electronics

MDPI

Article

Machine-Learning-Enabled Intrusion Detection System for Cellular Connected UAV Networks

Rakesh Shrestha [1,*,†], Atefeh Omidkar [2,†], Sajjad Ahmadi Roudi [3,†], Robert Abbas [4,*,†] and Shiho Kim [1,*,†]

[1] Yonsei Institute of Convergence Technology, Yonsei University, Incheon 21983, Korea
[2] Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N1L6, Canada; ato947@usask.ca
[3] Department of Science and Engineering Firoozkooh Branch, Islamic Azad University, Tehran 1313965677, Iran; sajad.ahmadi.roudi@gmail.com
[4] School of Engineering, Macquarie University, Sydney, NSW 2109, Australia
* Correspondence: rakez_shre@yonsei.ac.kr (R.S.); robert.abbas@mq.edu.au (R.A.); shiho@yonsei.ac.kr (S.K.)
† These authors contributed equally to this work.

check for updates

Citation: Shrestha, R.; Omidkar, A.; Roudi, S.A.; Abbas, R.; Kim, S. Machine-Learning-Enabled Intrusion Detection System for Cellular Connected UAV Networks. *Electronics* **2021**, *10*, 1549. https://doi.org/10.3390/electronics10131549

Academic Editor: Amir Mosavi

Received: 28 May 2021
Accepted: 22 June 2021
Published: 26 June 2021

Abstract: The recent development and adoption of unmanned aerial vehicles (UAVs) is due to its wide variety of applications in public and private sector from parcel delivery to wildlife conservation. The integration of UAVs, 5G, and satellite technologies has prompted telecommunication networks to evolve to provide higher-quality and more stable service to remote areas. However, security concerns with UAVs are growing as UAV nodes are becoming attractive targets for cyberattacks due to enormously growing volumes and poor and weak inbuilt security. In this paper, we propose a UAV- and satellite-based 5G-network security model that can harness machine learning to effectively detect of vulnerabilities and cyberattacks. The solution is divided into two main parts: the model creation for intrusion detection using various machine learning (ML) algorithms and the implementation of ML-based model into terrestrial or satellite gateways. The system identifies various attack types using realistic CSE-CIC IDS-2018 network datasets published by Canadian Establishment for Cybersecurity (CIC). It consists of seven different types of new and contemporary attack types. This paper demonstrates that ML algorithms can be used to classify benign or malicious packets in UAV networks to enhance security. Finally, the tested ML algorithms are compared for effectiveness in terms of accuracy rate, precision, recall, F1-score, and false-negative rate. The decision tree algorithm performed well by obtaining a maximum accuracy rate of 99.99% and a minimum false negative rate of 0% in detecting various attacks as compared to all other types of ML classifiers.

Keywords: UAV; machine learning; intrusion detection system; cybersecurity attacks; software-defined security

1. Introduction

Recently, Unmanned Aerial Vehicles (UAVs) or drones have become popular and have been used for a variety of purposes in terms of everyday flying objects connected to the internet and able to identify themselves to other devices by sharing information via smart devices such as mobile phones and tablets. UAVs are flying objects that can fly independently or with the assistance of human pilots. UAVs are used for package delivery, aerial mapping, irrigation, environmental management, aerial photography, monitoring, rescue operations, and other critical applications. The reliability of UAV and its wireless communications is important for those critical applications. Security schemes and intrusion detection techniques are used to ensure critical security features. For communication, UAVs can communicate with terrestrial networks such as ground Base Stations (BSs) and non-terrestrial networks such as low-altitude satellites.

Space-based technologies, which include a newly named network of communications satellites for non-terrestrial uses, permit global telecommunications systems to transmit their signals based on voice, video, and other data from multiple access points [1]. Terrestrial infrastructure requirements can be reduced by using communications satellites so that more cost-effective service delivery options can be provided. Other applications of remote sensing satellites are in the agricultural area (for monitoring soil, drought, and crop development), environment (surveying water cycles, air quality, forests and state of ecosystems), UAV (communication), global health objectives (monitoring disease patterns, predict risk areas and define regions), and Internet of Things (IoT), where things are connected to the internet, which can be implemented in factory automation, smart homes, etc. [2]. According to [3], it is expected that there will be more than 27 billion devices connected to the internet by 2026 [4]. 5G networks will be able to support massive IoT (mIoT) devices, which will change the way society can interact with basic everyday objects. The majority of mobile networks are terrestrial, so they cannot cover remote areas, e.g., deserts, oceans, and forests, as they lack terrestrial infrastructure and universality of telecommunication software [5]. Similarly, UAVs flying at low altitudes cannot operate beyond cellular coverage area. Thus, satellite infrastructures, which extend and complement the terrestrial network, play a vital and crucial role in UAV networks [6]. The term Internet of Remote Things (IoRT) has been studied in [7], which reviews satellite-based IoTs in terms of Media Access Control (MAC) protocols for the sensor of satellite networks, supporting IPv6, heterogeneous networks interoperability, and managing Quality of Service (QoS) criterion. It can also be used in drones to remotely control them. A constellation of Low Earth Orbit (LEO) satellite architecture, efficient spectrum allocation, heterogeneous networks agreement, access, and routing protocols has been presented in [8]. Another novel architecture is designed to obtain intelligent, rapid, and efficient Heterogeneous Space and Terrestrial Integrated Networks (H-STINs) [9], which includes a proposed intelligent data center based on Software-Defined Networking (SDN)/Network Functions Virtualization (NFV) technology. The fifth generation (5G) networks can be deployed in satellite-based communication to achieve high bandwidth, low latency, and increasing coverage [10–12]. The introduction of UAVs into telecommunications networks, either as aerial users or as communication platforms, introduces new design possibilities as well as hurdles. Because of the high elevation and mobility of UAVs, service quality requirements, and the high chance of UAVs to ground Line of Sight (LoS) networks, both cellular-connected UAV communication and UAV-assisted wireless communication differ significantly from their terrestrial counterparts. In this paper, we focus on UAVs as a cellular user. Moreover, due to complexity, heterogeneity, and many interconnected resources, providing security in the UAV networks has become a big challenge in space-based networks. Some security issues that imply security in space-based information networks are hand-off security, transmission control security, and routing protocol security, which have been addressed in [13]. In [14], the security requirements of satellite-based wireless networks have been studied systematically, and the appropriate model for attacks are proposed for a satellite platform according to the MIL-STD-1553B bus, which is significantly used in an internal bus of spacecraft. There are two main ways for increasing security in satellite-based UAV networks: (1) encryption algorithms and (2) machine learning (ML) techniques, which are used to mitigate anomalies [15,16]. Encryption algorithms can protect the networks against external attacks by using authenticated packets from the source node and have two classifications, which are content-based and stream-based. An automation level should be considered in 5G-based UAV networks because of its complexity, volatility, and scalability, which has never been seen earlier. Blockchain, which has been presented as a secure, decentralized, and distributed ledge, r can be used for ensuring satellite security [17,18]. Unlike previous methods, which uploaded data to a cloud server or stored it in a single location, blockchain creates several small chunks of an original block and distributes them to the different parts of networks. Blockchain can be deployed as a means for providing secure transactions in the network infrastructures; on the other hand, machine learning and artificial intelligence methods can be used to prevent the network from violent threats [19,20]. Blockchain can encrypt everything that exits in the network so that the data cannot be altered across the transmission. In the encryption-based security methods, the key management scheme consumes energy in nodes, and it is considered a noticeable challenge. Furthermore, these methods should maintain all the keys in the whole network that waste the limited energy of the nodes [21]. Thus, cryptographic approaches increase network cost and node overhead in order to support internal attacks with cryptographic keys. In this paper, we will discuss alternative security solutions based on Intrusion Detection System (IDS) integrated with machine learning methods. Machine learning is a significant method for security protection, which can provide security orchestration required to detect new threats in the UAV networks supported by satellite communication. Moreover, the IDS is effectively useful as a security scheme to increase accuracy in the networks and protect drones against intruders because both internal and external attacks can be accurately detected by IDS [22,23]. Anomaly detection is one of the IDS approaches for detecting new attacks that had never been seen before that instance. Anomaly detection uses a supervised machine learning (ML) algorithm [24–26] to create normal data behavior patterns. A real-time model was presented for detection of outgoing Denial of Service (DoS) attacks in [27], in which many ML algorithms are used and compared in terms of speed, accuracy, and weighting calculation. Machine learning algorithms are an important approach to handle the security problems in 5G-based UAV networks [28]. ML is a kind of artificial intelligence that applies various learning techniques to train devices without definite programming. ML can be employed efficiently in the UAV networks for the following reasons: (1) There is no need for a mathematical model for complex UAV environments. (2) Some applications, which require datasets, can be correlated. (3) ML algorithms are able to adjust with the dynamics and unforeseen patterns of UAV systems. (4) ML algorithms can eliminate human interventions, which does not fit for the UAV networks. The main contribution of this paper is as follows:

