Experiment 4 – Normalisation: Power Attributes and Metrics**

The aim of *Experiment 4* is to create classifiers for each ML algorithm based on a normalised, pre-processed dataset for power attributes and metrics. Results are captured using 10-fold cross validation techniques for both SMOTE and Spread Sub Sample

balancing techniques.

The overall aim of *Experiment 4* is to:

- Compare each classifier against *Experiment 2* and *3* results to understand the significance of power attributes and metrics on the classifier performance
- Compare balancing techniques
- Identify highly efficient algorithms for detecting specific attacks for classifier ensemble
- Carry results forward to *Experiment 5* to compare against classifier ensemble techniques to improve performance for power metrics

As can be seen in Table 9, reducing the dataset to only include power metrics significantly decreased performance for all classifiers during *Experiment 4*. It is worth noting that despite reducing the performance significantly each classifier performed better than the ZeroR baseline of 20.59%, demonstrating that power metrics obtain predictive power. Additionally, SMOTE outperformed Sub Sampling techniques.

Given the confusion matrix, there were no highly efficient classifiers for detecting specific attacks. However, there were some moderately efficient classifiers that should be considered for evaluation during *Experiment 5* and *6*.

Decreased Path Metric and Rank & Version attacks were detected based on power consumption metrics during *Experiment 4*. SMOTE – NB detected Decreased Path Metric attack with a ROC of 0.90. SMOTE – MLP detected Decreased Path Metric and Rank & Version attacks with a ROC of 0.91 and 0.94 respectively. SMOTE – RF detected Decreased Path Metric and Rank & Version attacks with a ROC of 0.93 and 0.96 respectively.

In conclusion, decreased Path Metric and Rank & Version attacks were detected based on power consumption metrics during *Experiment 4*.

Table 9: Experiment 4 Results

	RSME	MAPE (%)	ROC Average	Correctly Classified Instances (%)	Balancing Technique Average (%)
SMOTE – NB	0.38	73.45	0.80	47.06	**52.45**
SMOTE – SVM	0.45	63.58	0.70	49.02	
SMOTE – MLP	0.33	61.44	0.86	55.88	
SMOTE – RF	0.33	62.01	0.86	57.84	
Sub Sample – NB	0.25	78.88	0.73	32.00	37.00
Sub Sample – SVM	0.52	84.82	0.60	32.00	
Sub Sample – MLP	0.42	79.08	0.71	40.00	
Sub Sample – RF	0.43	77.07	0.68	44.00	

✓ **Experiment 5 – Feature Selection**

The aim of *Experiment 5* is to use feature selection based on the results of *Experiment 4*. NB, MLP and RF algorithms will be considered with SMOTE sampling. It was clear that Decreased Path Metric and Rank & Version attacks were detectable based on power consumption metrics. Investigating results of Rank & Blackhole and Rank & Sybil attacks, it is likely that using ML to detect specific attacks against OF0 and MRHOF based on power consumption may not be possible.

The overall aim of *Experiment 5* is to:

- Remove Rank & Blackhole and Rank & Sybil features from the dataset
- Identify highly efficient algorithms for detecting Decreased Path Metric and Rank & Version attacks based on power consumption for use in *Experiment 6*

Table 10: Experiment 5 Results

	RSME	MAPE (%)	ROC Average	Correctly Classified Instances (%)
SMOTE – NB	0.32	59.61	0.85	68.33
SMOTE – MLP	**0.12**	**44.20**	**0.91**	**75.00**
SMOTE – RF	0.26	46.28	0.91	73.33

As can be seen in Table 10 and after considering evaluation metrics, MLP performed best based on power consumption metrics post feature selection. It is noted that RF performed similarly to MLP with a larger RSME margin.

Given the confusion matrix, the following highly efficient algorithms for detecting Decreased Path Metric and Rank & Version attacks were identified: SMOTE – NB detected Decreased Path Metric attack with ROC of 0.93. SMOTE – MLP detected Decreased Path Metric and Rank & Version attacks with a ROC of 0.89 and 0.90 respectively. SMOTE – RF detected Decreased Path Metric and Rank & Version attacks with ROC of 0.89 and 0.91 respectively.

Overall, Decreased Path Metric and Rank & Version detection improved significantly based on power consumption metrics post feature selection and MLP performed best in total.

- ✓ **Experiment 6 – Classifier Ensemble**

The aim of *Experiment 6* is to exploit AdaBoost, Bagging, Stacking and Voting classifier ensemble methods to increase the likelihood of detecting Decreased Path Metric and Rank & Version attacks based on power consumption metrics.

