informatics

MDPI

Article

Uncovering the Limitations and Insights of Packet Status Prediction Models in IEEE 802.15.4-Based Wireless Networks and Insights from Data Science

Mariana Ávalos-Arce [1], Heráclito Pérez-Díaz [1], Carolina Del-Valle-Soto [1,*] and Ramon A. Briseño [2]

[1] Facultad de Ingeniería, Universidad Panamericana, Álvaro del Portillo 49, Zapopan 45010, JA, Mexico; 0197495@up.edu.mx (M.Á.-A.); 0190575@up.edu.mx (H.P.-D.)
[2] Centro Universitario de Ciencias Económico Administrativas, Universidad de Guadalajara, Zapopan 45180, JA, Mexico; alejandro.bmartinez@alumnos.udg.mx
* Correspondence: cvalle@up.edu.mx; Tel.: +52-33-1368-2200

Abstract: Wireless networks play a pivotal role in various domains, including industrial automation, autonomous vehicles, robotics, and mobile sensor networks. This research investigates the critical issue of packet loss in modern wireless networks and aims to identify the conditions within a network's environment that lead to such losses. We propose a packet status prediction model for data packets that travel through a wireless network based on the IEEE 802.15.4 standard and are exposed to five different types of interference in a controlled experimentation environment. The proposed model focuses on the packetization process and its impact on network robustness. This study explores the challenges posed by packet loss, particularly in the context of interference, and puts forth the hypothesis that specific environmental conditions are linked to packet loss occurrences. The contribution of this work lies in advancing our understanding of the conditions leading to packet loss in wireless networks. Data are retrieved with a single CC2531 USB Dongle Packet Sniffer, whose pieces of information on packets become the features of each packet from which the classifier model will gather the training data with the aim of predicting whether a packet will unsuccessfully arrive at its destination. We found that interference causes more packet loss than that caused by various devices using a WiFi communication protocol simultaneously. In addition, we found that the most important predictors are network strength and packet size; low network strength tends to lead to more packet loss, especially for larger packets. This study contributes to the ongoing efforts to predict and mitigate packet loss, emphasizing the need for adaptive models in dynamic wireless environments.

Keywords: packet loss; packet sniffer data; binary classification interference; wireless communications

Citation: Ávalos-Arce, M.; Pérez-Díaz, H.; Del-Valle-Soto, C.; Briseño, R.A. Uncovering the Limitations and Insights of Packet Status Prediction Models in IEEE 802.15.4-Based Wireless Networks and Insights from Data Science. *Informatics* **2024**, *11*, 7. https://doi.org/10.3390/informatics11010007

Academic Editor: Alessandro Pozzebon

Received: 7 October 2023
Revised: 12 January 2024
Accepted: 18 January 2024
Published: 26 January 2024

1. Introduction

Currently, wireless networks have evolved to become an integral part of our daily lives, finding applications across diverse domains such as industrial automation, autonomous vehicles, remote surgery, robotics, mobile sensor networks, and internet-based applications [1]. These networks facilitate seamless communication by connecting two or more nodes, enabling the exchange of information without the need for physical links or connector cables. Nowadays, wireless networks are used in industrial automation, vehicles, remote surgery, robots, mobile sensor networks, and internet-based applications. A wireless network is then a connection of two or more nodes where information is sent or received (or both) without the use of any physical link (i.e., connector cables) between them.

A packet is a small segment of a larger message. Data sent over a network are divided into packets, which are later recombined by the destination node into the originally sent file [2]. Theoretically, it could be possible to send data over a network without chopping them down into small packets of information. However, such an approach becomes

impractical when more than two nodes are involved in the network: whenever any long line of bits is passed over two nodes (one sender and one receiver), the other nodes need to wait for this communication to finish so they can make use of said channel [3]. In other words, packet division of data makes a network able to exchange billions of files simultaneously instead of just a few. It also means that it is possible for packets to take multiple paths through the network and reach the same destination, as long as all of them arrive successfully at their destination.

In order to determine the traveling process of the packets, we need to consider the network's topology, which tells us the way that nodes are physically (and logically) placed and the connection rules of the network; each topology solves different connection problems, and because of that, it needs different communication rules. Two different roles are usually present in these communications, that is, the sender and the receiver, and network nodes can play one or both of them, once again depending on the network's topology.

IEEE 802.15.4 serves as a crucial standard delineating the physical layer and medium access control for low-rate wireless personal area networks (LR-WPANs) [4]. This standard plays a foundational role in shaping the ZigBee specification, aiming to provide a comprehensive solution for networks by constructing upper-level protocol stack layers not covered by the standard. The primary objective of IEEE 802.15.4 is to define basic network layers catering to a specific type of wireless personal area network (WPAN), facilitating communication among ubiquitous, cost-effective, and low-speed devices. Emphasizing economical communication with nearby nodes and minimal to no infrastructure, the standard prioritizes energy efficiency. In its fundamental form, it envisions a communication range of 10 m with a transfer rate of 250 kbps. Multiple physical layers are defined to accommodate varying requirements, initially introducing alternative rates of 20 and 40 kbps, and the current version incorporates an additional rate of 100 kbps. Lower rates can be achieved, further reducing energy consumption. Notably, the key feature of 802.15.4 within WPANs is the attainment of exceptionally low manufacturing costs through technological simplicity without sacrificing generality or adaptability.

The application of AI for packet interference detection and collision analysis in small-scale wireless networks holds significant importance in optimizing network performance and reliability. In the realm of modern wireless communication, where small-scale networks are prevalent, the accurate identification of interference and potential collisions is paramount for maintaining seamless and efficient data transmission [5]. This AI-driven analysis contributes by providing a sophisticated and automated means to discern interference patterns, predict potential collisions, and enhance the overall robustness of wireless networks. By leveraging AI techniques, this study not only addresses the challenges posed by packet interference but also represents a progressive step towards developing intelligent solutions for improving the reliability and performance of small-scale wireless communication systems. The management of information as independent pieces that travel in any order involves multiple problems when in practice: packets may collide between nodes or get lost in the process of arriving at their destination, and this is called packet loss. Intermittent data packet losses and network-induced time delay are known to be two of the main causes for the performance deterioration or even instability of any controlled networked system.