- Initially, we use a dataset named CSE-CIC-IDS2018 [29] on Amazon Web Service (AWS) for training and testing, which are performed once per iteration. The same training and test data are considered for all models to provide a fair comparison between them. Eighty percent of the dataset is assigned to training set, and 20% is assigned to the test set.
- Model creation can be defined as feature selection, implementation, refinement, and comparison. We propose a 5G satellite-based UAV model. We implemented security based on ML algorithms in gateways. To increase the accuracy of our system model as well as implementing it in the real world, we consider the features that we mention here. Some of the most important features include flow duration, total packets in the forward and backward direction, maximum and minimum size of the packet in the forward and backward direction, average and total size of the packet in forward direction, standard deviation packet size in forward and backward direction, etc. We consider zero (0) for normal and one (1) for attack records.
- In order to provide security in this paper, data packets are encrypted initially, and then ML algorithms are used to increase the level of accuracy of packets to identify which one is the correct packet and which one is fake or attack packet. The ML algorithms such as Logistic Regression (LR), Linear Discriminant Analysis (LDA), KNN, Decision Tree (DT), Gaussian Naive Bayes (GNB), Stochastic Gradient Descent (SGD), and K-mean are used.
- Finally, we compare the output of the above ML algorithms for above-mentioned attacks based on their precision, recall, F1-score parameters, accuracy rate, false-negative rate, correctly classified records, and incorrectly classified records, which will be explained in Section 4.

The remainder of the paper is organized as follows. Section 2 describes the background of satellite backhaul connectivity based on the 5G system. Section 3 presents the system model and IDS for satellite-based UAV security. Section 4 is related to an intrusion detection system based on ML approaches to detect various types of attacks. Section 5 provides the experimental results using ML techniques for various attacks. Section 6 provides discussion on the experimental results obtained using ML, while Section 7 presents the future works, and finally, Section 8 provides the conclusion.

2. Background and Related Works

2.1. Satellite Architecture

A Public Land Mobile Network (PLMN) can have both terrestrial 3GPP access and satellite 3GPP access. However, the coverage of the satellite access network may span over the coverage of the terrestrial access network, as shown in Figure 1 [30].

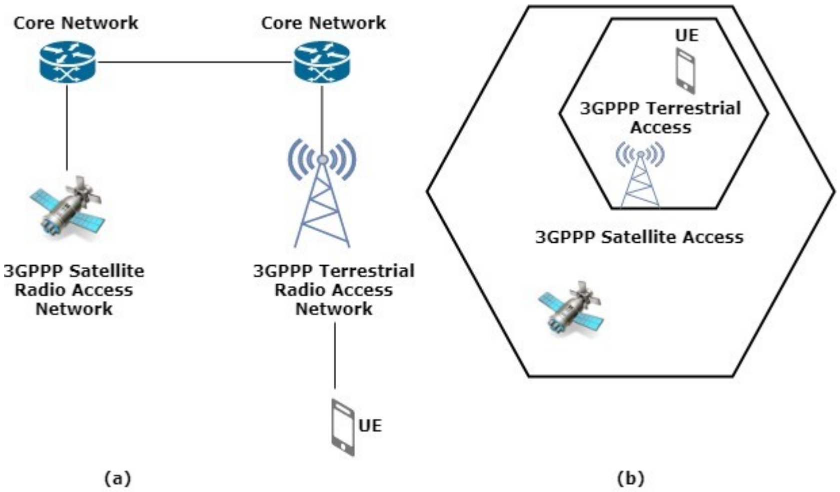

Figure 1. Satellite and terrestrial 3GPP access networks within a PLMN. (**a**) architecture, (**b**) coverage.

A satellite access network is shared between multiple core networks in a 5G Multi-Operator Core Network (MOCN) sharing architecture. In this case, the shared satellite Radio Access Network (RAN) broadcasts the system information for both PLMNs, whose core networks are available. According to Figure 2, these PLMNs might have different Mobile Country Codes (MCCs) [30].

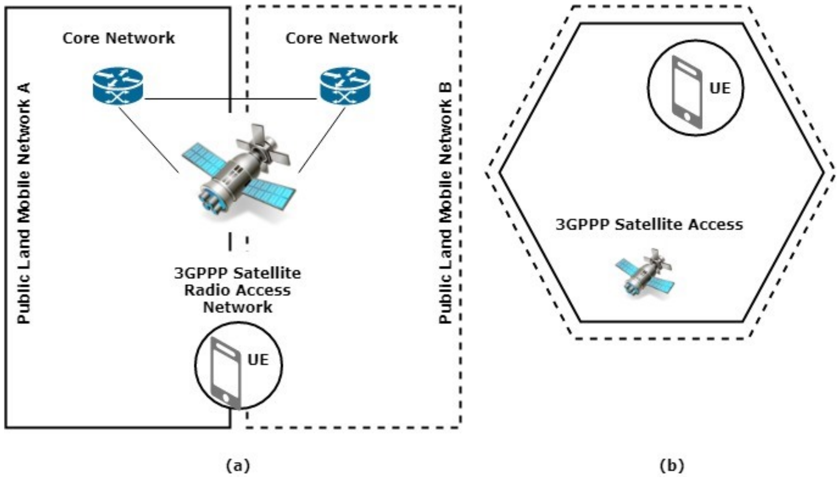
Core Network
Core Network
Public Land Mobile Network A
Public Land Mobile Network B
3GPP Satellite Radio Access Network
UE
(a)
UE
3GPP Satellite Access
(b)

Figure 2. Multi Operator Core Network sharing architecture with satellite radio access network: (**a**) architecture, (**b**) coverage.

A satellite back haul is used between the core and terrestrial access network, providing a backup transport for the N2/N3 reference points as demonstrated in Figure 3. The N2/N3 reference points are generally used for connecting standalone non-3GPP accesses (e.g., WLAN access) to the 5G core network via control plane and user plane functionality, respectively. The User Plane Function (UPF) is one of the Network Functions (NFs) of a 5G core (5GC) network. The 5GC network consists of more than one Access and Mobility

Management Functions (AMFs) and UPFs. The 5GC is linked with the distributed gNB through standard N2 and N3 interfaces [31]. The satellite system transparently carries the communication payload of the 3GPP reference points [30].

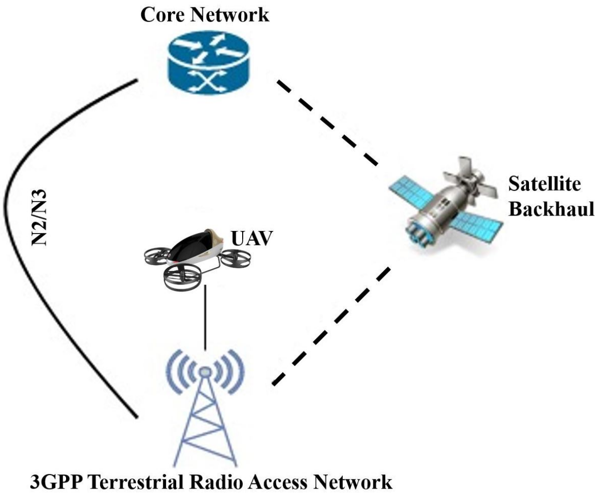

Figure 3. 5G System with a satellite backhaul.