The overall aim of this experiment is to:

- Use AdaBoost and Bagging classifier techniques to increase the likelihood of detection
- Use Stacking and Voting classifiers for classifier ensemble

As can be seen in Table 11, AdaBoost, Bagging and Stacking while utilising NB, MLP and RF classifiers, were unable to increase performance beyond what had already been captured during *Experiment 5*. Voting was established using combinations of NB, MLP, RF, and SVM.

Table 11: Experiment 6 Results

	RSME	MAPE (%)	ROC Average	Correctly Classified Instances (%)
Voting – MLP and RF	**0.24**	**45.60**	**0.92**	**84.21**
Voting – SVM and NB	0.32	53.51	0.87	69.47

A combination of MLP and RF with minimum probability selected as the combination rule provided the best results. Addressing the confusion matrix, voting with MLP and RF classifiers detected Decreased Path Metric and Rank & Version attacks with a ROC of 0.86 and 0.96 respectively. Voting with SVM and NB classifiers detected Decreased Path Metric and Rank & Version attacks with a ROC of 0.81 and 0.91 respectively. Overall, Decreased Path Metric and Rank & Version detection improved significantly based on power consumption metrics post classifier ensemble utilising Voting with MLP and RF.

- ✓ **Experiment 7 – Detecting Attacks Against MRHOF and OF0**

The aim of *Experiment 7* is to understand the success rates of detecting attacks against MRHOF and OF0. As the most successful classifier identified during experimentation, MLP will be used to assess if there is any difference in detecting attacks against the two objective functions.

The overall aim of this experiment is to:

- Assess success rate of detection for MRHOF and OF0 for network and power metrics
- Assess success rate of detection for MRHOF and OF0 for network metrics
- Assess success rate of detection for MRHOF and OF0 for power metrics post feature selection

Experiment 7 was designed to understand the success rate of detecting attacks against MRHOF and OF0. For each experiment, MRHOF and OF0 instances were removed independently of one another allowing the MLP classifier to train, test and cross validate results. MLP was used to assess if there was any variance in detecting attacks between the two objective functions. All network and power metrics were considered initially, removing MRHOF and OF0 instances independently of one another, with malicious and benign activity remaining as the selected class.

Removing MRHOF reduced instances to 178. Removing OF0 reduced instances to 293. As can be seen in Table 12, the ML model was better at detecting attacks against MRHOF than it was against OF0.

Table 12: Experiment 7 Results						
		RSME	MAPE (%)	ROC Average	Correctly Classified Instances	OF0 and MRHOF Comparison Average (%)
OF0	**MLP ALL**	0.40	54.21	0.75	55.88	49.67
	MLP NETWORK	0.39	52.85	0.80	56.86	
	MLP POWER	0.43	80.70	0.66	36.28	
MRHOF	**MLP ALL**	0.22	21.17	0.97	84.31	**71.24**
	MLP NETWORK	**0.21**	**19.34**	**0.97**	**87.25**	
	MLP POWER	0.42	77.20	0.71	42.16	

✓ **Experiment 8 – Detecting Attacks Against MRHOF and OF0 with a Balanced Class**

The aim of *Experiment 8* is to understand the success rates of detecting attacks against MRHOF and OF0 with a balanced class. As the most successful classifier identified during *Experiment 7*, MLP using network metrics will be used to assess if there is any difference in detecting attacks against the two objective functions. The overall aim of this experiment is to:

- Assess success rate of detection for MRHOF and OF0 with a balanced class for network metrics

Before each experiment, MRHOF and OF0 were balanced using SMOTE oversampling technique. Instances were then removed independently of one another allowing the MLP classifier to train, test and cross validate results.

Table 13: Experiment 8 Results					
		RSME	**MAPE (%)**	**ROC Average**	**Correctly Classified Instances (%)**
OF0	MLP NETWORK METRICS	0.38	52.82	0.75	57.84
MRHOF	MLP NETWORK METRICS	**0.21**	**19.34**	**0.97**	**87.26**

In general, *Experiment 8* was designed to understand the success rate of detecting attacks against balanced MRHOF and balanced OF0 objective functions. This experiment builds on results captured during *Experiment 7* to understand if MLP classifier is better at detecting combined attacks against MRHOF than OF0. As can be seen in Table 13, the ML model was better at detecting attacks against MRHOF than it was against OF0 despite objective function being balanced. This supports findings identified during *Experiment 7*.

VI. Discussions

In this section, the research questions stated at the beginning of this paper are addressed as follows.

- RQ1. Is there an available IoT dataset that is suitable to meet the research scope, or is the development of a novel dataset required?