1.1. Hypothesis

Due to the nature of shared communication paths in wireless networks, as described above, the study of packet loss has garnered considerable research interest. With the purpose of further investigating this problem and aiming to provide new insights into the conditions under which packet loss occurs, the hypothesis of this study is as follows:

> There exists a condition or set of conditions in a network environment that, when present, lead to packet loss.

In the following sections, a description of the methodology and experimentation applied to a testing network environment will be presented, along with the prediction model for the Packet Status (OK or ERROR) that was trained using the collected data.

1.2. Contribution

This work emphasizes the critical role of packetization in modern wireless networks, highlights the challenges posed by packet loss, particularly in the context of interference, and introduces the central hypothesis driving this research—that specific conditions within a network's environment are linked to packet loss occurrences. The subsequent sections delve into the methodology, experimentation, and data analysis that form the foundation of this study, aiming to advance our understanding of the conditions leading to packet loss in wireless networks.

The contribution of this work lies in highlighting the critical role of packetization in modern wireless networks and emphasizing the challenges posed by packet loss, especially in the context of interference. This study aims to provide new insights into the conditions under which packet loss occurs. It also delves into the methodology, experimentation, and data analysis that form the foundation of the research, contributing to our understanding of the conditions leading to packet loss in wireless networks.

The tests included are as follows:

- Control Test 1: Baseline test with no interference.
- Control Test 2: Another instance of the baseline test.
- Double Network Test: Involving interference from another nearby network.
- Radio Frequencies Test: Involving interference from common-use machines producing radio frequencies.
- No Restriction Test: Involving interference from people using personal devices without restrictions.
- Wind Test: Involving interference from fans blowing air across the space.
- Wireless Test: Involving interference from devices using WiFi protocol.

2. Related Work

In recent years, the proliferation of electronic and electrical devices has introduced new challenges in maintaining the reliability of wireless networks based on the IEEE 802.15.4 standard. Packet loss can be particularly problematic in scenarios involving interference from neighboring networks [6], as such interference can disrupt the seamless flow of data packets. Identifying the specific conditions or sets of conditions within a network's environment that lead to packet loss is a critical research endeavor. Packet loss, a critical performance metric, becomes increasingly unpredictable as the interference from colocated devices grows. Researchers have been actively investigating methods to predict and mitigate packet loss under such dynamic conditions. Notably, Mindo et al. presented an innovative approach using machine learning techniques to predict packet loss in the presence of electronic devices in their work titled "Machine Learning-Based Prediction of Packet Loss in IEEE 802.15.4 Networks with Device Interference" [7]. Their study leveraged historical network data and achieved promising results in predicting packet loss probabilities, contributing to the understanding of how electronic devices impact network reliability.

Another significant work focuses on modeling the behavior of wireless networks with IEEE 802.15.4 devices in the presence of electrical interference. In their work, Sun et al. explored the impact of electrical noise on packet loss in "Characterizing the Effects of Electrical Noise on IEEE 802.15.4 Wireless Networks" [8]. They conducted comprehensive experiments to measure the effect of various electrical devices on packet loss rates. Their findings highlighted the need for accurate models to predict packet loss, especially in scenarios where sensitive applications, such as healthcare monitoring or industrial automation, rely on wireless communication.

Machine learning and statistical approaches have become prominent in developing predictive models for packet loss in IEEE 802.15.4 wireless networks. Ayeesha et al. in-

troduced a predictive model based on a long short-term memory (LSTM) neural network in their research paper "LSTM-Based Packet Loss Prediction for IEEE 802.15.4 Wireless Sensor Networks in Electrical Environments" [9]. This LSTM-based model demonstrated remarkable accuracy in forecasting packet loss in the presence of electrical interference. Their work exemplifies the growing interest in harnessing deep learning techniques to tackle the packet loss prediction challenge in complex wireless environments. As wireless networks based on the IEEE 802.15.4 standard continue to evolve, researchers are also exploring the potential of reinforcement learning (RL) algorithms in addressing packet loss issues. Recent work by Gonzalez et al. in "Reinforcement Learning-Based Packet Loss Mitigation in IEEE 802.15.4 Networks with Device Interference" [10] introduced an RL framework to adaptively optimize network parameters to minimize packet loss. Their study demonstrated the adaptability of RL in dynamically managing network resources and mitigating the impact of colocated electronic and electrical devices on packet loss. Furthermore, as wireless networks based on the IEEE 802.15.4 standard continue to advance, researchers are exploring diverse approaches to address the challenge of packet loss. In addition to LSTM-based models, recent work by Raza et al. [11] delved into the application of reinforcement learning algorithms for mitigating packet loss in IEEE 802.15.4 networks with device interference. Their study introduces an RL framework designed to adaptively optimize network parameters, showcasing the potential of RL in dynamically managing resources to minimize the impact of colocated electronic and electrical devices on packet loss. This research expands the repertoire of strategies available for enhancing the reliability of wireless networks in the face of evolving technologies and network conditions.

The effective operation of wireless networks is grounded in a fundamental principle: packetization. In contemporary networks, information is segmented into smaller units known as packets, which are transmitted individually throughout the network. This strategy improves network scalability by enabling the simultaneous exchange of multiple files [12]. Moreover, it introduces the intriguing possibility of packets taking multiple paths through the network to reach their destination, thereby enhancing network robustness. As Wagner et al. pointed out in their seminal work, this packet division strategy enables billions of files to be transmitted simultaneously, a feat that would be practically impossible with a nonpacketized approach [13].