In the case of Non-Geostationary Satellite Orbits (NGSO) such as Low-Earth Orbiting (LEO), Medium-Earth Orbiting (MEO), and Highly Eccentric Orbiting (HEO), the attached cells and tracking areas move with the corresponding gNBs. The NGSO with beam steering is capable of seamless handover from one satellite to another to guarantee the connectivity service for moving gNBs when proceeding on non-geostationary satellites. In satellite access, the one-way propagation delay between a User Equipment (UE) and a satellite communication payload may range between 2 ms and 140 ms according to the satellite altitude and the relative location of the UE. In our case, the UAVs are considered UEs. It is possible that in a constellation of non-geostationary satellites including Inter Satellite Links (ISLs), the delay between a UE and functional elements of the Core Network will increase depending on the actual location of the communication endpoints. The delay also depends on the function and mode of operation of the configuration of the NGSO Access Network. The Non-access Stratum (NAS) is a functional layer that provides communication between the mobile user nodes and the core network nodes. The impacts of delays in the satellite access on the 5G system in the NAS are as follows:

- As mentioned above, the propagation delay between the UE and access node can change significantly, i.e., between 2 ms to 140 ms.
- The need for the 5G core network implies tackling different access capabilities such as propagation delays, coverage, etc., which can satisfy a terrestrial network.
- UEs can utilize the concept of multi-connectivity, which is the capability of supporting simultaneous UEs by multiple sessions that can take advantage of various 3GPP access networks (terrestrial and satellite in the forward and backward direction), as shown in Figure 4.

While 4G is the most common network in the world, it will not be able to manage the huge number of connections that will be on the network in the future, at which point 5G will come into existence. Unique Radio Frequencies (RF) are used by 5G networks to gain what 4G networks were unable to obtain. Each radio spectrum includes several bands from low frequencies to high frequencies that uniquely have particular features.

The 4G network uses frequencies lower than 6 GHz, while 5G uses frequencies from low-band, 600 MHz, to mid-band, and to very high frequencies between 30 GHz to 300 GHz, and they differ from country to country [32]. These low-band frequencies have high 5G coverage and are suitable for rural areas. These high frequencies have great advantages, the most important of which is high capacity and throughput. In comparison to 5G, 4G cells transmit data covering the cell area, which is a waste of both energy and cell power, while 5G cell transmits only small beams in the direction of users only [33]. Furthermore, much smaller antennas are installed in 5G because of shorter wavelengths while still supplying directional control. In terrestrial mobile networks, one BS can effectively be equipped with even more directional antennas for supporting over 1000 additional devices per square meter in comparison with 4G. Thus, many more users can use 5G with enhanced Mobile Broad Band (eMBB), high precision, and very low latency. Moreover, 5G networks can easily receive the required type of data and switch to a lower power when lower rates are needed, and then it switches to a higher-powered mode [34]. A 5G satellite network can be based on a constellation of one or multiple satellites. The satellites are placed in LEO in order to permit connectivity of users, which have constrained RF and energy capabilities. The constellation of satellites may provide a continuous service, with a satellite covering any user with a continuous global coverage. The satellites that are not within range of a ground station can use ISL to communicate (via indirect means) to the ground station. When a UE moves from one static tracking area to another, the tracking area is updated. A Heterogeneous Space and Terrestrial Integrated Networks (H-STIN) architecture has been proposed according to advancement procedure of the UAV, mobile networks, and satellite network [7].

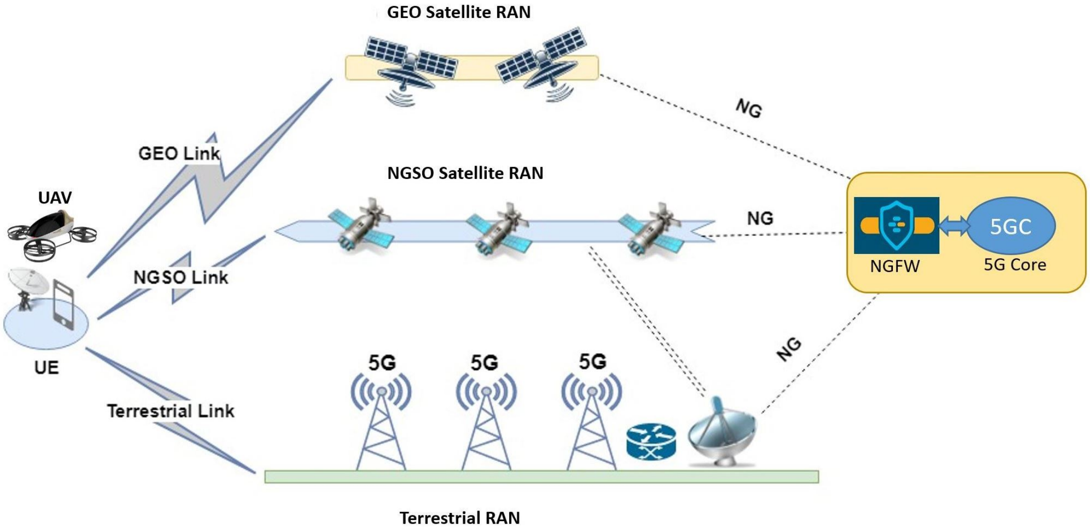

Figure 4. Multi-connectivity architecture with terrestrial, NGSO satellite, and GEO satellite RANs.

The integration of UAVs into cellular networks provides significant advantages with several applications and use cases. With the new paradigm of integrating UAVs with cellular networks, UAVs can be used in two categories. In one category, UAVs can be used as aerial users where the UAVs use the cellular network for connectivity also known as cellular-connected UAVs. The second category is that UAVs can be used as aerial platforms for communication, i.e., UAVs can be used as cellular BS or relays, to provide extended communication to the terrestrial networks and users, known as UAV-assisted wireless communication [35]. In Release 17, there is a 5G enhancement for UAVs to provide extended service to mobile users by using on-board UAV access nodes (UxNB). The UxNB provides extended coverage in scenarios such as natural disasters, temporary coverage for mobile users, and other emergency situations [36]. With the help of on-board UxNB access nodes, the UAV can act as either a base station, where it is connected with the 5G core network, or as a relay, where the UAV is connected with the terrestrial BS to provide extended coverage, as shown in Figure 5. The telecommunications community has acknowledged the importance of providing communication support to low-altitude UAVs in achieving beyond-LoS control and developing a secure communication network. Only terrestrial or satellite communication cannot satisfy the connectivity issues for terrestrial, aerial vehicles, and mobile devices. 5G communication needs to have non-terrestrial support such as integration of satellite system to enhance its communication range and provide guaranteed service. The satellite enhances the 5G system by providing satellite access to allow a radio coverage extension to the terrestrial networks, as well as extension to other 5G terrestrial networks through a roaming agreement. The 5G systems define conditions to avoid instability of the offered Quality of Services (QoS) when switching from the 5G satellite access network to the terrestrial access network and vice versa. Security is a very important issue for a UAV system, where it flies autonomously and beyond LoS communication. The UAV traffic management (UTM) system provides Command and Control (C2), navigation, airspace management, traffic management, route planning, monitoring, etc. to autonomous UAVs [37]. The UTM provides continuous C2 to the autonomous UAVs based on the pre-schedule flying route and monitors its flight status, as shown in Figure 6. However, we will concentrate on cellular-connected UAVs flying at a very low altitude only. A cellular-enabled UAV allows the ground pilot to remotely control and operate the UAV over an LoS range. It also offers an efficient way to establish wireless communication between UAVs, end users, and UAV traffic controllers, regardless of their locations. Even though cellular-enabled UAV communications provide advantages, there are still instances where cellular networks are inaccessible, such as in remote locations, including the sea, desert, or mountains. In such situations, the cellular networks integrated with satellite systems can be used to enable UAV communications outside the terrestrial coverage of cellular networks. In 5G systems, it is possible to integrate satellite communication to extend its connectivity and communicate with UAVs.

