Toward Effective Framework for Wireless Intrusion Detection System in Detecting Krack and kr00k attacks in IEEE 802.11

Zaher Salah * and Esraa Abu Elsoud

Posted Date: 25 July 2023

doi: 10.20944/preprints202307.1619.v1

Keywords: Wireless; IDS; Machine Learning; Krack; Kr00k; IEEE8021.11

Article

Toward Effective Framework for Wireless Intrusion Detection System in Detecting Krack and kr00k attacks in IEEE 802.11

Zaher Saleh Salah [1,*] and Esraa Taher Abu Elsoud [2]

[1] Department of IT, The Hashemite University, Jordan
[2] Department of CIS, The Hashemite University, Jordan; 2070606@hu.edu.jo
* Correspondence: zaher@hu.edu.jo

Abstract: The rise in Internet users has brought with it the impending threat of cybercrime as the Internet of Things (IoT) increases and the introduction of 5G technologies continues to transform our digital world. It is now essential to protect communication networksfrom illegal intrusions in order to guarantee data integrity and user privacy. In this situation, machine learning techniques used in data mining have proven to be effective tools for constructing intrusion detection systems (IDS) and improving their precision.We use the well-known AWID3 dataset, a comprehensive collection of wireless network traffic, to investigate the effectiveness of machine learning in enhancing network security. Our work primarily concentrates on the Krack and Kr00k attacks, which targetthe most recent and dangerous flaws in IEEE 802.11 protocols. Through diligent implementation, we were able to successfully identify these threats using an IDS model that is based on machine learning. Notably, the resilience of our method was demonstrated by our Ensemble classifier's astounding 99% success rate in detecting the Krack attack. The effectiveness of our suggested remedy was further demonstrated by the high accuracyrate of 96.7% displayed by our Neural Network-based model in recognizing instances of the Kr00k attack. Our research showsthe potential for considerably boosting network security in the face of new threats by leveraging the capabilities of machine learning and a diversified dataset. Our findings open the door for stronger, more proactive security measures to protect IEEE 802.11 networks' integrity, resulting in a safer online environment for all users.

Keywords: wireless; IDS; machine learning; krack; Kr00k; IEEE8021.11

I. Introduction

The next generation of wireless networks will require a unified platform to support the vast numbers of devices, users,and services with different data rates and latency requirements. Current wireless technologies like 3G and 4G-LTE haveseveral limitations that restrict any possible enhancement of the systems to meet these demands. Accordingly, researchers have developed an advanced wireless communication technology called (5G) to satisfy the requirements above. After several scientificresearch, it was found that the fifth-generation technology has limitations too, it can't be used for long-distance communication or low-power wide-area technology. This indicates that current communication technology will not be fully and effectivelyable to meet demands in the future. A highly advanced digital civilization backed by limitless wireless connectivity is also anticipated to have arisen by the year 2030 [1].

In particular, within the context of 5G technology, the Internet of Things (IoT) has arisen as a ground-breaking idea, where technologies and solutions are incorporated to connect objects, people, platforms, and software via the Internet. In orderto create comprehensive IoT networks, these devices will be endogenously fully equipped with IoT modules that enable D2D communication with one another [2]. Furthermore, RAT will be supported by 5G to connect these devices. New radio technologies, such as NOMA, massive MIMO, mmWave, and a number of other IoT communication technologies, will be introduced in the 5G network.

One of the primary requirements for 5G systems and beyond is security related to 5G technologies. To investigate security in networks 4G use cryptography protocols for user

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

authentication [3], while 3G networks use two-way authentication to prevent connection establishment with fake base stations [4]. A new age has begun with the introduction of 5G technology, which presents distinct security and privacy challenges beyond those faced by earlier systems. An innovative approach to security procedures is necessary given the changing architecture and the introduction of new services. Researchers have added newideas like visibility and centralized policy in addition to the fundamental security principles of Confidentiality, Integrity, and Availability. These updates are intended to strengthen data protection in the face of evolving threats by enhancing security safeguards [5], [6] as shown in Figure 1.

Toward Effective Framework for Wireless Intrusion Detection System in Detecting Krack and kr00k attacks in IEEE 802.11

Figure 1. Evolved Security requirements for 5G technologies.

One of the most important security requirements in the 5G security model is data confidentiality; it is the parameter that can protect the transmission data from disclosure unauthorized entities: it means ensuring that the sender message in the 5G networks is only readable by the proposed destination. For preserving data secrecy in the context of 5G network applications, encryption methods have evolved into essential instruments. Data within 5G networks can be safely secured and decryptedby using symmetric key encryption. Sensitive information is kept secure by being protected from unauthorized access or interception due to this critical security defend [7].

Data integrity relates to keeping data from being alert or modified while transformed from one location to another, to investigatethe integrity concept authentication technique used by 5G-AKA (Authentication and Key Agreement), The fact that the 5GNew Radio (NR) offers user integrity protection is a significant improvement in 5G security. This is significant because user plane integrity protection was not supported by 4G. Small data transmissions can take advantage of this new feature, especially for IoT devices with constrained bandwidth [8, 9].

Availability issues—in most cases, are related to DoS attacks and are conducted in a wireless network by jamming. Spread- spectrum techniques can prevent jamming effectively, but they cannot be used in IoT nodes with limited resources (e.g., sensors). 5G does not implement new techniques to test the availability.

A centralized Security Policy can protect all enterprise endpoints from external threats and can be investigated through authentication. All network layers must be covered by comprehensive end-to-end security strategies for 5G networks. 5G operators need to have full visibility, control, and monitoring of overall network layers to implement such a thorough security mechanism. To manage and control security policies, open Application Program Interfaces (APIs) should be combined with5G technologies. The 5G network can thus have uniform software and hardware security policies. The implementation of the security mechanism in new 5G services will become easier with high visibility across the network. Furthermore, improved visibility helps in prevent and isolate the threats before attacks happen [7]. However, most research focuses on the three

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

core requirements for 5G security the CIA [10–12]. Figure 2 presents the Core requirements for 5G security and the solutions to maintain these requirements.

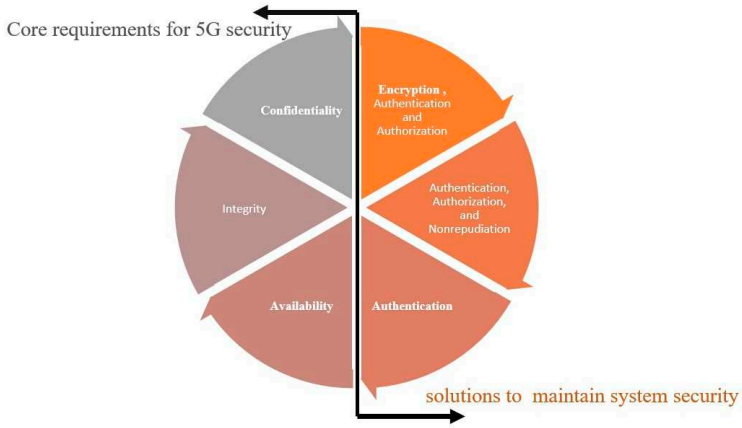

Figure 2. Core requirements for 5G security (left) and solutions to maintain system security (right).

A comprehensive security policy is the starting point for managing the countermeasures need to secure wireless networks. Technical countermeasures which help in wireless security environments include hardware and software. Hardware counter- measures like Smart cards, VPNs, and bio-metrics are hardware solutions. while proper AP configuration, software patches, authentication, intrusion detection systems (IDS), and encryption are all examples of software countermeasures [13].

The researchers put up suggestions based on cybersecurity defensive systems, particularly the well-known IDS, to protect againstcyber threats from wireless communications by anticipating and resolving flaws. Due to the heterogeneity and massive amountsof unstructured data present in the network, IDS procedures are useless for the real-time detection of potential intrusions in 5G [14]. One of the most promising methods in the field of artificial intelligence (AI) over the past few years has been machine learning (ML). Its remarkable abilities enable systems to learn from enormous amounts of data, considerably more than humansare able to process. Organizations can use ML to evaluate large datasets, improve their understanding, and provide accurate predictions. With its unique ability to predict and avert issues, this disruptive technology opens the door for more proactive and effective approaches in a variety of fields [15–17].