Notably, recent work by Bennis et al. [14] delves into the impact of interference from neighboring networks on packet loss rates. Their findings provide valuable insights into mitigating packet loss in crowded wireless environments. Moreover, Shah et al. delved into the impact of interference from neighboring networks on packet loss rates [15]. Their findings provide valuable insights into mitigating packet loss in crowded wireless environments. Furthermore, Freschi et al. made significant contributions to understanding the impact of interference from neighboring networks on packet loss rates [16]. Their recent work sheds light on innovative strategies for mitigating packet loss in densely populated wireless environments, offering additional perspectives. These studies contribute to a more comprehensive understanding of the challenges and potential solutions associated with packet loss in crowded wireless scenarios.

Packetization enhances network scalability as it allows multiple files to be exchanged concurrently [17]. Moreover, it introduces the intriguing possibility of packets taking multiple paths through the network to reach their destination, thereby enhancing network robustness. As Tuker et al. pointed out in their seminal work, this packet division strategy enables billions of files to be transmitted simultaneously, a feat that would be practically impossible with a nonpacketized approach [18].

Table 1 provides an overview of recent research efforts related to the prediction of packet loss in various experiment environments. As discussed earlier, packet loss is a critical issue in networked systems that can lead to performance deterioration and instability. The table lists different references to research papers, each focusing on specific experiment environments, such as wireless sensor networks, IoT devices, 5G networks, smart grids, edge computing, and mobile ad hoc networks. These studies employ diverse

techniques for packet loss detection, including machine learning, statistical analysis, deep learning, time series analysis, probabilistic models, and network simulation. However, it is important to note that each model has its limitations, which are also highlighted in the table. These limitations may include being specific to certain sensor types, assuming stable network conditions, having high computational complexity, requiring precise timestamp synchronization, having limited historical data availability, or not being suitable for real-time prediction. This table serves as a valuable reference for researchers and practitioners working to address packet loss issues in different networked environments by offering insights into the current state of research and the associated challenges and limitations. The occurrence of packet loss stands as a pivotal concern in networked systems, capable of instigating performance degradation and system instability [19]. The compilation includes diverse references to research papers, each centering on specific experimental contexts such as wireless sensor networks, IoT devices, 5G networks, smart grids, edge computing, and mobile ad hoc networks. These investigations employ a spectrum of techniques for packet loss detection, encompassing machine learning, statistical analysis, deep learning, time series analysis, probabilistic models, and network simulation [20–22].

Table 1. Current literature on packet loss prediction models.

Reference	Experiment Environment	Detection Loss Technique	Model Limitations
[23]	Wireless sensor network	Machine learning	Limited to specific sensor types.
[24]	IoT devices	Statistical analysis	Assumes stable network conditions.
[25]	5G network	Deep learning	High computational complexity.
[26]	Smart grids	Time series analysis	Requires precise timestamp synchronization.
[27]	Edge computing	Probabilistic models	Limited historical data availability.
[28]	Mobile ad hoc network	Network simulation	Not suitable for real-time prediction.

3. Materials and Methods

The experimentation involved various controlled-environment tests, each representing different interference scenarios, including a baseline test with no interference, interference from neighboring networks, radio-frequency interference, unrestricted device usage, wind interference, and wireless interference. This study captured and analyzed packet data, considering factors such as frame type, length, RSSI, and time.

To determine the dominant machine learning algorithm, the research applied several classification algorithms, including naive Bayes, SVM, neural network, logistic regression, random forest, gradient boosting, and decision tree. The models were trained and evaluated based on precision, recall, F1 score, and confusion matrices, with a focus on predicting packet loss (error class) in the presence of WiFi interference.

The results indicate that the gradient boosting and random forest models achieved the best performance, with precision rates of 97% and 96%, respectively. However, due to the dataset's imbalance, additional metrics such as F1 score and recall for the error class were considered. The confusion matrices provided insights into the models' tendencies to predict both correct and incorrect outcomes.

The network installation was completed using Texas Instruments' Evaluation Boards of models LAUNCHXL-CC1352R1 and LP-CC2652RB as nodes of the network, as observed in Figure 1. Both models are evaluation boards that support wireless connections with

protocols based on the IEEE 802.15.4 standard and a built-in temperature sensor [29]. The network had two types of nodes, the collector and sensor node, applying the simplest form of master–worker architecture vastly used in controlled network systems nowadays, which implements a star network topology (allowing a unique collector to whom multiple sensors are linked). The collector node was a CC26x Evaluation Board (1), and the sensor nodes were a mixture of CC26x and CC13x Evaluation Boards (8).

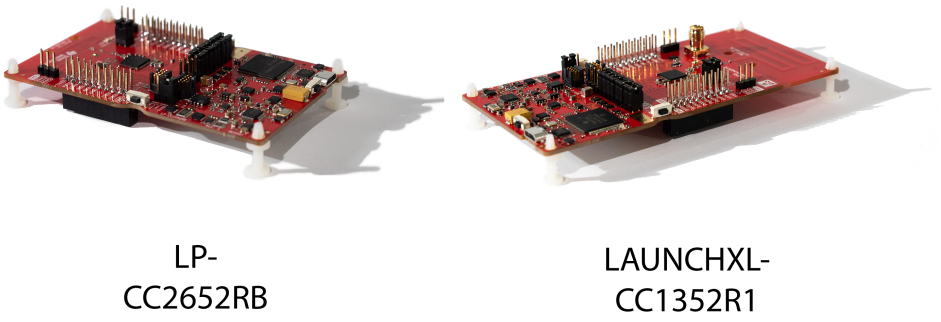

Figure 1. Models LAUNCHXL-CC1352R1 and LP-CC2652RB.

The nodes implemented a star network topology, as observed in Figure 2, since all sensor nodes were in charge of periodically sending the environment temperature in degrees Celsius. The collector node was then responsible for managing the orchestration of the sensor nodes' requests and monitoring the number of orphan nodes that may be disconnected from the network.