(Alam et al., 2016) and (Buczak and Guven, 2015) discuss research into ML, IDS and IoT, identifying that DARPA and KDD datasets are often used since collecting, labelling and pre-processing IoT data is difficult and time consuming. A novel approach was taken to identify and develop a dataset focused on IoT features and attacks. The raw dataset that we provided in this paper included IoT features and attacks. Pre-processing, normalisation and sampling of raw data was time consuming; however, it was worthwhile. Furthermore, the novel dataset can be shared for further research in this field since correctly labelled IoT datasets are a scarce resource within the research community. The dataset contained a number of limitations that could be improved upon in future to enhance performance. Dataset limitations are discussed in the next section.

RQ1 answer summary: In this paper, a novel dataset was developed focused on IoT features and attacks.

- RQ2. What is the impact of pre-processing, normalisation, feature selection, and sampling on classifier performance?

(Yin and Gai, 2015) discuss challenges to be considered when developing an imbalanced dataset with a focus on pre- processing, normalisation, sampling and feature selection. Experiments 1, 2 and 5 were designed to understand the impact pre-processing, normalisation and feature selection have on performance. Experiments 1 – 4 were designed to understand sampling strategy.

Experiment 1 and 2 concluded that each classifier post pre-processing and normalisation improved performance by 19.29% on average based on balancing techniques for each ML algorithm. Experiments 1 – 4 concluded that SMOTE performed better than Spread Sub Sample by 46.32% on average based on balancing techniques. Experiment 5 concluded that feature selection can be used to remove IoT attacks that were not relevant to detection through power consumption metrics. Post feature selection, including the removal of Rank & Blackhole and Rank & Sybil attacks, the attack detection is increased by 29.67% based on correctly classified instances.

RQ2 answer summary: Pre-processing, normalisation, feature selection and sampling techniques are critical processes that provide significant impact on overall ML performance.

- RQ3. What is the most successful deployment of ML algorithms and classifiers?

(Haq et al., 2015) reviewed 49 related studies and discuss classifier deployments including single and ensemble methodologies. SVM is identified as the most common algorithm for IDS. When considering classifier ensemble Neutral Network and Fuzzy Logic combinations are most common. Results from Experiments 2, 3 and 5 concluded that MLP, followed closely by RF, was the most successful ML model for time series events with SVM performing worst in contrast to (Haq et al., 2015). Table 14 presents average results from experiments 2, 3 and 5 for RSME, MAPE, ROC and correctly classified instances displaying overall performance.

Overall performance of ML algorithms appeared less accurate than (Napiah et al., 2018) at 99.44% and in some instances (Nannan et al., 2018) at 86.78%, since power statistics have been included, lowering average results significantly. Including the power metric results was important since they provide an honest evaluation of the project and results can be developed upon in future research resolving limitations.

RQ3 answer summary: Classifier ensemble voting technique, using the top two performing models MLP and RF, was the most successful deployment of ML algorithms and classifiers with a ROC of 0.97.

Table 14: Overall Performance of ML Algorithm

	RSME	**MAPE (%)**	**ROC Average**	**Correctly Classified Instances (%)**	**Overall Performance of ML Algorithm**
Voting (MLP and RF)	**0.19**	**21.38**	**0.97**	**87.08**	**1**
MLP	0.14	21.83	0.96	87.07	2
RF	0.20	30.45	0.96	85.53	3
NB	0.25	34.17	0.94	79.64	4
SVM	0.32	30.09	0.83	71.89	5

- RQ4. Are ML algorithms more successful in detecting combined attacks against MRHOF or OF0?

(Airehrour et al., 2018) and (Mehta and Parma, 2018) identify a gap in research regarding IDS for combinations of IoT attacks using ML. Confusion matrices for MLP and RF were reviewed to understand what attacks were successfully detected based on network and power metrics. Reviewing network metrics, it was identified that Rank & Blackhole attacks were detected 100% of the time with no errors. Other attacks were detected successfully based on network metrics with a ROC score of 0.99 or above. Overall performance was reduced as benign activity was often incorrectly classified as an attack with a ROC score of 0.96 and precision rate of 78.15%. As indicated during the conclusion of Experiment 4, Rank & Blackhole and Rank & Sybil attacks were not successfully detected based on power metrics with true positive rates of 33.33% and 50.00%, respectively.

Reviewing confusion matrices for MLP and RF in Experiment 6, it was clear that the ensemble techniques significantly enhanced performance beyond results captured during Experiment 4 taking power metrics into consideration. Overall performance was improved from 57.84% correctly classified instances with a ROC of 0.86 to 84.21% and 0.93 respectively. Demonstrating power metrics can be used to detect a combination of IoT attacks. Decreased Path Metric and Rank & Version attacks were detected with true positive rates of 70.00% and 81.31%, respectively.