The machine learning field has enabled different paths that are effective in handling network intruders. Therefore, the employ- ment of ML tools in 5G systems has attracted much interest from international projects and research, such in [18–21].

ML techniques can examine the features of the network data to distinguish between attacks and normal traffic. Some attributes improve the accuracy of the intrusion detection system, whereas the network data contains noisy attributes well decreasesthe detection accuracy. The feature selection techniques are considered crucial for an IDS because FS help in increasing the accuracy of detection which is used as the input to the learning approaches.

A reliable intrusion detection system must be implemented because the IEEE 802.11 protocol-based short-distance transmis- sion wireless network has faced security issues. Many researchers have proposed various IDS systems, and it has been found that data-mining-based techniques are very effective at detecting abnormal network behavior in these systems [22]. Intrusion detection has new difficulties as network data keeps growing exponentially, which are caused by the data's nonlinear structure and enormous volume. The characteristics of the data, notably the existence of redundant features, have a substantial impacton the efficacy of existing approaches. The AWID dataset has become a popular experimental dataset in the field of wireless network environments. The "curse of dimensionality" problem is exacerbated by the large complexity and intrinsic duplication of this dataset. Designing effective intrusion detection strategies requires addressing

 doi:10.20944/preprints202307.1619.v1

these issues [22]. The large volume of the blind spot regions in the dimensionality dataset can lead to significantly variable estimates of actual model performance when algorithm designers use insufficient sample sizes to train and evaluate algorithms for finding patterns in a complex construct. Due to this variability, it is challenging to predict how well a model performs on data that has not yet been observed [23]. The imbalance between benign and attack samples might increase the false positive rate in addition to high dimensionality, which can influence the effectiveness of any suggested IDS [24].

AWID dataset was released in 2016 [25]. It's the first dataset of its kind that focuses on IDS, more specifically its focuson WIDS. Studying the AWID dataset will help researchers to be familiar with 802.11 network vulnerabilities and attacks,as well as educate them on the true effects of these attacks on daily life. The AWID dataset has 156 features and 37 million packets. A new version of this data was published in 2021 [26] by capturing and analyzing the traces of attacks that were sent into the IEEE 802.1X Extensible Authentication Protocol (EAP) environment. It focuses on WPA2 Enterprise, 802.11w, and Wi-Fi 5. It includes multi-layer attacks like Krack and Kr00k.

This study focuses on analyzing the most recent assaults on IEEE 802.11, especially Krack and Kr00k. Our main goal is tocreate an Intrusion Detection System (IDS) model using MATLAB to successfully detect and counteract these attempts.

The remainder of the paper is divided into the following sections:

Section 2 offers a thorough analysis of the relevant work, examining research that has been done in this area.

The technique used in this study is presented in Section 3, along with the step-by-step process we used to create our IDSmodel.

We show and discuss the findings from the experiments we conducted in Section 4, highlighting how well our IDS modelperformed in identifying Krack and Kr00k attacks.

Finally, we draw conclusions from our research in Section 5, where we go over the consequences, restrictions, and possibledirections for further study in the area of IEEE 802.11 security.

By arranging our work in this way, from the initial literature review to the concluding reflections on the significance of ourfindings, we hope to provide a thorough and coherent discussion of our research.

II. Literature Review

The evolution of communication networks has resulted in an exponential rise in the number of Internet of Things (IoT) devices that are connected to Wi-Fi networks. These devices generate enormous amounts of data traffic, which can be maliciousand makes it difficult to detect such attacks. Feature selection is used to reduce the amount of data for intrusion detectionmodel classifiers by removing noisy information and choosing the best features in the data, which participates in improving the IDS performance and solving these challenges.

We will focus on the AWID dataset, a Wi-Fi network intrusion benchmark dataset introduced due to the lack of the dataset in wireless intrusion detection systems (WIDSs) where the oldest datasets were about IDSs in general [27]. Moreover, this dataset was the first dataset produced using wireless network traffic.AWID has been enhanced and extended to AWID3 by capturing and examining the traces of cyberattacks sent into the IEEE 802.1X Extensible Authentication Protocol (EAP) environment. It concentrates on 5G wireless networks, 802.11w, and WPA2 Enterprise. It includes multi-layer and modern attacks like Krack and Kr00k [26]. This section will present some previous studies that used the AWID dataset in their research.

[28] In this study, the authors present a novel method that combines stacked feature extraction with weighted feature selection using deep learning techniques. This is a ground-breaking strategy for a Wi-Fi Impersonation Detection attack. The goal is to increase the detection of such attacks' precision and effectiveness. Three distinct algorithms—ANN, C4.5, and SVM—were used for the experiments and were developed and assessed using the AWID datasets as the basis.

 doi:10.20944/preprints202307.1619.v1

By utilizing this method, the authors were able to identify and detect Wi-Fi Impersonation assaults with an astounding accuracy rate of 99.918%. This represents an important development in the area and illustrates the effectiveness and potentialof the suggested deep-feature extraction and selection strategy. The results of this study emphasize the significance of using sophisticated methodologies and utilizing extensive datasets to address the constantly changing problems caused by Wi-Fi Impersonation attacks. The achieved accuracy rate underscores the value of ongoing research and development in increasing Wi-Fi security in addition to demonstrating the usefulness of the suggested approach. A. Diro et.al [29] used deep learning algorithms to identify and analyze critical attacks and threats on Internet of Things devices, particularly those that take advantageof weaknesses in wireless communications. They have achieved a high accuracy with 99.91% for the ISCX dataset and 98.22% for the AWID dataset.

In [30] the authors built a wireless intrusion detection system (IDS) designed specifically to operate access points in passive mode. The complex nature of wireless attacks, which frequently include the deceitful fabrication of fake access points to dupe unwary users, served as the inspiration for this strategy. The proposed IDS sought to efficiently detect and counteract such malicious activity by concentrating on the passive mode. When experimenting with the AWID dataset, the approach proposed by the authors produced encouraging results, obtaining an exceptional accuracy rate of 98%. These results demonstrate howwell the suggested method works for precisely recognizing and thwarting wireless attacks. The IDS demonstrated its capacity todistinguish between legal and fraudulent access points by utilizing the passive mode and applying complex detection algorithms.

S. M. Kasongo et.al [31] proposed a feed-forward deep neural network and feature extraction-based wireless intrusion detectionsystem (IDS). Two datasets, UNSW-NB15 and AWID, were used to assess the system's performance. They also contrasted theirfindings with those of well-known machine learning algorithms including Random Forest (RF), Support Vector Machine (SVM),Naive Bayes (NB), Decision Tree (DT), and k-Nearest Neighbor (kNN). Four categories—binary and multiclass attacks, full features, and selected features—were used to categorize the experimental research. They used the Extra Trees (ET) approachto reduce the number of features in the AWID dataset to 26. The test set had 115,128 instances, which made up 20% of the AWID-CLS dataset, whereas the training set had 359,115 instances. The suggested model's binary classification accuracy was 98.6% on the validation data and 98.69% on the test data. The highest accuracy for multiclass classification was 98.47% on the validation data and 98.59% on the test data. Furthermore, the performance of the suggested model showed increased accuracy outcomes when the number of attributes was decreased. The authors' high precision percentage for binary classification was 99.67% on the validation data and 99.66% on the test data. Similar results were obtained for multiclass classification, where 99.78% of the validation data and 99.77% of the test data were correctly classified. These results demonstrate the efficiency of the suggested wireless IDS system by demonstrating its precision in classifying and identifying various attack types. Deep neural networks and feature extraction were combined, and the accuracy significantly increased, especially when the numberof characteristics was decreased.