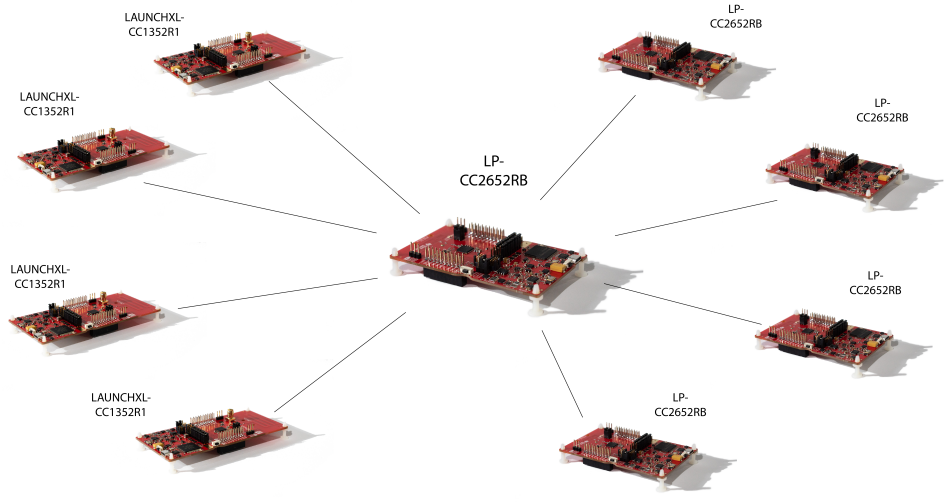

Figure 2. Star network topology with LAUNCHXL-CC1352R1 and LP-CC2652RB.

In order to capture packet data across the network, a sniffer board was used. This study employed a Texas Instruments' CC2531 USB Dongle, which is based on the IEEE 802.15.4 wireless standard [30].

3.1. Experimentation

The nodes of the network were placed in a closed environment consisting of a 10 m × 30 m rectangular closed room. The experimentation involved 7 controlled-environment tests, each consisting of variations in environmental conditions and the presence or absence of interference, with an individual duration of one hour. This means that the following tests were each performed over a 60 min time period, during which the network nodes and sniffer board worked simultaneously to send temperature data to the collector and capture the packet information sent, respectively. The tests are as follows:

- Control Test 1: This is the control or baseline network data for the experiments. In this test there were no interference or barrier conditions at all to stress the network so that it represented the ideal conditions of the system.
- Control Test 2: this is another instance of the previous test type, captured right after to later be averaged with the first control test and form a single Control Test.
- Double Network Test: this involved another star network nearby the analyzed one, becoming a test for IEEE 802.15.4 network interference.
- Radio Frequencies Test: this involved five common-use machines that produce radio frequencies of different kinds inside a household: a microwave, an air fryer, a blender, a toaster, and a sandwich maker were working at the same time and place that the network in question was operating in.
- No Restriction Test: this involved four people navigating the room and making use of their personal devices (cellphone, laptop) without restrictions, with the purpose being that this test would represent the network and interference traffic in an average household.
- Wind Test: this involved two fans blowing air across the space where the collector node was installed at a speed of 20 km/h.
- Wireless Test: this involved different machines that used specifically WiFi network protocol, which were 1 SmartTV, 3 smartphones, 1 game console, and 2 personal computers, all downloading content from the internet in the same room and at the same time the network was operating.

While each of the tests was being performed, the sniffer board captured, in real time, the information of traveling packets during traffic. Once each individual test was finished, the captured data were stored in a file format named packet sniffer data (PSD). The PSD files were processed and translated into a single tabular file from which the prediction model gathered its training data.

3.2. Measured Data Description

The wireless protocol IEEE 802.15.4 standard suggests [31–33] that every packet contains its Frame Control Type inside the packet's payload byte block. This frame type takes eight different values [34] in the captured data for the present study: 1_OCT_HEADER (1 Octet MAC Header for Low Latency) [34], CSL_WAKEUP, CSL_SECURE_ACK, and RFID_BLINK, all of which are coded to fit in 3 bits inside the payload byte block [32,35] as mentioned. The frame type is crucial since it refers to the type of command that sent a packet [34], which in turn specifies if the sender is either the collector/master (Beacon) or one of the sensor nodes.

There is a section of the packet's bytes consisting of two bytes called Status Bytes destined to store either 0 (ERROR) or 1 (OK) [32]. For all packets captured, this number becomes the class of the packet in the present study: OK (1), when a packet was captured as it is meant to be received, and ERROR (0), when a packet is captured after a collision or fragmentation that causes the loss of a packet in the network. Thus, given the classification of every packet captured, a supervised prediction model can be trained. Table 2 shows the dataset features description.

However, first, we assessed describing how the signals happen throughout time, and thus a time series plot of the packets per frame type lets the reader visualize the experiment's trends. For that, a time series using the packet number of every captured packet and the packet's frame type was used. The average number of packets through time per frame control field (FCF) type is shown in Figure 3.

Table 2. Description of Dataset.