RQ4 answer summary: The ML algorithms were better at detecting attacks against equally balanced MRHOF than OF0.

VII. Conclusion, Limitations & Recommendations

This paper aims to detect IoT combined attacks of: Rank & Version, Rank & Blackhole, Decreased Path Metric, as well as Rank

& Sybil against two IoT's popular objective functions of OF0 & MRHOF using machine learning algorithms. This aim was stablished based on a comprehensive gap analysis across high quality research papers in the field. In order to successfully achieve this aim and due to lack of suitable IoT datasets, a novel dataset was developed focused on IoT's network and power features as well as IoT combined attacks. Pre-processing, normalisation, feature selection and sampling were identified as critical processes significantly impacting performance. Voting as a classifier ensemble technique, using top performing models MLP and RF, was identified as the most successful deployment of ML classifiers. Specific attacks were detected successfully based on network and power metrics; benign activity was also detected successfully and could be employed to prevent zero-day IoT attacks. The ML model was better at detecting attacks against equally balanced MRHOF than OF0. Addressing our captured results, our machine learning approach was successful in detecting all combined attacks against OF0 and MRHOF based on the network and power metrics in which MLP and RF algorithms were the most successful classifier deployment for single and ensemble models.

Although our initial aims were achieved, there were limitations in research and simulated experiments that present opportunities for future researchers to consider. Areas of the project that provide opportunities in future include the continued development of an IoT dataset, ML algorithms and classifiers, sampling, feature selection and novel MRHOF, OF0 and RPL attacks. For instance, our dataset contained four implemented combined attacks that were successfully identified using network metrics. Only two combined attacks were able to be detected using power metrics. It is recommended that further research is conducted to understand attacks that can be identified by power metrics, for instance Distributed Denial of Service (DDoS) attacks. We recommend the implementation of an IoT sensor lab in order to produce a large IoT dataset based on project limitations. Additionally, the literature review acknowledged a large range of MRHOF and OF0 attacks that could be used to meet project scope. The selected attacks were useful for detecting network metrics but provided limited success based on power metrics. It is recommended that a wider range of MRHOF and OF0 attacks are included in future datasets with a focus on those attacks that impact power metrics.

VIII. Acknowledgment

This research is funded by the School of Computing at Edinburgh Napier University.

References

[1] Airehrour, D., Gutierrez, J. A. and Ray, S. K. (2017) 'A Trust-Aware RPL Routing Protocol to Detect Blackhole and Selective Forwarding Attacks', Australian Journal of Telecommunications and the Digital Economy. doi: 10.18080/ajtde.v5n1.88.

[2] Alam, F. et al. (2016) 'Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)', in Procedia Computer Science. doi: 10.1016/j.procs.2016.09.068.

[3] Anthéa Mayzaud, Anuj Sehgal, Rémi Badonnel, Isabelle Chrisment, J. S. (2016) 'Using the RPL Protocol for Supporting Passive Monitoring in the Internet of Things', in. Istanbul, Turkey: IEEE/IFIP Network Operations and Management Symposium.

[4] Atzori, L., Iera, A. and Morabito, G. (2017) 'Understanding the Internet of Things: definition, potentials, and societal role of a fast evolving paradigm', Ad Hoc Networks. doi: 10.1016/j.adhoc.2016.12.004.

[5] Belgiu, M. and Drăgu, L. (2016) 'Random forest in remote sensing: A review of applications and future directions', ISPRS Journal of Photogrammetry and Remote Sensing. doi: 10.1016/j.isprsjprs.2016.01.011.
[6] Borton, T. (2014) 'Reach touch and teach', YNEDT. doi: 10.1016/j.nedt.2013.11.003.

[7] Buczak, A. and Guven, E. (2015) 'A survey of data mining and machine learning methods for cyber security intrusion detection', IEEE Communications Surveys & Tutorials. doi: 10.1109/COMST.2015.2494502.

[8] Airehrour, D., Gutierrez, J. A., & Ray, S. K. (2018). SecTrust-RPL: A secure trust-aware RPL routing protocol for Internet of Things. Future Generation Computer Systems, 93, 860-876. doi:10.1016/j.future.2018.03.021

[9] Haixiang, G. et al. (2017) 'Learning from class-imbalanced data: Review of methods and applications', Expert Systems with Applications. doi: 10.1016/j.eswa.2016.12.035.

[10] Haq, N. F. (2015) 'Application of Machine Learning Approaches in Intrusion Detection System: A Survey', International Journal of Advanced Research in Artificial Intelligence.