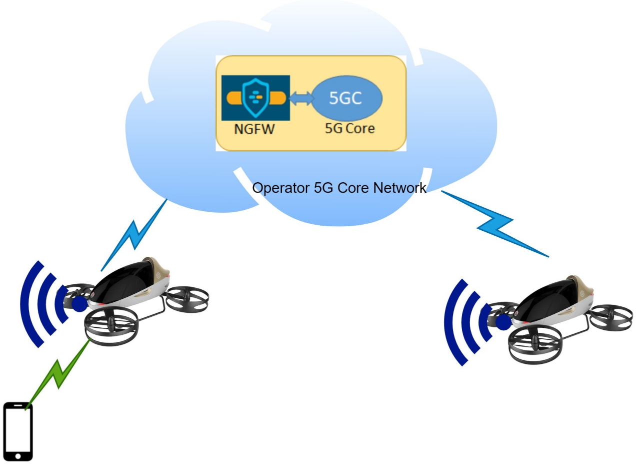

Figure 5. UAVs acting as aerial platform (BS) using UxNB access nodes.

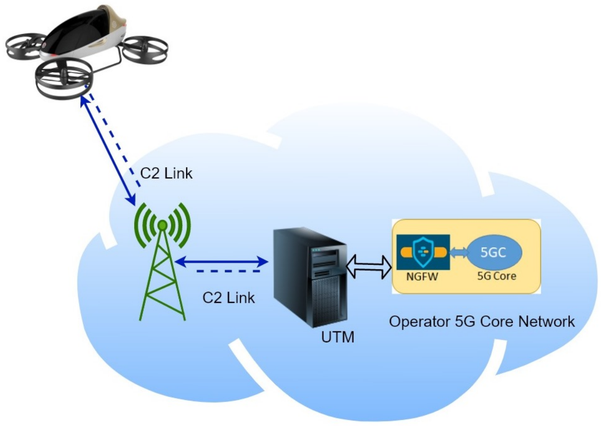

Figure 6. Cellular-connected autonomous UAVs controlled by UTM system.

2.2. Related Works

There are three major types of ML algorithms, namely supervised learning, unsupervised learning, and semi-supervised learning, which can be widely applied in various networks including UAV networks to increase network security. These ML techniques are also used in intrusion detection systems. An IDS has several benefits, including attack detection, protection against violations, and recording existing threats to protect satellite networks. Moreover, it acts as high-quality control for safe format and administration and furnishes useful records about intrusions that occur. There are two main approaches to detect intrusions, and they are based on signature and statistical anomaly. The authors in [38] present an exhaustive survey on IDS based on CICIDS-2018 datasets. The CICIDS2018 is the most comprehensive Big Data, publicly available intrusion detection dataset that encompasses a broad range of types of attacks. These authors examined numerous research papers and compared their performance based on their ML models, computing environments and several performance parameter scores such as accuracy, precision, recall, area under curve, etc. The CSE-CIC-IDS-2018 datasets can be a convincing dataset to evaluate ML-based IDS in UAVs [39,40].

2.2.1. Related Works for CSE-CIC-IDS2018 Dataset

This section summarizes the research that has been done that leverages the CSE-CIC-IDS2018 dataset to employ machine learning techniques. It also gives a quick review of the main machine learning techniques and demonstrates how the CSE-CIC-IDS2018 dataset can be used to evaluate and test different types of machine learning methods. To detect network intrusion traffic and identify attack types, the authors of [41] used a variety of deep learning frameworks. For training and testing, ten-fold cross-validation with an 80–20 or 70–30 split was utilized. The main drawback of this study is the use of only one classifier. In [42], the authors analyzed how well the results of an intrusion detection dataset can be generalized by integrating both CIC-IDS-2017 and CIC-IDS-2018. The authors employed 12 supervised learning algorithms from various families to assess performance. The assumption that some categorical characteristics, such as destination port, have the same number of unique values in both datasets is a shortcoming of this study. This study presents a taxonomy of deep learning intrusion detection models as well as a summary of pertinent research publications. The KDD Cup 1999 [43], NSL-KDD [44], CICIDS2017, and CICIDS2018 datasets were then used to test four deep learning models (feed-forward neural network, auto encoder, deep belief network, and LSTM). The usage of KDD Cup 1999 and NSL-KDD, both of which are outdated and have recognized faults, is one of the study's drawbacks. The biggest issue with KDD Cup 1999 is a large number of duplicate records [44]. NSLKDD is a better variant that avoids the problem of duplicated instances, although it is still far from ideal. For instance, in the NSL-KDD test dataset, some attack classes have no records. In [45], to detect Botnet attacks, the authors trained a two-layer MLP using Python and Scikit-learn. The AUC for this study was one, which is a perfect score. All of the related accuracy, precision, and recall ratings were perfect. The article is four pages in length (with two references), and there is a noticeable lack of depth. Another disadvantage is that it used just one classifier to evaluate the performance. In [46], the authors used DoS datasets from the KDD Cup 1999 and CIC-IDS-2018 to train a CNN. Python and Tensor Flow were used to create the model. The train-to-test ratio was 70–30 for both datasets. The authors employed around 283,000 samples in KDD and approximately 11,000,000 in CIC-IDS-2018. The use of the KDD Cup 1999 dataset, which, as previously mentioned, is an older dataset with a significant number of redundant instances. This is one of the fundamental flaws of the KDD Cup 1999 dataset.

Some papers [47] used outdated datasets to evaluate the IDS system using machine learning such as KDD Cup 1999, NSL-KDD, and ISCX2012. These datasets are obsolete, with a huge number of redundant occurrences compared to the rapid development of new types of network technologies and introduction of newer cybersecurity attacks. Several recent research papers detect IDS and malware utilizing various ML techniques. One of them is [48], which proposed a multi-dimensional feature fusion and stacking ensemble mechanism (MFFSEM) machine learning in Network IDS to detect anomalous behaviors. They used their proposed scheme on multiple feature datasets to achieve global multi-dimensional anomaly detection model in the real world. They claimed that their scheme is superior to other ensemble approaches; however, they used old datasets such as KDD Cup 99, NSL-KDD, UNSW-NB15, and CIC-IDS2017. They also did not include decision-tree-pruning methods or optimal feature selection strategies. The authors in [49] proposed ensemble-based classification using stacked ensemble of dense, convolutional neural networks (CNN), and a meta-learner for malware detection in Windows Portable Executable (WinPE) small operating system. They used Classification of Malware with PE headers (ClaMP) dataset for this type of malware detection. Similarly, the authors in [50] used ensemble-based ML methods such as random forest, extremely randomized tree, and voting mechanism for web injection or webshell detection in lightweight and heavyweight IoT computing scenarios. The authors used 1551 malicious PHP webshells and 2593 normal PHP scripts for IoT security testing. The authors of [51] used DenseNet-based deep learning model to classify malware by handling imbalanced data issues. This model was evaluated on four malware datasets and can detect malwares move efficiently than conventional malware detection. However, this paper needs to improve optimize the false negative rates in detecting the malwares. Most of the above-mentioned related works are based on malware detection, which is similar to host-based intrusion detection that needs an agent or host on the machine. However, network-based intrusion detection is an advanced and as precise detection system that can detect any type of intrusions on any systems (i.e., network- or host-based). They can analyze outgoing and incoming traffic on network interfaces. On the other hand, malware detection has difficulty detecting intrusion based on network traffic only, and if the malware detection host is compromised, then the attacker can disable the malware detection agent.