Variable	Description	Values
TIME(MS)	Packet time in the network	0 to 16751.18238 milliseconds
LENGTH	Length of the package	1 to 109 bytes
RSSI	Network signal strength	−199 to 53 decibels
Status bytes	Package reception status	0 Error (1.95)% 1 OK (98.05)%
DATA	Data frames	0 not a DATA frame type (79.01%) 1 DATA frame type (20.99%)
CMD	Command frames	0 not a CMD frame type (71.28%) 1 CMD frame type (28.72%)
ACK	Acknowledgment frames	0 not an ACK frame type (50.91%) 1 ACK frame type (49.09%)
BCN	Beacon frames	0 not a BCN frame type (99.17%) 1 BCN frame type (0.83%)
1_OCT_HEADER	Frames with a one-octet header	0 not a 1_OCT_HEADER frame type (99.79%) 1 1_OCT_HEADER frame type (0.21%)
CSL_WAKEUP	Wake-up frames for the Listen Unit (CSL)	0 not a CSL_WAKEUP frame type (99.97%) 1 CSL_WAKEUP frame type (0.03%)
CSL_SECURE_ACK	Secure acknowledgment frames for the Listen Unit (CSL)	0 not a CSL_SECURE_ACK frame type (99.95%) 1 CSL_SECURE_ACK frame type (0.05%)
RFID_BLINK	Radio-Frequency Identification (RFID) frames	0 not an RFID_BLINK frame type (99.93%) 1 RFID_BLINK frame type (0.07%)
CONTROL	CONTROL frame test type	0 not a CONTROL frame test type (81.79%) 1 CONTROL frame test type (18.21%)
DOUBLE NETWORK	DOUBLE NETWORK frame test type	0 not a DOUBLE NETWORK frame test type (79.97%) 1 DOUBLE NETWORK frame test type (−299.51%)
ELECTRO	Radio-frequency frame test type	0 not a Radio-frequencies frame test type (82.24%) 1 Radio-frequencies frame test type (17.76%)
NORMAL	No restriction frame test type	0 not a No restriction frame test type (84.06%) 1 No restriction frame test type (15.94%)
WIND	WIND frame test type	0 not a WIND frame test type (85.34%) 1 WIND frame test type (14.66%)
WIRELESS	WIRELESS frame test type	0 not a WIRELESS frame test type (86.60%) 1 WIRELESS frame test type (13.40%)

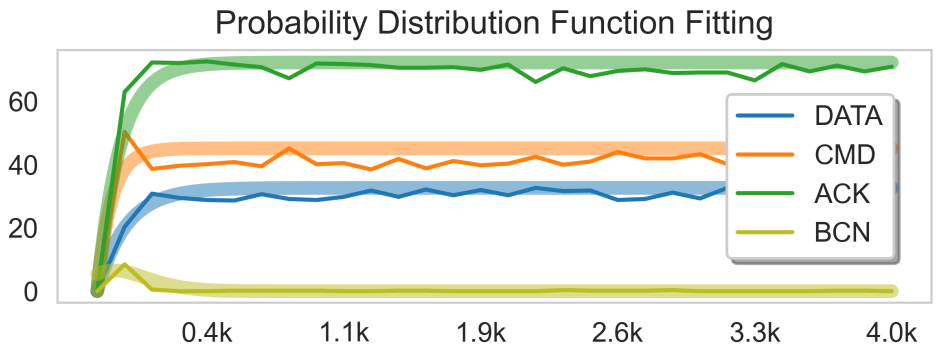

Figure 3. Average number of packets' arrival (y axis) through time (x axis) per frame type, and its corresponding parametric distribution fitting.

3.3. Measured Data Analysis

In Figure 3, a parametric distribution fitting is presented, showcasing the tendency of the tests performed: interarrival events are the best example of an event that occurs with periodicity following an exponential distribution, and such a distribution was used for modeling DATA, CMD, and ACK packet arrival behavior, as those packets are continuously present in the network. BCN frame type packets, representing the Beacon or collector node, only appear at the beginning of the network pipeline, and thus, their arrival periodicity follows a Poisson distribution. The exact analytical functions are expressed below for DATA, CMD, ACK, and BCN frames, respectively.

$$f_{data}(x) = 1 - e^{-\frac{1}{120}x} \quad (1)$$

$$f_{cmd}(x) = 1 - e^{-\frac{1}{75}x} \quad (2)$$

$$f_{ack}(x) = 1 - e^{-\frac{1}{150}x} \tag{3}$$

$$f_{bcn}(x) = \frac{1.15^x}{x!}e^{-1.15} \tag{4}$$

The remaining frame types, namely, 1_OCT_HEADER, CSL_WAKEUP, CSL_SECURE_ACK, and RFID_BLINK, were not included in this description since they were barely present in the experiments to have a mathematically defined behavior.

To decide which model to use for the classifier, all columns for each captured packet had to be analyzed individually. Most of the columns in the dataset of all captured packets are binary data (except for RSSI, Time, and Packet Number), explained by the fact that the packet sniffer data (PSD) format stores data in bytes [32], and therefore, most fields in the dataset are binary or Boolean representatives. Thus, the only continuous fields in the dataset are the LENGTH, RSSI, and TIME(MS) columns. Consequently, a box plot was created for each continuous field, but before doing so, the fields were grouped by TEST_TYPE (interference type) to produce Figure 4.

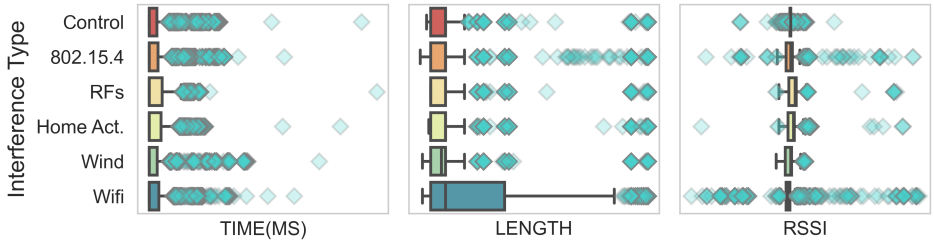

Figure 4. Box plot of the continuous fields in every captured packet per test type.

From Figure 4, we can conclude that WiFi interference generated the largest fluctuations in all three fields. This is reflected by a larger difference in the range of the box and the outliers concerning the ranges of the other experiments across the three fields. WiFi interference was also the test type that showed the largest absolute linear correlation to the target field (-0.134) out of all types of interference in this study. The linear correlation between the type of interference with respect to the target field, which is the Packet Status, is shown in Table 3. From the correlation values, there is a tendency when there is the use of WiFi access devices, which is present in the Home Activities and WiFi tests: the correlation with respect to the Packet Status value appears negative in the presence of WiFi interference, suggesting a weak negative association.

Table 3. Linear correlation between interference and packet status.