[11] Kanchymalay, K. et al. (2017) 'Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques', in IOP Conference Series: Materials Science and Engineering. doi: 10.1088/1757-899X/226/1/012117.

[12] Le, A. et al. (2016) 'A specification-based IDS for detecting attacks on RPL-based network topology', Information (Switzerland). doi: 10.3390/info7020025.

[13] Lee, T.H., Wen, C.H., Chang, L.H., Chiang, H.S. and Hsieh, M. . (2014) 'A Lightweight Intrusion Detection Scheme Based on Energy Consumption Analysis in 6LowPAN', in Advanced Technologies, Embedded and Multimedia for Human-centric Computing. Dordrecht: Springer, pp. 1205–1213.

[14] Liang, D. et al. (2018) 'A novel classifier ensemble approach for financial distress prediction', Knowledge and Information Systems. doi: 10.1007/s10115-017-1061-1.

[15] Linn, K. A. et al. (2016) ‘Control-group feature normalization for multivariate pattern analysis of structural MRI data using the support vector machine’, NeuroImage. doi: 10.1016/j.neuroimage.2016.02.044.

[16] Lucas, P. (2012) ‘Critical reflection. What do we really mean?’, Proceedings of the 2012 Australian Collaborative Education Network National Conference.

[17] Mayzaud, A., Badonnel, R. and Chrisment, I. (2016) ‘A taxonomy of attacks in RPL-based internet of things’, International Journal of Network Security.

[18] Mehdiyev, N. et al. (2016) ‘Evaluating Forecasting Methods by Considering Different Accuracy Measures’, in Procedia Computer Science. doi: 10.1016/j.procs.2016.09.332.

[19] Mehta, R. and Parma, M. (2018) ‘Security Attacks and Countermeasures in RPL for Internet of Things’, in 11th International Conference on Recent Innovations in Science, Engineering and Management. Pune, pp. 55–70.

[20] Nannan, L. et al. (2018) ‘Intrusion Detection System Based on Evolving Rules for Wireless Sensor Networks’, Journal of Sensors, 2018(146), p. 8.

[21] Napiah, M. N. et al. (2018) ‘Compression Header Analyzer Intrusion Detection System (CHA - IDS) for 6LoWPAN Communication Protocol’, IEEE Access. doi: 10.1109/ACCESS.2018.2798626.

[22] Pham, B. T. T. et al. (2016) ‘Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS’, Journal of Geomatics.

[23] Ramírez-Gallego, S. et al. (2017) ‘A survey on data preprocessing for data stream mining: Current status and future directions’, Neurocomputing. doi: 10.1016/j.neucom.2017.01.078.

[24] Rehman, A. et al. (2016) ‘Rank attack using objective function in RPL for low power and lossy networks’, in 2016 International Conference on Industrial Informatics and Computer Systems, CIICS 2016. doi: 10.1109/ICCSII.2016.7462418.

[25] Shafique, U. and Qaiser, H. (2014) ‘A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)’, International Journal of Innovation and Scientific Research ISSN.

[26] Sharma, D., Mishra, I. and Jain, S. (2017) ‘A Detailed Classification of Routing Attacks against RPL in Internet of Things’, International Journal of Advance Research Ideas and Innovations in Technology.

[27] Sheikhan, M. and Bostani, H. (2017) ‘A Security Mechanism for Detecting Intrusions in Internet of Things Using Selected Features Based on MI-BGSA’, International Journal of Information and Computer Technology Research, 9(2), pp. 53–62.

[28] Sousa, N. et al. (2017) ‘ERAOF: A new RPL protocol objective function for Internet of Things applications’, 2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech).

[29] Svinicki, M. D. and Dixon, N. M. (1987) ‘The Kolb Model Modified for Classroom Activities’, College Teaching. doi: 10.1080/87567555.1987.9925469.

[30] Tang, B., Kay, S. and He, H. (2016) ‘Toward Optimal Feature Selection in Naive Bayes for Text Categorization’, IEEE Transactions on Knowledge and Data Engineering. doi: 10.1109/TKDE.2016.2563436.

[31] Yin, H. and Gai, K. (2015) ‘An Empirical Study on Preprocessing High-dimensional Class imbalanced Data for Classification’, in 2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), and 2015 IEEE 12th International Conf on Embedded Software and Systems (ICESS). IEEE, pp. 1314–1319.

[32] Zarpelão, B. B. et al. (2017) ‘A survey of intrusion detection in Internet of Things’, Journal of Network and Computer Applications. doi: 10.1016/j.jnca.2017.02.009.

[33] Weka; Retrieved from https://www.cs.waikato.ac.nz/ml/weka/; last accessed: 10 June 2019.