2.2.2. Recent Public Datasets

Recently, a few newer public datasets available based on network intrusion detection have been introduced. One of them is the Boğaziçi University (BOUN DDoS) dataset, which is of resource-depletion DDoS attacks [52]. It was generated and recorded from router backbone mirrored ports at the Bogazici University campus environment. The datasets include non-attack and attack traffic such as TCP SYN, and UDP flooding packets based on Hping3 traffic generator software by flooding. There are some advantages of BOUN DDoS datasets. It provides simple resource depletion-type DDoS attacks on a campus network, which are suitable for generating and analyzing network-based attack detection methods. It consists of different intensities of attacks to help researchers to train and estimate their IDS methodologies for different attack densities. BOUN dataset consists of genuine background internet traffic combined with DDoS attack traffic. These datasets provide easier simulation and analysis because of small file sizes and fewer packets compared to other datasets, which helps researchers to import datasets in different research software platforms easily. However, BOUN DDoS dataset has limitations when it comes to achieving the task we are trying to solve in this paper. It only consists of only basic attack types such as DDoS attack, TCP SYN flood, UDP flood attack. BOUN DDoS has not been widely adopted by the research community as a benchmark dataset. The datasets have been used in academia but only by the authors who generated them.

Similarly, the LITNET-2020 dataset is a new annotated network benchmark dataset that contains real-world network traffic data and under-attack data samples from the academic networks environment captured over 10 months [53]. It consists of 85 network flow features that can be used to recognize 12 network attacks. The dataset features were analyzed based on statistical analysis and clustering methods. Some advantages of these datasets are as follows: It contains real-world network traffics, unlike other datasets, which were generated synthetically. The datasets are freely available for research purpose and can be used to benchmark network intrusion datasets. The datasets were accumulated over a longer period than other datasets, i.e., 10 months. It is therefore very helpful for researchers and academicians working in the cybersecurity domain. However, for the task we are trying to solve, they have some limitations. These datasets are new datasets that have not yet been widely adopted by the research community as benchmark datasets. For analysis, we used datasets that have been widely accepted, used, and analyzed by researchers and academicians such as CIC-IDS2018 datasets. The LITNET-2020 dataset lacks some of the popular attack types such as DDoS attacks, brute-force attacks, BoTnet, and infiltration attack types. Nevertheless, LITNET-2020 dataset might present an important contribution to the research community by enriching the number of datasets accessible for the development and refinement of new network-attack identification systems. This dataset has the potential to be adopted in new research for NIDS.

3. System Model

There will be a large number of UAVs, terrestrial vehicles, and smart devices in urban cities in the near future, and there are already millions of smart phones. It is important to ensure security of UAVs against attackers: if the attackers compromise UAVs, then they might crash into urban locations, causing serious damage. We need to provide an efficient security mechanism to the UAV system, and there are some requirements to be fulfilled. Thus, the potential requirements of the 5G-satellite system can be defined as follows:

- A 5G system supporting satellite access and massive Machine-Type Communications (mMTC) should also support UAV communication based on the 5G-satellite access network.
- A 5G system should have multiple access points including satellite networks and terrestrial access mobile networks, combined with a machine learning based firewall. In 5G Core (5GC), a machine learning-based, intelligent Next Generation Firewall (NGFW) provides protection across all these access points. Thus, NGFW helps to achieve multiple network slices, as shown in Figure 7.
- One of the requirements is the selection of satellite and terrestrial access networks. The selection should be based on operator policy, subscription settings, QoS settings, and security policies.

The description of our security system model is shown in Figure 7 and is discussed below. In this approach, different types of traffic from various devices/services can be divided into slices, from slice 1 to slice n. As the 5G core and RAN are software-defined, it is feasible to implement NGFW based on machine learning techniques and AI. The AI enabled software-defined help to examine network packet flows for anomalies. In this model, traffic is fed into the firewall component and analyzed with various machine learning techniques. The flows that are identified as anomalies, i.e., the packet flows that behave abnormally, are flagged as malicious, and the policies are updated to terminate these flows. The policy updates are then sent to the SDN controller to terminate the appropriate flows or drop packets. The SDN controller then provides proper routing and management of traffic entering into virtualized core network components.

Some of the key threats based on DDoS and DoS attacks that impact the 5G networks security (including data integrity protection, and data encryption) can be mitigated by using this model. By using a slice-based approach, security policies can be customized and configured based on the sensitivity of the data within the slice. This approach can assist in providing a greater degree of protection for a large variety of services that are expected to operate on 5G networks. The threats and vulnerabilities in 5G and UAV networks are shown in Figure 8. Additionally, by looking at lower-level network traffic such as flow-based statistics and not using deep packet inspection, network traffic can be analyzed in an encrypted state, removing the overhead and additional complexity of decrypting data for analysis and then re-encrypting, which will reduce latency as well. The effectiveness of flow-based analysis using machine learning is demonstrated in the results section.

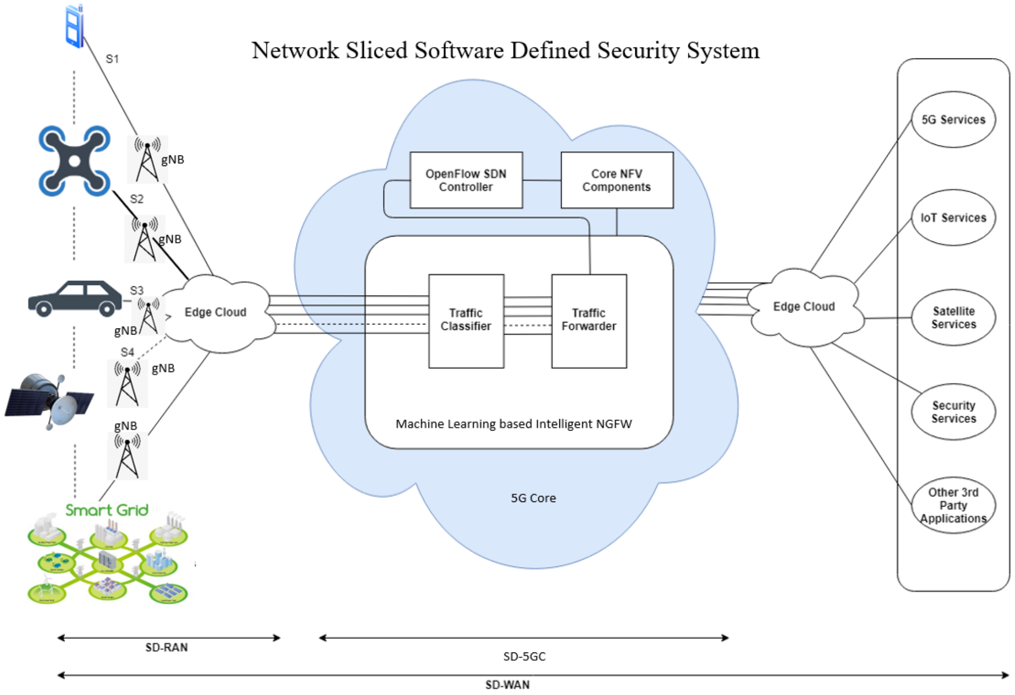

Figure 7. End-to-end model describing slice-based software defined security.

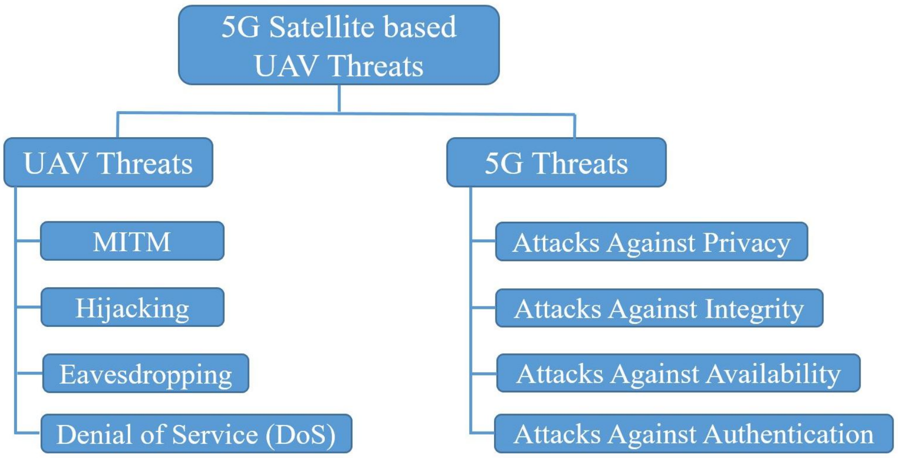
5G Satellite based UAV Threats
UAV Threats
MITM
Hijacking
Eavesdropping
Denial of Service (DoS)
5G Threats
Attacks Against Privacy
Attacks Against Integrity
Attacks Against Availability
Attacks Against Authentication

Figure 8. Threats in the 5G and UAV networks.