Test	Control	802.15.4	RFs	Home Activities	Wind	WiFi
Corr.	0.053	0.023	0.043	−0.027	0.027	−0.134

4. Results and Discussion

We created a predictive model with the aim of explaining the situations that cause packet loss in the WiFi interference scenario. We chose the WiFi interference data because, according to the previous section, it is the type of interference with a major correlation with the target field (packet status) and consequently a major predictive tendency. In that way, we took LENGTH, RSSI, TIME (MS), and frame type as predictors and the status byte (OK or error) as the target variable. For the frame type, we only took the DATA, CMD, ACK, and BCN frames. It is important to note that the not-included frame type 1_*OCT_HEADER* stands for 1 Octal Header Low Latency [34], and it has a linear correlation (0.44) to the target class, but it is a trivial result since low latency packets are sent when the conditions of the network force a low response from the devices, and, therefore, packet loss is expected.

Using the data mining Orange software, we connected seven classification algorithms (naive Bayes, SVM, neural network, logistic regression, random forest, gradient boosting,

and tree) to compare their performance for the packet loss prediction task. Furthermore, we split the dataset into training (80%) and test (20%). Figure 5 shows the visual programming of the prediction algorithms in the Orange software. The best results were obtained by gradient boosting and random forest with a precision of 91% and 81%, respectively. However, since the dataset is unbalanced, with many more packets that reached their destination than lost packets, it is important to look at the F1 and recall in the Table 4 metrics for the error class and the confusion matrices of each model. F1 and recall are metrics with sensitivity to minority classes, balancing precision with completeness. On the other hand, the advantage of computing the confusion matrix is the fact that we evaluate the model's tendency to predict mistakenly and use that evaluation in combination with the evaluation of the model's tendency to predict correctly, with the objective of computing the AUC metric, which represents how the binary classification by the model tends to give correct answers for both classes. Table 5 shows the confusion matrix computed for the gradient boosting model, and Table 6 shows the confusion matrix for the random forest.

Table 4. Model metrics.

Model	F1	Precision	Recall
Gradient boosting	72	91	59
Random forest	67	81	57

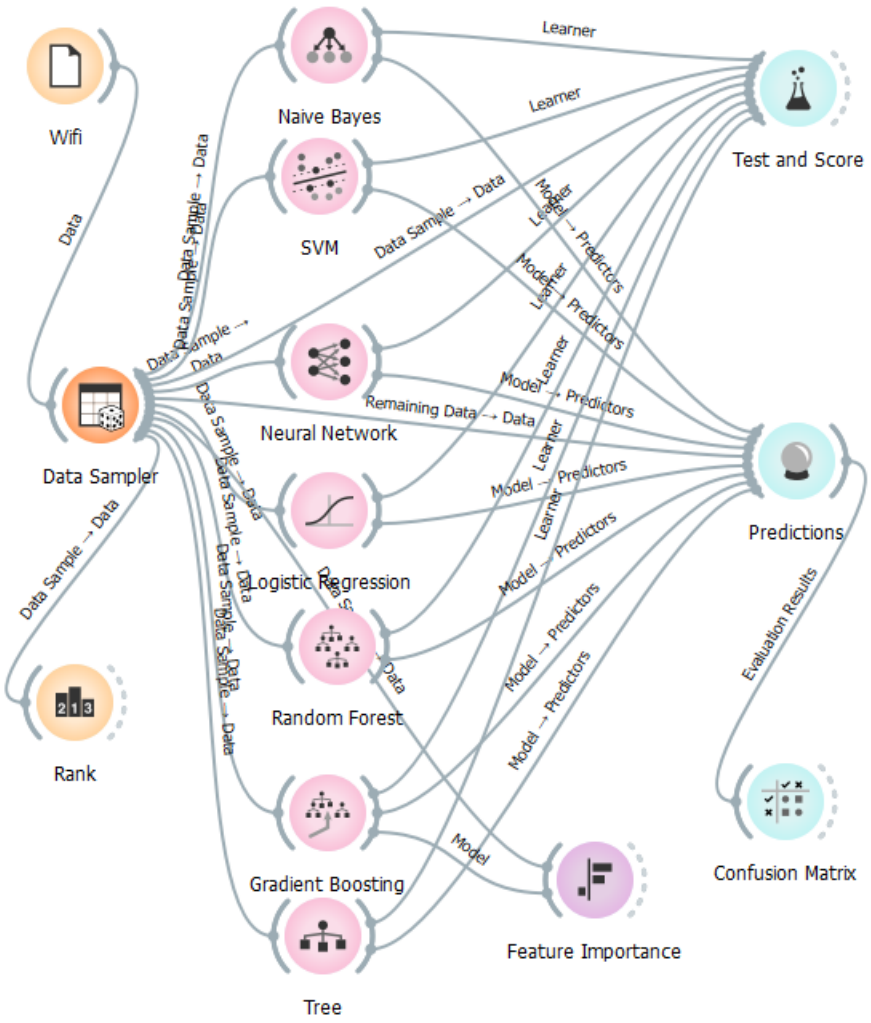

Figure 5. Visual programming of the prediction algorithms in Orange software.

Table 5. Confusion matrix for gradient boosting classifier.

	Predicted	
Real ↓	**1**	**0**
1	TP = 823	FN = 3
0	FP = 22	TN = 32

Table 6. Confusion matrix for random forest classifier.

	Predicted	
Real ↓	**1**	**0**
1	TP = 819	FN = 7
0	FP = 23	TN = 31

The total number of packets with class 1 (OK) and 0 (error) was 4142 and 259, respectively. The confusion matrices provided for the test split (880 records) show a high TPR (true positive rate) and almost a midvalued FPR (false positive rate), which is the desired behavior for a failure-related model: false positives are more dangerous to studies that rely on packet status than false negatives. It is possible to see that the F1 and recall of both models are above 50%. This is confirmed in the confusion matrix as the models correctly predicted 60% of the error class, very good metrics for a class that is present only in 6.1% of the data.