In [32] the authors addressed the imbalanced and high-dimensional network traffic issues with a system for intrusion detection.To improve classification accuracy, the suggested system makes use of feature selection and ensemble learning approaches.In this respect, the authors devised a hybrid strategy that combines the Bat Algorithm (BA) and Correlation-based Feature Selection (CFS), which maximizes the effectiveness of the feature selection process and enhances classification accuracy. The authors applied their suggested model to the NSL-KDD, AWID-CLS-R-Tst, and CIC-IDS2017 datasets in order to assess its efficacy. The AWID-CLS-R-Tst dataset underwent a number of preprocessing processes. This required both the replacement of missing values with zeros and the filtering out of features with constant values. The original 155 features were consequently

condensed to 84 features. An ensemble classifier was used to build the classification model. The superiority of the suggested strategy was shown by the experimental findings. With regard to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

AWID dataset in specific, a subset of just 8 chosen attributeswas used to obtain a remarkable accuracy rate of 99.52%. In contrast, the accuracy was 98.2% when feature selection was not used.

Hence, many literatures studied the impact of network intrusion detection systems; Z. Aydın et.al [33] mentioned in their research both the wireless and the wired intrusion detection system using the AWID dataset which includes the wireless networkattacks, and the UNSW-NB15 dataset which consists of the wired network attack. Also, they focused on the other performance matrices which are critical in intrusion detection systems such as F1, recall, and precision. After they preprocessed the datasets they remedied the imbalance problem via SMOTET (Synthetic Minority Over-Sampling Technique), the feature selection was performed using XGBoost, and finally, Bayesian optimization was applied before applying different ML algorithms.

D. L. Robert Wilson [34] attempted to address the issue of dataset imbalance and enhance the performance measures of machine learning methods applied to intrusion detection in their study. The AWID-CLS-F-Trn and AWID-CLS-R-Trn subsets of the AWID dataset were the subject of the investigation. These subsets represent various techniques for classifying assaults and actual attacks, respectively. It is noteworthy that the AWID dataset comprises two identical datasets that only vary in labeling strategies. Wilson used feature selection (FS) approaches and included all features in the analysis to improve the performance metrics. After that, the collected findings were contrasted, paying close attention to flooding and impersonation attacks. During the training phase, independent feature drop and group feature drop techniques were used. The results showed that when employing the Random Forest (RF) classifier, the independent features approach performed better. The group feature drop, on the other hand, showed enhanced dropping patterns for the flooding attack. However, when using the Logistic Regression (LR) classifier, the drop patterns for both independent and group feature drops were constant and comparable. The results showedthat the independent features strategy performed better when the Random Forest (RF) classifier was used. On the other hand, the flooding attack's group feature drop showed enhanced dropping patterns. The drop patterns for both independent and group feature drops were constant and similar when the Logistic Regression (LR) classifier was used, though.

In [35] the authors utilizing tree-based classification algorithms like Random Forest, XGBoost, LightGBM, and CatBoost, researchers examined the effects of feature selection. The feature set of the AWID dataset was reduced by the authors from155 to 15 features using the Shapley Additive exPlanations (SHAP) approach. Their investigation findings showed which characteristics had the greatest impact on detection models. The features wlan.da, wlan.fc-subtype, and wlan.lc.ds were shownto be the most significant in the detection process. The AWID dataset's patterns of intrusion and attack were distinguished and identified with the use of these attributes. The results indicate the value of particular features in enhancing the performance of intrusion detection models and emphasize the effectiveness of feature selection based on tree-based classification methods. The authors were able to improve the accuracy and efficacy of the detection procedure by limiting the feature set and concentrating on the most useful qualities.

Regarding a scalable ML-based intrusion detection system for IoT. The authors in [36] addressed the disadvantages of centralizedIDS for devices with limited resources, by utilizing two approaches—semi-distributed and distributed. The authors used feature selection approaches followed by classification in their investigation of the AWID dataset. The dataset was split into threeparts, each of which has 68 features. Seven features were chosen from each dataset after performing feature selection, givingthe full training set a total of 21 features. A semi-distributed strategy was used to find the most accurate feature selection technique for each dataset. The distribution method's classifier was also chosen to be a multilayer perceptron (MLP) classifier. Experiments were used to determine the efficacy of the two proposed structures. Even with a substantial CPU time of 186.26 seconds, an excellent accuracy of 99.97% was attained using the semi-distributed technique. While retaining a detection accuracy of 97.80%, the distributed technique showed the lowest CPU time of 73.52 seconds. These outcomes show that bothof the suggested topologies performed well in the tests that were conducted. The semi-distributed method demonstrated great accuracy, albeit

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

requiring more CPU time. In contrast, the distributed technique reduced CPU time while preserving acceptable detection accuracy.

The AWID2 dataset, which is the foundation of the wireless IDS literature and comprises a substantial collection of packets and its WEP-based infrastructure, has more than 150 different features. AWID has been enhanced and AWID3 has been introduced by capturing and examining the traces of cyberattacks that were transmitted into the IEEE 802.1X Extensible Authentication Protocol (EAP) environment.

III. Methodology

This chapter outlines the process for implementing the Framework for Intrusion Detection in Wireless Networks usingmachine learning techniques.

Wireless technologies have increased rapidly in recent years. While serious efforts have been made to secure these tech- nologies, most security measures have proven inadequate in practice. The AWID project aims to provide a solid basis for researchers to develop robust security mechanisms for current and future generations of wireless networks by providing tools, methodologies, and datasets, as the previous datasets weren't specific to wireless networks.

WiFi (IEEE 802.11) has taken over as the standardized technology for connecting digital devices in Wireless Local Area Networks due to the rise of smart portable devices such as smartphones, tablets, and Internet of Things (IoT) devices. WiFiis frequently used in critical locations as well as in homes, businesses, and organizations. Unsurprisingly, extensive academic research has focused on 802.11 protocol security as well as WiFi network security. With frequent modifications and corrective actions, vulnerabilities have been found in even the most recent versions of the software although these vulnerabilities have been existing for more than 20 years. Security in wireless technology is a major issue that has long gone unresolved. External security measures should therefore be used as crucial elements of 802.11 wireless networks for defending against known or unknown attacks [26].

AWID dataset was extracted in 2016 then it has been developed into a new version in 2021 called AWID3, the main difference between the old version and the new one can be summarized as follow:

1. AWID3 includes recently identified attacks against the 802.11 protocol, including well-known instances like Krack and kr00k. This inclusion enables researchers to investigate and create practical defenses against these particular dangers withinthe framework of the dataset.
2. A network's packet-level details are contained in the pcap format used to supply the data in AWID3. Researchers now have access to extensive data that can be utilized to assess network features and meet specific research objectives. The dataset also includes the Pairwise Master Key (PMK) and TLS keys.
3. The enterprise versions of the 802.11 standards are the main emphasis of AWID3. Stronger security features, like supportfor alternative network architectures and the use of Protected Management Frames (PMF), which were introduced with the 802.11w revision, are often present in these versions. By focusing on enterprise versions, the dataset is more applicable toactual security issues.
4. The link layer of the 802.11 protocol is initially targeted via attacks on the AWID3 dataset. These assaults, nevertheless, quickly spread to higher layers, affecting protocols that run at different levels of the network stack. Researchers can examinethe interrelated nature of network vulnerabilities due to this comprehensive perspective of attack propagation.
5. Every scenario in the dataset is covered in great detail by AWID3. Researchers may undertake detailed analysis and evaluationwith the help of this documentation, which also helps them grasp the nuances of assault scenarios.