3.1. UAV Threats and Vulnerability

Due to the UAV wireless communication system and its unmanned nature, UAV is not free from security attack and vulnerabilities, but instead the security issue that is even more serious. There are several security attacks issues in UAV as well as in the 5G network that have received considerable attention in recent years. UAVs occasionally face security threats of various types, such as malicious messages being sent to UAVs, and hackers interfering with ECUs and attempting to reverse engineer their micro-controllers, software, and so on. We will discuss some of the UAV security issues and threats in this section [54,55].

3.1.1. Man-in-the-Middle (MITM) Attack

In an MIMT attack, the malicious attacker places rogue access point between the endpoints of the target communication; i.e., the attack is carried out on the legitimate Wi-Fi links between the UAVs and the pilot. The attacker can gather active network information using wireless monitoring equipment and then reads and potentially changes the message exchanged between the nodes. As a result, the attacker takes over the UAVs under his control. This attack leads to eavesdropping, hijacking, and data tampering.

3.1.2. Hijacking

The adversary can hijack the radio or connection links between the UAVs and the ground controller by de-authenticating the management frames, which disconnects the connection. As a result, the adversary might take control of the UAVs and operate them according to his or her wish. As a result, this might cause the UAVs to crash or cause serious injuries.

3.1.3. Eavesdropping and Spoofing

Another prominent attack issue is eavesdropping and spoofing when attackers obtain critical information by listening to the communication between source and destination points via spoofing the Address Resolution Protocol (ARP) packets. In UAVs, once a hacker obtains the secret keys of the UAV, the whole device is compromised. The hacker can eavesdrop and steal their data through the open communication channel. In case of spoofing, the intruder will impersonate other UAVs and then take control of the UAV system by providing false information. GPS spoofing is a typical example of a spoofing or forgery attack in UAVs.

3.1.4. Denial of Service (DoS)

In a DoS attack, the attackers flood the controller with numerous requests, causing a network overload that depletes the bandwidth and resources to the UAV. The adversaries might use Telnet software to send several requests to the controller. Thus, the communication between the UAV and its controller is disrupted, and as a result, the UAVs may behave abnormally and might crash. Some other effects of DoS attacks on UAVs can be battery exhaustion, poor performance, latency, and system seize.

3.2. 5G Threats

The 5G telecom network is divided into four major network elements: RAN, core network, transportation network, and interconnection network. Again, each of the network elements consists of three planes for carrying various types of network traffics: control plane, user plane and management plane. These planes are vulnerable to new threats if they are exposed to the attackers. Furthermore, there are threats in 5G cellular networks, which are classified by authentication and privacy approaches, meaning pattern behavior of the attacks in 4G and 5G networks [56]. There are other classifications according to various metrics including passive or active, internal, external, etc. Four clusters of attacks have been described in [12,56], (1) attacks against privacy, (2) attacks against integrity, (3) attacks against availability, and (4) attacks against authentication. All of the mentioned threats are given in Figure 8. Some of the threats in 5G are discussed below.

3.2.1. Attacks against Privacy

In this category, there are fourteen attacks including MITM, eavesdropping, parallel session, reply attack, impersonation attack, collaborated attack, tracing attack, spoofing, privacy violation, adaptive chosen cipher text attack, chosen-plaintext cipher text, stalking, masquerade, and disclosure attacks [57,58]. It should be mentioned that the most important attack among them is MITM, when the false BS acts as a real BS [57].

3.2.2. Attacks against Integrity

According to [56], there are six attacks in this category, which are as follows: tempering attack, message insertion attack, message modification attack, cloning attack, message-blocking attack, and spam attack. The attack against integrity occurs when data are transmitted between the 5G nodes and mobile users although hash functions are mostly used for assuring the integrity of exchanged data.

3.2.3. Attacks against Availability

This category has six classes of attacks, including First in First out (FIFO), redirection, physical, and free-riding attack. This category can make a service such as a data routing service unavailable [56]. The FIFO attack can occur via robust adversary when the entering time and exiting time intervals are gathered. When an adversary obtains the information of the correct user, it can amplify its wrong signal strength to redirect or can impersonate itself as a right BS in 5G cellular networks.

3.2.4. Attacks against Authentication

This category includes ten different types of attacks, which are password reuse, password stealing, dictionary attack, brute force attack, de-synchronization attack, verifier leakage, forgery attack, partial-message collision, and stolen smart card attack. The authentication attack disrupts the authentication of the client to the server and vice versa. Password reuse and password stealing occur when an attacker shows itself as a legitimate user in order to log in to the server by guessing various passwords. In the stolen smart card attack, an attacker can disrupt the smartcard-based user password authentication schemes and then remotely achieve vital information without having access to the real passwords [57].

As a result, in order to provide an acceptable level of security in UAVs, several important factors, including reliable ID, reliable SW, secure configuration, trustworthy data, safe communication, privacy, and physical security, should be take into account, as shown in Figure 9. The 5G systems are expected to provide connectivity and other types of services to a large number of devices simultaneously. Such networks, including UAVs, may send or receive an infrequent or frequent small numbers of data, which are transmitted over the air interface and are vulnerable to eavesdropping. In addition to the protection of small data from eavesdropping at the application layer, it also protects the lower layers and protects against eavesdropping of headers such as IP headers.

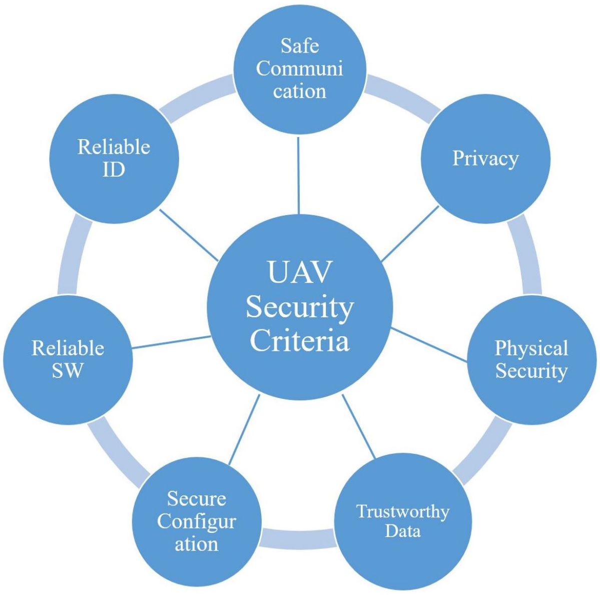

Figure 9. UAV security criteria.

A large number of UAVs acting as UEs performing similar actions at the same time can easily lead to a signaling attack on the network. If such an attack persists and is not dealt with appropriately, it brings a risk for other users in the network. As such, mitigating measures should be designed to protect the network against such attacks. For this key issue, it is assumed that the malicious behavior on the UAV is the result of an attacker with access to the UAV application, which can instruct to make certain requests to the network. An attacker could have obtained this access through the over-the-top service and could for example instruct the UAVs to set up dedicated bearers or request access to certain network slices [59].