Finally, we observed the random forest trees formed in the training process looking for patterns. We calculated the correlation of the predictors with the target variable with six different coefficients (info gain, gain ratio, Gini, chi-squared, relief, and FCBF) using the rank Orange widget. Additionally, we calculated and interpreted the important features for the gradient boosting and random forests models with the feature importance Orange widget. In the decision trees created by the random forest algorithm, as well as in the correlation coefficients and the important features of the gradient boosting model, the most important variable for predicting packet loss is RSSI, followed by LENGTH. Furthermore, in decision trees formed by the random forest model, it can be observed that packet loss decisions are oriented toward when the network is perceived as weak and the length of the packets is large. This can mean that the use of different devices using the same WiFi network tends to cause interference between them and in some occasions weakens the network.

On the other hand, the time it takes for a packet to be perceived as lost or received, and the type of ACK frame, despite not being among the variables with the highest correlation with the target variable, are important variables for the prediction of gradient boosting and random forest models. It seems logical to think that the longer it takes for a packet to be perceived as received or lost, the more likely it is that no node will receive the packet. Similarly, the type of ACK frame has the highest presence in the network; it could be that the preparation interval for communication between devices is the period of loss of this type of frame.

5. Conclusions and Future Work

This research presents a packet status prediction model for data packets in wireless networks based on the IEEE 802.15.4 standard, considering various types of interference in a controlled environment. This study's hypothesis, which suggests that specific conditions in a network's environment can lead to packet loss, is validated through experimentation. Notably, interference from WiFi devices shows a weak negative association with packet status, indicating that WiFi interference may contribute to packet loss.

The application of this work lies in improving the reliability of wireless networks, especially in environments with potential interference sources like WiFi devices. By understanding the conditions that lead to packet loss, network administrators can take proactive measures to minimize disruptions and optimize network performance. Additionally,

the methodology and experimentation described in this study provide a framework for future research in the field of wireless network reliability and packet loss prediction.

The analysis of packets lost in different types of interferences suggests that there exists, in fact, a relevant relationship between the combination of fields of a sniffed packet and the prediction of its status, as stated in the hypothesis.

As a side note, among all the interference tests performed, WiFi interference appeared to induce the most significant fluctuations in data, resulting in a higher number of outliers, as illustrated in Figure 4. This observation aligns with the interference type that exhibits the highest linear correlation with the class out of all interferences to which the network was exposed.

When analyzing the patterns that lead to packet loss in a WiFi interference environment, we observe that the network is occasionally perceived as weak, and this is when large packets are lost.

In future work, our objective is to identify the types and quantities of devices causing WiFi interference, and consequently, packet loss. Additionally, we intend to develop predictive models by employing balancing techniques and predicting rare events to enhance the precision metrics presented in this study.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/informatics11010007/s1.

Author Contributions: Conceptualization, C.D.-V.-S. and H.P.-D.; methodology, M.Á.-A.; software, M.Á.-A. and H.P.-D.; validation, C.D.-V.-S. and R.A.B.; formal analysis, R.A.B.; investigation, C.D.-V.-S.; resources, C.D.-V.-S.; data curation, M.Á.-A.; writing—original draft preparation, M.Á.-A.; writing—review and editing, C.D.-V.-S.; visualization, M.Á.-A.; supervision, C.D.-V.-S.; project administration, C.D.-V.-S.; funding acquisition, C.D.-V.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Integrity Code of the Universidad Panamericana was observed, validated by the Social Affairs Committee and approved by the Governing Council through resolution CR 98-22, on 15 November 2022.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

Acknowledgments: This work was supported in part by collaboration with REDTPI4.0 CYTED program.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fanibhare, V.; Sarkar, N.I.; Al-Anbuky, A. A survey of the tactile internet: Design issues and challenges, applications, and future directions. *Electronics* **2021**, *10*, 2171. [CrossRef]
2. Ye, T.; Veitch, D.; Bolot, J. Improving wireless security through network diversity. *ACM SIGCOMM Comput. Commun. Rev.* **2008**, *39*, 34–44. [CrossRef]
3. Nour, B.; Mastorakis, S.; Ullah, R.; Stergiou, N. Information-centric networking in wireless environments: Security risks and challenges. *IEEE Wirel. Commun.* **2021**, *28*, 121–127. [CrossRef] [PubMed]
4. Sanyal, I.; Rao, D.P.; Gunasekaran, R.; Sachin, S.; Prabhakar, T. Lessons Learnt From the Implementation of the IEEE 802.15. 4e-TSCH MAC. In Proceedings of the 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS), Bangalore, India, 3–8 January 2023; pp. 658–666.
5. Kafetzis, D.; Vassilaras, S.; Vardoulias, G.; Koutsopoulos, I. Software-defined networking meets software-defined radio in mobile ad hoc networks: State of the art and future directions. *IEEE Access* **2022**, *10*, 9989–10014. [CrossRef]
6. Felser, M.; Rentschler, M.; Kleineberg, O. Coexistence standardization of operation technology and information technology. *Proc. IEEE* **2019**, *107*, 962–976. [CrossRef]
7. Mindo, K.G. A Fused Machine Learning Intrusion Detection Model In Manets. Ph.D. Thesis, Kabarak University, Nairobi, Kenya, 2019.
8. Sun, W.; Yuan, X.; Wang, J.; Li, Q.; Chen, L.; Mu, D. End-to-end data delivery reliability model for estimating and optimizing the link quality of industrial WSNs. *IEEE Trans. Autom. Sci. Eng.* **2017**, *15*, 1127–1137. [CrossRef]