A. Structure of AWID3 Dataset

The AWID3 dataset has been carefully curated to record and examine the traces of different assaults within the IEEE 802.1X Extensible Authentication Protocol (EAP) environment. It is valuable

and publicly available. It is significant for being the first dataset to offer a review of the IEEE 802.11w standard, which is necessary for hardware to be approved for use with theWPA3 protocol. The AWID dataset, from which AWID3 was built, has 254 features, of which 253 are general features andone is used for labeling. The dataset is offered in CSV format for simple access and interoperability with many different data analysis tools and methodologies. A thorough understanding of network activity and attack patterns is made possible by the extracted features, which cover both the MAC (Media Access Control) layer and the application layer. The dataset consists of (36,913,503) instances (30,387,099) normal traffic, and (6,526,404) malicious ones. The malicious traffic includes 13 types of attacks;

- Deauthentication Attack.
- Disassociation Attack.
- Re-association Attack.
- Rogue AP Attack.
- Krack Attack.
- Kr00k Attack.
- SSH Brute Force Attack.
- Botnet Attack.
- Malware.
- SSDP Amplification.
- SQL Injection Attack.
- Evil Twin.
- Website spoofing.

In our research, we used two types of these attacks which are krack and kr00k, since these two types are the most recenttype of attacks discovered in IEEE 802.11.

1) *Krack Attack:* The Krack attack has been noted as a potential security risk to the current encryption techniques usedto preserve and protect Wi-Fi networks for the past 15 years. Publicly available information on the Krack attack includes information about the attack itself. There is no guarantee that every device will have a patch and be protected from these attacks coming from any networked point [37, 38]. The four-way handshake procedure, which is a crucial part of the IEEE
802.11 protocol, has a serious weakness that allows any attacker to decode a user's communication without eavesdropping on the handshake or knowing the encryption key, according to In [39] study. This flaw results from the Pairwise Transient Key (PTK) installation process' use of a particular message counter. It is vital to look at how keystreams are used in the encryption process in order to comprehend the decryption process. The plaintext and keystream are merged using the XOR (exclusive OR) technique to create the encrypted message that is sent from the client to the Access Point (AP). The PTK, which is derivedusing the AES (Advanced Encryption Standard), is scrambled with a number of other factors to create the keystream. The vulnerability, though, only exists in the XOR operation's last phase. The logic flow of this step is connected to a fundamental mathematical feature that is exploited by the KRACK vulnerability. Equation (1) shows how the plaintext (P) and keystream(KS), as shown in the paper, are combined to create the ciphertext (E). The KRACK hack uses this defect in the XOR methodto decrypt the encrypted communications, putting the security of wireless networks using the IEEE 802.11 standard at risk.

$$E = P \oplus KS \tag{1}$$

An attacker could use two captured encrypted packets to decrypt them. Since the keystreams are identical, XORing the twociphertexts results in the keystreams being canceled and leaving two plaintexts.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

$$E1 = P1 \oplus KS1 \tag{2}$$

$$E2 = P2 \oplus KS2 \tag{3}$$

$$KS1 = KS2 = KS \tag{4}$$

Then:

$$E1 \oplus E2 = (P1 \oplus KS) \oplus (P2 \oplus KS) = P1 \oplus P2 \tag{5}$$

If the attacker were to accurately estimate or know P1, they could decrypt P2. The well-known first message that the AP or client sends after connecting can be used for this. The key WPA2 stream was designed to stop this exploitation, but KRACK researchers have found a way around it. Most of the keystream is made up of the static variables PTK, GTK, flags, MAC addresses, and counters. The only variable that alters when communications are encrypted is the packet number. Because every encrypted communication will have a different packet number and unique keystream, XOR cancellation is not conceivable [40].

2) *Kr00k Attack:* Some WiFi traffic that has been encrypted with WPA2 can be decrypted by a vulnerability called Kr00k. Security company ESET discovered the vulnerability in 2019. According to ESET, this loophole affects more than a billion devices. Devices with Wi-Fi chips that have not yet received a patch from Broadcom or Cypress are vulnerable to Kr00k. The majority of modern Wi-Fi-enabled devices, including smartphones, tablets, laptops, and Internet of Things (IoT) devices, use these Wi-Fi chips [41]. Table I highlighted the main difference between Krack and Kr00k attacks.

Table I. Comparing Krack and Kr00k.

krack	kr00k
KRACK is a series of attacks, exploited by attackers	Kr00k is a vulnerability in WPA2
The basic idea of KRACK is that the attacker can use the keystream to know the plain and the cipher text.	The encryption employed to secure data packets transmitted over a WiFi connection is impacted by Kr00k. Typically, a unique key determined by the user's WiFi password is used to encrypt these packets. Researchers from ESET claim that during the "disassociation" process, this key is reset for Broadcom and Cypress Wi-Fi chips to an all-zero value.
Exploited during the 4-way handshake	Exploited after a disassociation
Because it exploits implementation flaws in the WPA2 protocol itself, it affects the vast majority of Wi-Fi-capable devices.	Identified in Broadcom and Cypress components used in mobile phones, tablets, laptops, and IoT devices.

B. Preprocessing steps

As we mentioned before the dataset consists of (36,913,503) instances (30,387,099) normal traffic, and (6,526,404) maliciousones.

(49,990) instances for the Krack attack while (186,173) instances for the kr00k attack.

We will implement our experiment in two phases; the first phase consists of two classes (Krack, normal) and (kr00k, normal).While the second phase consists of muli-class(Krack, kr00k, and normal).

To highlight the importance of the preprocessing of the dataset before using it in the proposed model, we have used the chosen sample without any preprocessing and feature selection techniques.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

The first sample consists of (106971 kr00k traffic and 106791 normal one), while the second sample consists of (33180 Krack traffic) and (34000 normal one), with 254 features for both samples, as Shawn in Table II.

Table II. Training and Testing Samples in Awid3 Dataset.

Samples	Attack	Normal
First Sample	106971 kr00k traffic	128093
Second Sample	33180 Krack traffic	34000

We chose the following machine learning algorithms:

1) Decision tree: the process for building a decision tree and the most commonly used criteria for splitting the data [42]:
   - Calculate an impurity measure for the entire dataset (e.g., Gini impurity or entropy).
   - For each feature, calculate the impurity measure of splitting the data based on the values of that feature.
   - Choose the feature that produces the lowest impurity measure after splitting the data.
   - Split the data based on the chosen feature and repeat the process for each resulting subset of data until a stopping criterionis met (e.g., a maximum depth is reached or the number of samples in a leaf node is below a certain threshold).

The equations for calculating impurity measures depend on the specific criterion being used. For example, the Gini impuritymeasure for a set of samples S with C classes is:

$$G(S) = 1 - \sum_{i=1}^{C} p_i^2 \tag{6}$$

where p_i is the proportion of samples in S that belong to class i.The entropy impurity measure for the same set of samples S is:

$$H(S) = -\sum_{i=1}^{C} p_i \log_2 p_i \tag{7}$$

where p_i is the same as above.

These impurity measures are used to evaluate the quality of each split and to choose the feature that produces the lowest impurity measure.

2) Ensemble classifiers: combine multiple individual classifiers into a single ensemble classifier to improve the overall predictiveperformance. There are different types of ensemble classifiers, such as bagging, boosting, and stacking, and the equations used for each type can vary.
3) SVM: Support Vector Machine (SVM) is a popular machine learning algorithm for classification, regression, and outlier detection. The main idea behind SVM is to find a hyperplane that separates the data into different classes with the largestmargin possible. The equations used in SVM [43]:

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right) \tag{8}$$

where $f(x)$ is the predicted class label, α_i is the Lagrange multiplier for the i-th training sample, y_i is the class labelof the i-th training sample (either +1 or -1), $K(x, x_i)$ is the kernel function that maps the input features x and x_i to a higher-dimensional space, and b is the bias term.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

$$\gamma = \frac{1}{|\mathbf{w}|} = \frac{1}{\sqrt{\sum_{i=1}^{N} \alpha_i^2 y_i^2}} \tag{9}$$

where **w** is the weight vector of the hyperplane.

$$\underset{\alpha}{\text{maximize}} \quad W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad \text{subject to} \quad 0 \le \alpha_i \le C \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{10}$$

where $W(\alpha)$ is the objective function to be maximized, C is a user-defined parameter that controls the trade-off betweenthe margin and the number of training errors, and the constraints ensure that the Lagrange multipliers are non-negative andsum up to zero.