The mitigation of the attack by software and appliances is usually deployed at the central position of the architecture. Thus, latency can be seen because network traffic has to be changed and prepared from the initial main path and then sent back to the destination, which is not the optimum approach. Furthermore, other sections such as core routers, switches, and firewalls must be pre-configured to mitigate attacks and allow the traffic diversion after the mitigation has been done. The SDN-based approach is a networking paradigm that has gained traction due to its dynamic functionality in programming networks and increasing network visibility. It is gaining popularity due to its ability to separate control and data planes of the networking infrastructure and assists in minimizing security vulnerabilities in various networks, such as UAV networks. One of the SDN controller advantages is that it knows the network topology and infrastructure and thus can monitor the traffic network. The SDN controller offers integrated security functions, which are routing, firewalling policies, and service chaining enablement, which provides dynamic security in the network via the controller. An NFV can be used in coordination with an SDN to assist in attack avoidance and network analysis. The NFV concepts deploy complicated network functions in commodity hardware and direct the traffic flows to the right network elements through the application of service chaining dynamically. In other words, NFV enables the development of network-based softwareized tools that can enhance in the security of data transmission networks. When anomalies are discovered, NFV functions will be used to mitigate potential threats. The final model is presented as Software-Defined Security (SDSec). According to the proposed architecture, designing a security approach, which protects the systems from DDoS and Malware attacks, becomes more complicated and dynamic. By introducing the concepts of SDN and NFV, the design of SDSec will be as follows:

- The softwareized components of the network, i.e., the NFV version of routers, switches, and firewalls, should be integrated with the SDN environment, which enables the monitoring topology types and manages the softwarized devices (NFV functionalities) directly and indirectly via their own element managers.
- SDN controllers have the capability to control the traffic flows and communication between points and to implement the security policy. Additionally, information on the network and the traffic analytics can be collected and processed by the SDN.
- Network security components can be applied through northbound APIs with the SDN controller in order to detect and respond to spoofing DDoS attacks. The advantage of SDN is that it can efficiently detect the DoS attacks and achieve optimal network wide effectiveness; however, it enforces overhead to the network access as well as overhead to network utilization performance.

3.3. Intrusion Detection System

Using IDS has several benefits, including attack detection and protection against violations, and recording existing threats to protect satellite networks. Moreover, it acts as high-quality control for safe format and administration, furnishes useful records about intrusions that occur. There are two main approaches to detect the intrusion, and they are based on signature and statistical anomaly as shown in Figure 10. The signature-based IDS is able to evaluate the data traffic in the behavior of signature, known identity, or patterns that have similarity with existing signatures. There are many definite and distinguished signatures, which are known for attackers; thus this method can broadly be applied. The statistical anomaly-based technique can be applied for new kinds of attacks; thus much greater overhead and processing capability is required in comparison to the signature-based approach. However, the anomaly-based technique defines and characterizes accurate static form and ideal dynamic behavior of the system. It is popular among researchers due to its potential in detecting new types of attacks efficiently. Basically, there are three types of IDS, and they are classified as network-based, host-based, and application-based.

The network- based IDS (NIDS) can reside on computer or appliances, which are connected to a segment of an organization's network and search for attack patterns when examining packets. Another advantage of NIDS is that it can be installed at a specific place, where it controls incoming and outgoing traffic. The Host-based IDS (HIDS) detects those types of attacks where the intruder creates, modifies, or eliminates the authentic system files or log files. In comparison to NIDS, it can usually be installed at any place, so it provides encrypted information access when transmitting over the network. Application-based IDS (AppIDS) investigated applications consisting of database management systems, content management systems, and accounting systems for abnormal events. In addition, AppIDS can be designed to block requests such as file system, network, configuration and execution space. One of the significant advantages of AppIDS is that it can interact with users and applications as well as operate on incoming encrypted data.

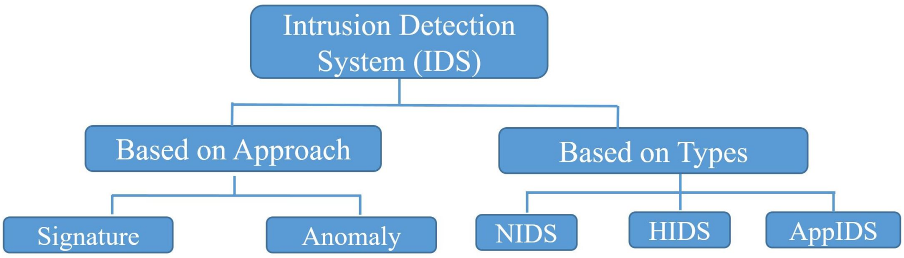

Figure 10. Intrusion detection system classification.

In this paper, we use network-based anomaly detection techniques to detect any new types of intrusion in the UAV networks. The UAVs fly in groups and communicate with each other by sharing critical information such as route information, traffic payload (such as multimedia and images), command and control information, and location information. As a result, it is critical to protect these information exchanges against malicious attackers by using IDS, who might try to leverage the vulnerabilities of wireless networks to disrupt the UAV operations. The anomaly-based technique used in this paper is an attempt to detect all the malicious traffics that harm the networks as well as the UAVs as early as possible to decrease the number of adverse effects. In the next section, we discuss the different types of machine learning techniques to detect various types of attacks in UAV system.

4. ML Approaches to Detect Attacks

There are many ways to create security in UAV networks, and among them, we use anomaly detection using ML algorithms in order to increase the accuracy of 5G-transmitted packets. Anomaly detection is not a novel field of research in machine learning systems, and recent research has focused on a wide range of machine -learning-based applications. First of all, 5G packets are recorded in the network. More clearly, we gather a total number of records, which are divided into two classes, including the number of normal records and attack records. In the considered dataset, 80% is allocated to train ML algorithms and 20% for testing algorithms. After training the mentioned ML algorithms, in testing records, a number of randomly selected packets are used for detecting legitimate or attack packets. Thus, UAVs based on 5G networks can be evaluated by these real data. Moreover, this structure operates like a firewall that controls and eavesdrops on 5G-based UAV data. If controlled data are confirmed as correct, then they can pass through the other nodes of networks, but if they are detected as an attack, then they is not allowed to enter the network. The rest of section provides a review of machine learning algorithms that are applied in this paper. Three major types of ML algorithms, namely supervised learning, unsupervised learning, and semi-supervised learning, can be widely applied in UAV networks in order to increase network security. The ML classifiers used in this paper are: Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-nearest Neighbor (KNN), Decision Tree (DT), Gaussian Naive Bayes (GNB), Stochastic Gradient Descent (SGD), and K-means (K-M). These ML classifiers add a label to the network features of UAV nodes in order to create a classification or regression model [60,61]. A brief description of the mentioned algorithm are discussed as follows:

- LR Algorithm: This method is applied for binary classification problems with two class values. Logistic regression is widely used to evaluate and explain the relationship between a binary real variable such as success or failure and predictor variables. It uses a logistic function for classification logistic regression. Modeling the mean of the response variable for a given set of predictor variables is one of the significant objectives of this algorithm [62].
- LDA: LDA is a well-known method to reduce and classify the projects that have high-dimensional data and create low-dimensional space to efficiently obtain a separate maximum class. In fact, an LDA classifier linearly combines original features. By simultaneously minimizing the samples of a class distance and maximizing distance between class categories, optimal design in LDA algorithm can be achieved [63].
- KNN: For both classification and regression objectives, KNN is an ideal choice in predictive problems. However, most of its applications are related to classification problems in the industry. It has three prominent benefits, including easy interpretation of output, predictive power, and calculation time. The classification is done based on the majority of neighbors of the considered case. This means the case is assigned to the class where the most similarities are observed among its K nearest neighbors, calculated using a distance function.
- DT: In DT, a decision-making method is used that is a tree-like model of the decisions and their potential outcomes that helps to reach a goal. In a DT classifier, a collection of test questions and conditions are designed in a tree shape. The internal nodes in DT include test conditions to divide records, which have different features. A class label including success or failure is assigned to all the terminal nodes. Then, DT recursively selects the best features to separate the data and develops the clusters as the leaf nodes of the tree until its iteration criterion is met. When the decision tree is built, a tree-pruning step can be applied to decrease the size of the decision tree. A decision tree model with many branches and leaves that is too large is known as overfitting [64].
- GNB: Another classification algorithm for binary (two-class) data is Naive Bayes, which is appropriate for multi-class classification problems. Initially, the Gaussian Naive Bayes classifier specifies the total number of classes and then computes the conditional probability for each dataset class. Then, for each feature, the conditional probability can be calculated.
- SGD: A stochastic gradient descent algorithm uses regularized linear models with stochastic gradient descent. In an SGD method, one random point is considered while changing weights, in contrast to gradient descent, which takes into account all of the training data. When there is a huge number of datasets, stochastic gradient descent is the faster choice than gradient descent.
- K-M: K-means clustering is classified as unsupervised learning, and it is used when data are unlabeled, such as data without definite categories or groups. The initial goal of this classifier is to find a cluster in the whole data that the number of clusters represents the variable named K. The algorithm iteratively performs until each data point is assigned to one of K clusters according to the features that are considered. In conclusion, data points are clustered based on the similarities that exist between features.