9. Ayeesha Nasreen, M.; Ravindran, S. Long short-term memory-based power-aware algorithm for prompt heterogenous activity. *Int. J. Commun. Syst.* **2022**, *35*, e5163. [CrossRef]
10. Gonzalez, I.A.M.; Turau, V. Comparison of WiFi Interference Mitigation Strategies in DSME Networks: Leveraging Reinforcement Learning with Expected SARSA. In Proceedings of the 2023 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), Dubrovnik, Croatia, 4–7 September 2023; pp. 270–275.
11. Raza, S.; Faheem, M.; Guenes, M. Industrial wireless sensor and actuator networks in industry 4.0: Exploring requirements, protocols, and challenges—A MAC survey. *Int. J. Commun. Syst.* **2019**, *32*, e4074. [CrossRef]
12. Landaluce, H.; Arjona, L.; Perallos, A.; Falcone, F.; Angulo, I.; Muralter, F. A review of IoT sensing applications and challenges using RFID and wireless sensor networks. *Sensors* **2020**, *20*, 2495. [CrossRef]
13. Wagner, J.P.; Hausheer, D.; Gartner, M. Improving Packet Processing Speed on SCION Endhosts. Ph.D. Thesis , Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany, 2021.
14. Bennis, M.; Debbah, M.; Poor, H.V. Ultrareliable and low-latency wireless communication: Tail, risk, and scale. *Proc. IEEE* **2018**, *106*, 1834–1853. [CrossRef]
15. Shah, S.M.; Sun, Z.; Zaman, K.; Hussain, A.; Ullah, I.; Ghadi, Y.Y.; Khan, M.A.; Nasimov, R. Advancements in Neighboring-Based Energy-Efficient Routing Protocol (NBEER) for Underwater Wireless Sensor Networks. *Sensors* **2023**, *23*, 6025. [CrossRef] [PubMed]
16. Freschi, V.; Lattanzi, E. A study on the impact of packet length on communication in low power wireless sensor networks under interference. *IEEE Internet Things J.* **2019**, *6*, 3820–3830. [CrossRef]
17. Woolsey, N.; Chen, R.R.; Ji, M. Towards finite file packetizations in wireless device-to-device caching networks. *IEEE Trans. Commun.* **2020**, *68*, 5283–5298. [CrossRef]
18. Tüker, M.; Karakış, E.; Sayıt, M.; Clayman, S. Using packet trimming at the edge for in-network video quality adaption. *Ann. Telecommun.* **2023** , 1–14. [CrossRef]
19. Seyfollahi, A.; Ghaffari, A. A review of intrusion detection systems in RPL routing protocol based on machine learning for internet of things applications. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 1–32. [CrossRef]
20. Shahraki, A.; Taherkordi, A.; Haugen, Ø.; Eliassen, F. A survey and future directions on clustering: From WSNs to IoT and modern networking paradigms. *IEEE Trans. Netw. Serv. Manag.* **2020**, *18*, 2242–2274. [CrossRef]
21. Narayanan, A.; De Sena, A.S.; Gutierrez-Rojas, D.; Melgarejo, D.C.; Hussain, H.M.; Ullah, M.; Bayhan, S.; Nardelli, P.H. Key advances in pervasive edge computing for industrial internet of things in 5g and beyond. *IEEE Access* **2020**, *8*, 206734–206754. [CrossRef]
22. Jawad, A.T.; Maaloul, R.; Chaari, L. A comprehensive survey on 6G and beyond: Enabling technologies, opportunities of machine learning and challenges. *Comput. Netw.* **2023**, *237*, 110085. [CrossRef]
23. Alruhaily, N.M.; Ibrahim, D.M. A multi-layer machine learning-based intrusion detection system for wireless sensor networks. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 281–288. [CrossRef]
24. Chen, Y.; Lu, L.; Yu, X.; Li, X. Adaptive method for packet loss types in IoT: An naive Bayes distinguisher. *Electronics* **2019**, *8*, 134. [CrossRef]
25. Dong, R.; She, C.; Hardjawana, W.; Li, Y.; Vucetic, B. Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 4692–4707. [CrossRef]
26. James, J.; Lam, A.Y.; Hill, D.J.; Hou, Y.; Li, V.O. Delay aware power system synchrophasor recovery and prediction framework. *IEEE Trans. Smart Grid* **2018**, *10*, 3732–3742.
27. Kochovski, P.; Drobintsev, P.D.; Stankovski, V. Formal quality of service assurances, ranking and verification of cloud deployment options with a probabilistic model checking method. *Inf. Softw. Technol.* **2019**, *109*, 14–25. [CrossRef]
28. Yin, M. Application of Active Learning Algorithm in Mobile Ad Hoc Network Intrusion Detection. In Proceedings of the 2023 World Conference on Communication & Computing (WCONF), Virtual Event, 14–16 July 2023; pp. 1–6.
29. Lindh, J.; Lee, C.; Hernes, M.; Johnsrud, S. *Measuring CC13xx and CC26xx Current Consumption*; Application Report; Texas Instrument Incorporated: Dallas, TX, USA, 2019.
30. Texas Instrument Incorporated. *SWRU222: CC USB Software Examples User's Guide*; Texas Instrument Incorporated: Dallas, TX, USA, 2009.
31. Rohde & Schwarz. *Generation of IEEE 802.15.4 Signals*; Rohde & Schwarz: Munich, Germany, 2012.
32. Texas Instruments Incorporated. *SmartRF Packet Sniffer User's Manual*; Texas Instruments Incorporated: Dallas, TX, USA, 2010.
33. Hedberg, M.F. PacketZapper: An Automated Collection and Processing Platform for IoT Device Traffic. Master's Thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, 2023.

34. Abdallah, A.E.; Hamdan, M.; Abd Razak, S.; Ghalib, F.A.; Hamzah, M.; Khan, S.; Ali, S.A.B.; Khairi, M.H.; Salih, S. Resource Exhaustion Attack Detection Scheme for WLAN Using Artificial Neural Network. *Comput. Mater. Contin.* **2023**, *74*, 5607–5623.
35. Huang, R.; Nie, Z.; Duan, C.; Liu, Y.; Jia, L.; Wang, L. Analysis and comparison of the IEEE 802.15. 4 and 802.15. 6 wireless standards based on MAC layer. In Proceedings of the International Conference on Health Information Science, Melbourne, VIC, Australia, 26–30 May 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 7–16.