4) Kernal: A kernel function is a function that maps the input data into a higher-dimensional space, where it is easier to finda separating hyperplane. The equation of linear Kernal [44]:

$$K(x_i, x_j) = x_i^T x_j \tag{11}$$

where x_i and x_j are the input features of the i-th and j-th training samples, respectively.

5) KNN: K-Nearest Neighbors (KNN) is a simple yet effective machine learning algorithm used for classification and regression tasks. The basic idea behind KNN is to find the K nearest training samples to a given test sample based on adistance metric, and then use the labels of the K nearest neighbors to predict the label of the test sample. The equation ofKNN can be represented as follow [42]:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \tag{12}$$

where p is the number of features in each sample.

6) Neural Network: Neural Networks are a powerful class of machine learning algorithms that are inspired by the structure and function of the human brain. A neural network consists of multiple layers of interconnected processing units called neurons,and the input data is processed through the network in a forward pass, with the output of each layer serving as the input to the next layer. Some equations used in neural networks [45]:

- Activation function://

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{13}$$

where z is the input to the neuron. The ReLU function is given by:

$$ReLU(z) = \max(0, z) \tag{14}$$

where z is the input to the neuron.

- Forward pass:

$$y_j = f\left(\sum_{i=1}^{n} w_{ij} x_i + b_j\right) \tag{15}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

where y_j is the output of the j-th neuron in the layer, x_i is the i-th input to the layer, w_{ij} is the weight of the connection between the i-th input and the j-th neuron, b_j is the bias term of the j-th neuron, and f is the activation function.

- Backpropagation:

$$w_{ij} \leftarrow w_{ij} - \alpha \frac{\partial E}{\partial w_{ij}} \tag{16}$$

$$b_j \leftarrow b_j - \alpha \frac{\partial E}{\partial b_j} \tag{17}$$

where E is the error function, α is the learning rate, and $\frac{\partial E}{\partial w_{ij}}$ and $\frac{\partial E}{\partial b_j}$ are the partial derivatives of the error concerning the weights and biases, respectively.

After applying different ML algorithms, using the cross-validation K=10, the accuracy results were very low, which is to be expected. The results are presented in Table III

Table III. The Performance of the Learning Algorithms without Preprocessing.

Algorithm	Accuracy / kr00k attack	Accuracy / Krack attack
Decision tree	22.10%	50%
Ensemble classifier	44.20%	50%
SVM	12.50%	failed
Kernal	12.50%	failed

The accuracy results proved the importance of preprocessing steps since it's a constructive and essential step for obtainingthe correct data required to build a classifier, as shown in several types of research such as [46–48]. Data preprocessing,which aims to convert the raw data into a simpler and more efficient format for subsequent processing steps, is a crucial step in the knowledge discovery process because quality decisions must be based on quality data. Thus, the preprocessing procedures were carried out on the AWID3 dataset. The AWID3 consists of 13 CSV files with (36,913,503) instances, (30,387,099) normaltraffic, and (6,526,404) malicious ones. That was studied and well understood.

1) Detecting Krack attack: According to the importance of preprocessing step as we mentioned before, the preprocessing procedure for the Krack dataset sample was as follows:

1- Deleting the constant and empty features.
2- Ignoring features that have more than 60% missing values.3- Replace missing values with NaN.

The remaining dataset consists of 67 features and (67180 instances) (33180 Krack traffic and (34000 Normal traffic).

After preprocessing the data, we applied different ML algorithms. Table IV shows The performance of the learning algorithmsafter preprocessing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

Table IV. The performance of the learning algorithms after preprocessing-krack attack.

Algorithm	Accuracy	True Positive Rate	False Negative Rate
Decision tree	95.1%	90%	10%
Ensemble classifier	50.6%	50.6%	49.4%
KNN	73.8%	48.3%	51.7%
SVM	50.6%	50.6%	49.4%

For the same previous sample, we have used feature selection techniques to reduce the computing time and enhance the accuracy of the detection model. We chose the ANOVA FS technique, as Shawn in Figure 3 which is a widely used statistical approach for comparing different independent means.

$$F_i = \frac{MS_B}{MS_W} = \frac{\sum_{j=1}^{n_i}(\bar{x}ij - \bar{x})^2/(k-1)}{\sum j = 1^{n_i} \sum_{l=1}^{k}(x_{ijl} - \bar{x}_{il})^2/(N-k)} \quad (18)$$

In this equation, F_i represents the ANOVA F-value for the i-th feature, MS_B is the between-group mean square, MS_W isthe within-group mean square, n_i is the number of samples in group i, k is the total number of groups, N is the total numberof samples, $\bar{x}ij$ is the mean of the j-th sample in group i, $\bar{x}$ is the overall mean, and $xijl$ is the l-th feature value of the j-th sample in group i.

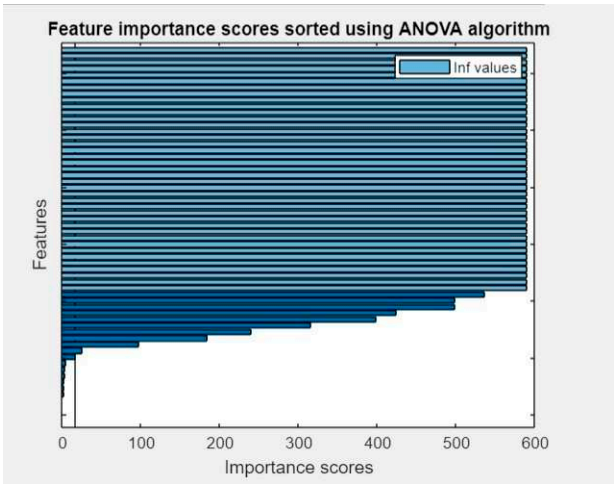

Figure 3. Feature importance scores stored using ANOVA algorithm.

The features ranked in the ANOVA method by calculating the variance ratio between and within groups [49]. The accuracyresults after applying ANOVA feature selection (FS=15), are shown in Table V.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

Table V. The performance of the learning algorithms after feature selection-krack attack.

Algorithm	Accuracy	True Positiv eRate	False Negativ eRate	AUC
Decision tree	93.2%	86.6%	13.4%	0.964
Naive Bayes	95%	97.7%	2.3%	0.9891
Ensemble classi-fier	99.1%	98.2%	1.8%	0.9998
KNN	68.2%	35.8%	64.2%	0.9995
SVM	73.8%	46.9%	53.1%	1

The results proved the necessity of processing the dataset since an efficient result depends on efficient data, furthermore,the results show improvement in accuracy results when we used feature selection techniques, as Figure 4 shows.

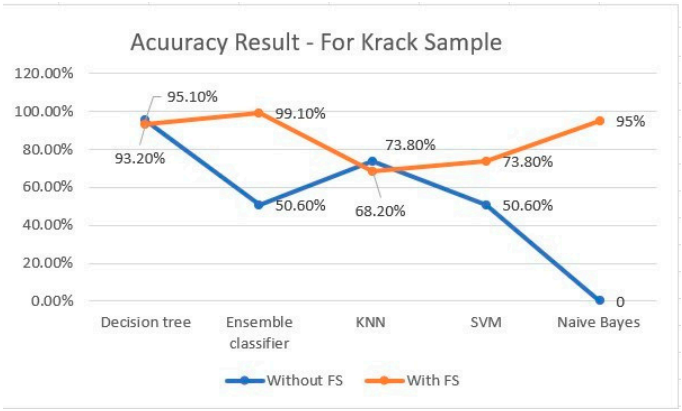

Figure 4. The performance of the learning algorithms- Krack attack.

The best accuracy results that we got were from the Ensemble classifier with 99.1% in addition to 1.8% False NegativeRate, followed by Naive Bayes with 95% accuracy result and 2.3% False Negative Rate.

2) Detecting Kr00k attack: This sample consists of 235,064 instances; (106971 kr00k traffic and 128093 normal ones), thepreprocessing procedure for the Krack dataset sample was as follows:

1- Deleting the constant and empty features.
2- Ignoring features with more than 60% missing values, the remaining features are 63.3- Replace missing values with NaN.

The remaining dataset consists of 63 features and 235,064 instances.

After preprocessing the data, we applied different ML algorithms. Table VI shows The performance of the learning algorithmsafter preprocessing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

Table VI. The performance of the learning algorithms after preprocessing-kr00k attack.

Algorithm	Accuracy	True Positive Rate	False Negative Rate
Decision tree	81.8%	60%	40%
Ensemble classifier	70%	45.3%	54.7%
KNN	53.3%	53.3%	46.7%
SVM	44.8%	40.7%	59.3%

For the same previous sample, we have used ANOVA feature selection techniques to reduce the computing time and enhancethe accuracy of the detection model. As Shawn in Figure 5 which is a widely used statistical approach for comparing different independent means.

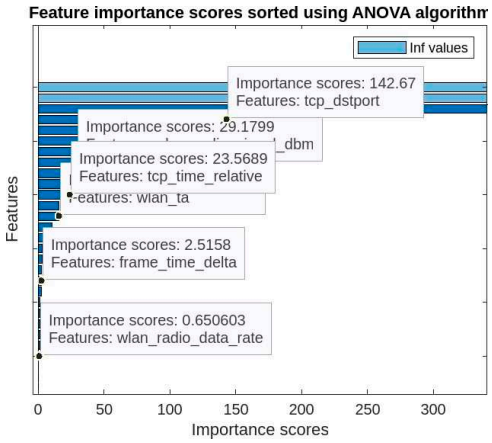

Figure 5. Feature importance scores stored using ANOVA algorithm- Kr00k.

The accuracy results after applying ANOVA feature selection (FS=15), are shown in Table VII. After applying the Ml algorithms on the chosen sample three times; without any process for the dataset, with preprocess and with FS, we can realize that the preprocess step is a critical and essential step in data mining to get accurate results. Especially in dealing with suchdata which suffer from high dimensionality imbalance and overfitting of the data. The accuracy results for the mentioned steps are presented in Figure 6. The best accuracy that we got was for Neural Network and SVM with 96.7%. We can conclude from Figure 6 how the accuracy affects by FS and prepossessing the dataset before applying any ML algorithms on it.

Table VII. The performance of the learning algorithms after feature selection-kr00k attack.

Algorithm	Accuracy
Decision tree	93.3%
Ensemble classifier	83.3%
KNN	86.7%
SVM	96.67%
Neural Network	96.7%
Kernal	83.3%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

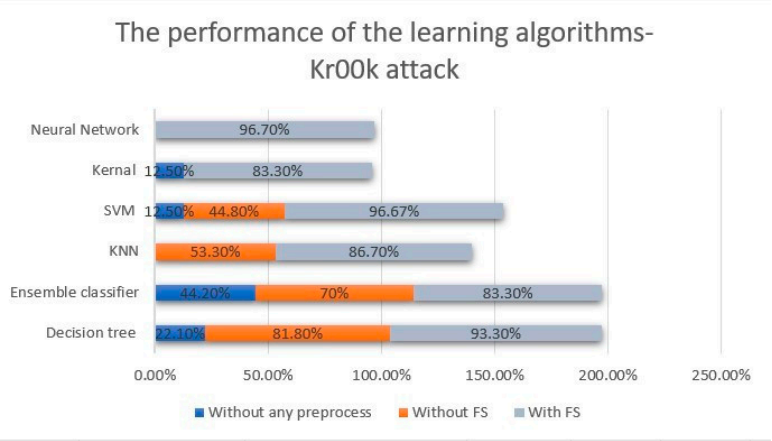

Figure 6. The performance of the learning algorithms- Kr00k attack.

Multi Class Detection: In this phase, we will use a sample consisting of three classes(Krack, Kr00k, and Nominal), 15,000 instances, and 254 features. Due to the importance of preprocessing as we noted in the previous subsections, we have applied the preprocessing steps in the chosen sample. Where we removed the empty features and the features with constant values, in addition to replacing all the empty cells in the remaining features with NaN. Then we applied ML algorithms usingthe classification linear application on MATLAB. The performance of the ML algorithms is presented in tableVIII, the table presented the accuracy results using FS with NOVA algorithm techniques and without using FS.

The accuracy results for the mentioned steps are presented in Figure 7.

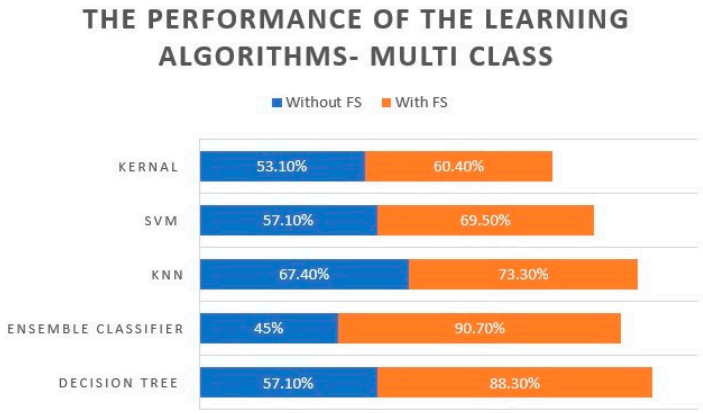

Figure 7. The performance of the learning algorithms- Multi class.

The best accuracy that we achieved without using FS was for KNN 67.4%, while when we applied the ANOVA FS the performance increased in all the used algorithms. The best accuracy was achieved by the Ensemble classifier with 90.7%, and for the Decision tree with 88.3%.

IV. Results and Discussion

We implemented our experiments in three phases; for the first phase, we used a sample of AWID3 dataset that includes a nominal two classes (Krack and Normal), the best accuracy that we achieved was for the Decision tree with 95.1% without using FS, while the best accuracy after using ANOVA FS was 99.1% for Ensemble classifier. For the second phase we used a sample with two nominal classes too (Kr00k and Normal), the best accuracy that we achieved was for the Decision tree as well, with 81.8% without using FS. The accuracy increased to 96.7% for Neural Network after applying ANOVA FS techniques. In the last phase, we used a multi-classes classifier, where the label

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

includes theree classes (Krack, Kr00k, and Nominal), and the accuracy results were determined low. However, after applying the FS techniques the accuracy increased to 90.7% and 88.3% for the Ensemble classifier and decision tree consecutively. Figure 8 summarized the results for the three mentioned phases.

Table VIII. The performance of the learning algorithms-multi class.

Algorithm	Accuracy without FS	Accuracy FS
Decision tree	57.1%	88.3%
Ensemble classifier	44.7%	90.7%
KNN	67.4%	73.3%
SVM	57.1%	69.5%
Kernal	53.1%	60.4%

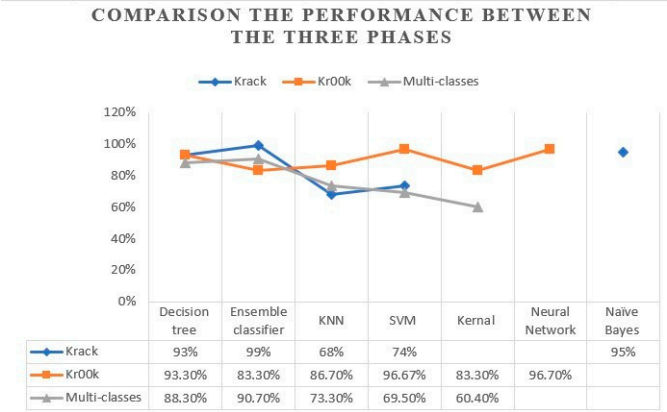

Figure 8. Comparison the performance between the three phases.

When looking at the results, it becomes clear that the Decision tree and Ensemble classifier have a high performance inthree experiments.

It's worth to a mention that the wireless dataset including the AWID3 dataset has many challenges; such as:

- High dimensionality refers to a high number of features in the dataset. So it's important to transform the data from a high- dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension [50]. we solved this problem using the ANOVA FS techniques,where the performance of the algorithms shows clear improvement in accuracy results when we reduce the high dimensionality for the dataset.
- Overfitting of the dataset: it occurs when a statistical model fits exactly against its training data [51]. Actually, when performingthe ML model to the data without solving the overfitting problem, the accuracy results will be almost 100% or 99.99% whichis not reliable performance. We solved this problem in the pre-processing step by getting rid of the features that coping thelabel, in addition to the importance of FS in solving this problem.
- imbalanced data: Unbalanced refers to a classification data set with skewed class proportions. We solved this problem by taking almost the same number of instances for attack and benign in the three experiments.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2023 doi:10.20944/preprints202307.1619.v1

Comparing our findings with previous studies

The authors in [37] used a state-machine architecture to find KRACK attacks by monitoring numerous wireless channels.To specifically identify the KRACK symptoms at various points of a handshake session, they have undertaken deep packet inspection and created a grouping method to group Wi-Fi handshake packets. They used supervised machine learning models based on gradient boosting, and their accuracy was around 93.39% with a false positive rate of 5.08%.

In [52] the authors proposed a framework for Unsupervised Classification and Data Mining of Tweets about Cyber Vulnera- bilities, this vulnerability included the Kr00K attack, which allows unauthorized decryption in WiFi chips. The best accuracy that they achieved was 88.52%

Chatzoglou et.al, applied Deep learning and machine learning techniques on the AWID3 benchmark dataset [53], in orderto answer questions about the competence of 802.11-specific and non-802.11 features when used separately and in tandemin detecting application layer attacks and to know which network protocol features are the most informative to the machine learning model for detecting application layer attacks, the performance for detection model achieved 96.7% accuracy.

Due to the increased urgency for unrestricted network data access to improve cyber-AI efficiency in unfamiliar threat scenarios,the authors in [54] proposed an automated network scanning and data-mining technique through open-source service discovery tools for deep reinforcement learning-based cognitive network intrusion detection systems. They obtain the lowest false alarm rate and a 98.68% accuracy.

In [55] the authors investigated how to map machine learning algorithms to programmable network devices. Furthermore, state-of-the-art and newly proposed in-network ML algorithms are evaluated and compared in terms of functionality, resources, scalability, and throughput. They used six datasets, including KDD99 and AWID3, for Intrusion Detection purposes. Their accuracy ranged from 97.47% for decision trees to 49.37% for KNN.

Table IX summarized the above-mentioned studies.

Table IX. Comparison of Our Findings with Other Studies.

Refrence	Year	Description	Accuracy
[37]	2022	They monitored numerous wireless channels to detect Krackattack	93.0%
[52]	2021	Proposed a framework for Cyber Vulnerabilities, includingKr00k	88.52%
[53]	2022	Proposed a model to detect the application layer attacks	96.7%
[54]	2021	proposed an automated network scanning and data-miningtechnique for Network IDS	98.68%
Our work		We focused on the IEEE 802.11 vulnerabilities, by proposinga model to detect Krack and kr00k attacks using ML tech- niques.	Phase 1: 99% Phase 2: 96.7% phase 3: 90.7%

V. Conclusion

Worldwide internet usage has significantly increased as a result of the extensive adoption of WiFi connectivity, but therehas also been a commensurate growth in cybercrimes that target the weaknesses of wireless systems. In this research, we concentrated on the recently discovered attacks known as Krack and Kr00k that were found in the IEEE 802.11 standard.The protocols for

constructing wireless local area networks, including the Media Access Control (MAC) and physical layer protocols, are specified in this standard, which is a part of the larger IEEE 802 collection of local area network technical standards.

We developed an Intrusion Detection System (IDS) model utilizing MATLAB's classification linear program to address the problems of these attacks. The AWID3 dataset, which is regarded as one of the most recent and well-liked wireless datasets accessible, was used to thoroughly test our suggested model. With the help of this research, we were able to successfullyhandle problems like excessive dimensionality and data imbalance that are frequently found in wireless datasets.

We detected these assaults with remarkable accuracy using our ML-based methodology. For example, our Ensemble classifier showed astounding 99% accuracy in recognizing Krack attacks, while our Neural Network classifier had the best accuracy in detecting Kr00K attacks at 96.7%.

This study highlights how important it is to comprehend the distinctive characteristics of wireless datasets and how they affect the effectiveness of detection models. We intend to address the issues posed by wireless datasets in the next work to improve wireless IDS performance even more. We work to contribute to the creation of stronger and more efficient wireless intrusion detection systems by constantly enhancing and modifying our methods.

References

1. S. Alraih, I. Shayea, M. Behjati, R. Nordin, N. F. Abdullah, A. Abu-Samah, and D. Nandi, "Revolution or evolution? technical requirements andconsiderations towards 6g mobile communications," *Sensors*, vol. 22, no. 3, p. 762, 2022.
2. L. Chettri and R. Bera, "A comprehensive survey on internet of things (iot) toward 5g wireless systems," *IEEE Internet of Things Journal*, vol. 7, no. 1,pp. 16–32, 2019.
3. V. T. Hoang Ahn and M. Ma, "A secure authentication protocol with performance enhancements for 4g lte/lte-a wireless networks," in *2021 3rdInternational Electronics Communication Conference (IECC)*, 2021, pp. 28–36.
4. P. Anantha Prabha, N. Arjun, J. Gogul, and S. Divya Prasanth, "Two-way economical smart device control and power consumption prediction system,"in *Proceedings of International Conference on Recent Trends in Computing*. Springer, 2022, pp. 415–429.
5. M. Liyanage, A. Braeken, A. D. Jurcut, M. Ylianttila, and A. Gurtov, "Secure communication channel architecture for software defined mobile networks," *Computer Networks*, vol. 114, pp. 32–50, 2017.
6. A. Gurtov, M. Liyanage, and M. Ylianttila, *Software defined mobile networks (SDMN): beyond LTE network architecture*. John Wiley & Sons, 2015.
7. J. H. Park, S. Rathore, S. K. Singh, M. M. Salim, A. Azzaoui, T. W. Kim, Y. Pan, and J. H. Park, "A comprehensive survey on core technologies andservices for 5g security: taxonomies, issues, and solutions," *Hum.-Centric Comput. Inf. Sci*, vol. 11, no. 3, 2021.
8. S. Gupta, B. L. Parne, and N. S. Chaudhari, "Security vulnerabilities in handover authentication mechanism of 5g network," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. IEEE, 2018, pp. 369–374.
9. R. Borgaonkar, I. Anne Tøndel, M. Zenebe Degefa, and M. Gilje Jaatun, "Improving smart grid security through 5g enabled iot and edge computing," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 18, p. e6466, 2021.
10. A. J. Gonzalez, P. Grønsund, A. Dimitriadis, and D. Reshytnik, "Information security in a 5g facility: An implementation experience," in *2021 JointEuropean Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2021, pp. 425–430.
11. H. Kim, "5g core network security issues and attack classification from network protocol perspective." *J. Internet Serv. Inf. Secur.*, vol. 10, no. 2, pp.1–15, 2020.
12. J. P. Mohan, N. Sugunaraj, and P. Ranganathan, "Cyber security threats for 5g networks," in *2022 IEEE International Conference on Electro InformationTechnology (eIT)*. IEEE, 2022, pp. 446–454.
13. K. Tsiknas, D. Taketzis, K. Demertzis, and C. Skianis, "Cyber threats to industrial iot: a survey on attacks and countermeasures," *IoT*, vol. 2, no. 1, pp.163–186, 2021.

14. S. Muthuramalingam, M. Thangavel, and S. Sridhar, "A review on digital sphere threats and vulnerabilities," *Combating Security Breaches and Criminal Activity in the Digital Sphere*, pp. 1–21, 2016.
15. P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017.
16. A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, and R. W. Heath, "5g mimo data for machine learning: Application to beam-selection using deep learning," in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–9.
17. V. P. Kafle, Y. Fukushima, P. Martinez-Julia, and T. Miyazawa, "Consideration on automation of 5g network slicing with machine learning," in *2018ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*. IEEE, 2018, pp. 1–8.
18. I. B. Sofi and A. Gupta, "A survey on energy efficient 5g green network with a planned multi-tier architecture," *Journal of Network and Computer Applications*, vol. 118, pp. 1–28, 2018.
19. I. Ioannou, C. Christophorou, V. Vassiliou, and A. Pitsillides, "A distributed ai/ml framework for d2d transmission mode selection in 5g and beyond," *Computer Networks*, vol. 210, p. 108964, 2022.
20. O. Nassef, W. Sun, H. Purmehdi, M. Tatipamula, and T. Mahmoodi, "A survey: Distributed machine learning for 5g and beyond," *Computer Networks*, vol. 207, p. 108820, 2022.
21. K. V. Babu, S. Das, G. N. J. Sree, S. K. Patel, M. P. Saradhi, and M. Tagore, "Design and development of miniaturized mimo antenna using parasitic elements and machine learning (ml) technique for lower sub 6 ghz 5g applications," *AEU-International Journal of Electronics and Communications*, vol. 153, p. 154281, 2022.
22. L. Yang, J. Li, L. Yin, Z. Sun, Y. Zhao, and Z. Li, "Real-time intrusion detection in wireless network: A deep learning-based intelligent mechanism," *IEEE Access*, vol. 8, pp. 170 128–170 139, 2020.
23. V. Berisha, C. Krantsevich, P. R. Hahn, S. Hahn, G. Dasarathy, P. Turaga, and J. Liss, "Digital medicine and the curse of dimensionality," *NPJ digital medicine*, vol. 4, no. 1, pp. 1–8, 2021.
24. S. J. Lee, P. D. Yoo, A. T. Asyhari, Y. Jhi, L. Chermak, C. Y. Yeun, and K. Taha, "Impact: Impersonation attack detection via edge computing using deep autoencoder and feature abstraction," *IEEE Access*, vol. 8, pp. 65 520–65 529, 2020.
25. C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis, "Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 184–208, 2015.
26. E. Chatzoglou, G. Kambourakis, and C. Kolias, "Empirical evaluation of attacks against ieee 802.11 enterprise networks: The awid3 dataset," *IEEE Access*, vol. 9, pp. 34 188–34 205, 2021.
27. C. Kolias, V. Kolias, and G. Kambourakis, "Termid: A distributed swarm intelligence-based approach for wireless intrusion detection," *International Journal of Information Security*, vol. 16, no. 4, pp. 401–416, 2017.
28. M. E. Aminanto, R. Choi, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim, "Deep abstraction and weighted feature selection for wi-fi impersonation detection," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 621–636, 2017.
29. A. Diro and N. Chilamkurti, "Leveraging lstm networks for attack detection in fog-to-things communications," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 124–130, 2018.
30. S. C. Sethuraman, S. Dhamodaran, and V. Vijayakumar, "Intrusion detection system for detecting wireless attacks in ieee 802.11 networks," *IET networks*, vol. 8, no. 4, pp. 219–232, 2019.
31. S. M. Kasongo and Y. Sun, "A deep learning method with wrapper based feature extraction for wireless intrusion detection system," *Computers & Security*, vol. 92, p. 101752, 2020.
32. Y. Zhou, G. Cheng, S. Jiang, and M. Dai, "Building an efficient intrusion detection system based on feature selection and ensemble classifier," *Computer networks*, vol. 174, p. 107247, 2020.
33. Z. Aydın, V. Ç. Güngör *et al.*, "Intrusion detection with bayesian optimization on imbalance wired wireless and software-defined networking traffics."
34. D. L. Robert Wilson, "Towards effective wireless intrusion detection using awid dataset," 2021.
35. S. Bhandari, A. K. Kukreja, A. Lazar, A. Sim, and K. Wu, "Feature selection improves tree-based classification for wireless intrusion detection," in *Proceedings of the 3rd International Workshop on Systems and Network Telemetry and Analytics*, 2020, pp. 19–26.

36. M. A. Rahman, A. T. Asyhari, L. Leong, G. Satrya, M. H. Tao, and M. Zolkipli, "Scalable machine learning-based intrusion detection system foriot-enabled smart cities," *Sustainable Cities and Society*, vol. 61, p. 102324, 2020.
37. A. Agrawal, U. Chatterjee, and R. R. Maiti, "ktracker: Passively tracking krack using ml model," in *Proceedings of the Twelveth ACM Conference onData and Application Security and Privacy*, 2022, pp. 364–366.
38. R. dos Reis Fontes12 and C. E. Rothenberg, "On the krack attack: Reproducing vulnerability and a software-defined mitigation approach."
39. M. Vanhoef and F. Piessens, "Key reinstallation attacks: Forcing nonce reuse in wpa2," in *Proceedings of the 2017 ACM SIGSAC Conference on Computerand Communications Security*, 2017, pp. 1313–1328.
40. C. P. Kohlios and T. Hayajneh, "A comprehensive attack flow model and security analysis for wi-fi and wpa3," *Electronics*, vol. 7, no. 11, p. 284, 2018.
41. M. Čermák, S. Svorenčík, and R. Lipovskỳ, "Kr00k-cve-2019-15126–serious vulnerability deep inside your wi-fi encryption," *ESET Research White Paper*, 2020.
42. T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer,2009, vol. 2.
43. V. Cortesc, "Support vector networks. m achine learning," *vol*, vol. 20, pp. 273–297, 1995.
44. J. Shawe-Taylor, N. Cristianini *et al.*, *Kernel methods for pattern analysis*. Cambridge university press, 2004.
45. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
46. S. Alam and N. Yao, "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis," *Computational and Mathematical Organization Theory*, vol. 25, no. 3, pp. 319–335, 2019.
47. A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N project report, Stanford*, vol. 1, no. 12, p. 2009, 2009.
48. C. Kubik, S. M. Knauer, and P. Groche, "Smart sheet metal forming: importance of data acquisition, preprocessing and transformation on the performance of a multiclass support vector machine for predicting wear states during blanking," *Journal of Intelligent Manufacturing*, vol. 33, no. 1, pp. 259–282, 2022.
49. H. Nasiri and S. A. Alavi, "A novel framework based on deep learning and anova feature selection method for diagnosis of covid-19 cases from chest x-ray images," *Computational intelligence and neuroscience*, vol. 2022, 2022.
50. R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selectionand feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.
51. M. Belkin, D. Hsu, S. Ma, and S. Mandal, "Reconciling modern machine-learning practice and the classical bias–variance trade-off," *Proceedings of the National Academy of Sciences*, vol. 116, no. 32, pp. 15 849–15 854, 2019.
52. K. Alperin, E. Joback, L. Shing, and G. Elkin, "A framework for unsupervised classificiation and data mining of tweets about cyber vulnerabilities," *arXiv preprint arXiv:2104.11695*, 2021.
53. E. Chatzoglou, G. Kambourakis, C. Smiliotopoulos, and C. Kolias, "Best of both worlds: Detecting application layer attacks through 802.11 andnon-802.11 features," *Sensors*, vol. 22, no. 15, p. 5633, 2022.
54. E. Muhati and D. B. Rawat, "Asynchronous advantage actor-critic (a3c) learning for cognitive network security," in *2021 Third IEEE InternationalConference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*. IEEE, 2021, pp. 106–113.
55. C. Zheng, M. Zang, X. Hong, R. Bensoussane, S. Vargaftik, Y. Ben-Itzhak, and N. Zilberman, "Automating in-network machine learning," *arXiv preprintarXiv:2205.08824*, 2022.