In this paper, we use datasets from CSE-CIC-IDS 2018 on AWS, which provides a good understanding of intrusion configurations and characteristics. It is a collaborative project between Communications Security Establishment (CSE) and the Canadian Institute for Cybersecurity (CIC) that began in 2018 [29]. A systematic approach was used to generate the datasets for testing, analyzing, and evaluating the IDS considering the network-based anomaly detectors. It uses the profile concept to produce datasets in a systematic manner that gives comprehensive explanations of intrusions as well as abstract distribution models for programs, protocols, or low-level network entities. It provides an extensive benchmark dataset for IDS that comprises representations of events and behaviors observed in the network. Individual operators may use these profiles for a diverse range of network protocols with different topologies to create network events because of the abstract nature of the profiles. The applied dataset has a complete description of intrusions for applications, protocols, or lower-level network entities and is widely used for test and evaluation of intrusion detection algorithms. The dataset consists of six different attack scenarios such as Botnet attack, HTTP denial of service, Collection of web application attacks, infiltration of the network attacks, brute force attacks, and DDoS attacks. A detail information on these attack scenarios can be found in [29]. It includes 6,437,330 normal records and 1,656,840 total attack records, while it has 20% test records and 80% training records. The complete illustration of attack distribution types within CSE-CIC- IDS2018 dataset is demonstrated in Table 1.

Table 1. Table of attack records.

Category	Attack	Number	Train	Test
Botnet	Bot	286,191	838,860	209,715
	Benign	762,384		
Dos	GoldenEye	41,508	1,677,719	419,430
	Hulk	461,912		
	SlowHTTPTest	139,890		
	Slowloris	10,990		
	Benign	1,442,849		
Web	Brute Force-Web	611	1,677,720	419,430
	Brute Force-XSS	230		
	SQL Injection	87		
	Benign	2,096,222		
Infilteration	Infilteration	161,934	755,336	188,835
	Benign	782,237		
BruteForce	FTP-BruteForce	193,360	838,860	209,715
	SSH-Bruteforce	187,589		
	Benign	667,626		
DDos	HOIC	686,012	838,860	209,715
	LOIC-UDP	1730		
	Benign	360,833		

5. Experimental Results

5.1. Experimental Setup

In this sub-section, we discuss the experimental setup. We use Python programming language because, it is easy to use and is a desired application development platform for many application areas. Python has many ML libraries, including internet protocols, string operations, web services tools, and operating system interfaces. In Python, Scikit-Learn is available, whic is an open-source library. It permits implementation of several ML algorithms such as classification and clustering. Some of the libraries used to process the datasets are Pandas, Numpy, and Sklearn. In particular, n-fold cross-validation (usually 10-fold) and train-test split (normally 70–30 or 80–20) are two typical schemes for evaluating machine learning models. When the number of samples in certain categories is small or disproportionate, n-fold cross-validation is typically employed, but the train–test split is often used when the dataset has a significant number of samples in each category. We used cross-validation with 10 folds in this paper. As mentioned in the previous sections, logistic regression, linear discriminant analysis, KNN, decision tree, and Gaussian Naive Bayes have been used and compared in terms of accuracy, precision, recall, F1 Score, false negative rate, etc.

A confusion matrix, which is also known as an error matrix, includes prominent information about real and predicted output classes. A confusion matrix is a table that is used to represent the output of a classification model (or "classifier") on a collection of test data whose true values are known. The outcomes of the classification are divided into two classes, i.e., correct and incorrect classes. The confusion matrix for the intrusion detection is given in Table 2. A confusion matrix with specific layout visualizing the performance of ML algorithm is created for each ML classifier. The main elements of the confusion matrix are presented as follows:

- True Positive (TP): A TP rate shows the number of attack packets that are correctly classified as attacks.
- True Negative (TN): The number of normal packets that correctly classified as normal packets is known as the TN rate.
- False Negative (FN): FN is an incorrect classification where the attack packets are considered normal packets. The FN rate will increase when the number of attack packets that are incorrectly classified as normal packets grows, such it will be anticipated that a serious problem occurs in network resources in terms of confidentiality and availability.
- False Positive (FP): FP refers to when the normal packets are incorrectly classified as the attack packets. The value of FP will grow, which leads to an increase in the computation time. Clearly, the effect of this incorrect classification is less harmful than increasing the FN value.

Table 2. Confusion Matrix.

Confusion Matrix		**Prediction**	
		Positive Class	**Negative Class**
Actual	Normal	TP	FN
	Anomaly	FP	TN

5.2. Results

In general, for IDS, recall and precision values are appropriate choices, but other important valuesl including FP rate and FN rate, are serious factors. In IDS, FN and FP parameters should possibly be reduced, specifically, the FN parameter, which demonstrates that the portion of attacks classified as legitimate packets. According to the definition of precision, when the value of precision is low, it means the classifier has a high percentage of false-positive value. Hence, many normal packets are classified as attack packets, so it has a lower effect in comparison to the FN rate. For a better understanding of recall parameter, the lower percentage of it can be interpreted that the value of FN is high, and thus the huge portion of attacks can be found as a normal packet that shows this kind of classifier has a large value of attack classification process. In terms of F1 score, a higher value of F1 score means fewer incorrectly classified packets (i.e., normal-to-attack and attack-to-normal) and vice versa. Accuracy rate indicates correctly classified normal and attack packet to the total packets.

A Botnet attack is the first attack that was evaluated with the above-mentioned ML algorithms. As can be seen in Table 3, KNN and DT classifier have similar output, but DT has better FNR and incorrectly classified packet. K-M has the worst accuracy rate and the highest incorrect classification value.

In Table 4, we present the evaluation of DoS attacks with respect to various ML algorithms. The DT and KNN indicate high accuracy rates with the highest correctly classified data, respectively. The DT has 0 FNR with 1 precision, recall, and F1-score value.

However, the LDA classifier has a 99.02% accuracy rate, and its FNR is 0.014, which shows that a nearly huge portion of attacks are known as normal packets as compared to DT and KNN. The K-M classifier has the lowest accuracy rate at 37.67% and a high FNR of 0.897, as can be seen from the table.

Table 3. Botnet Attack.

Botnet	**AR**	**P**	**R**	**F1**	**FNR**	**CC**	**IC**
LR	88.06%	0.867	0.666	0.753	0.038	184,692	25,023
LDA	94.45%	0.84	0.984	0.907	0.07	198,095	11,620
KNN	99.99%	1	1	1	0.00004	209,705	10
DT	99.99%	1	1	1	0.00001	209,712	3
GNB	76.15%	0.534	0.999	0.696	0.328	159,708	50,007
SGD	90.30%	0.901	0.726	0.804	0.03	189,386	20,329
K-M	61.93%	0.002	0.001	0.001	0.148	129,891	79,824

Table 4. DoS Attack.

Dos	**AR**	**P**	**R**	**F1**	**FNR**	**CC**	**IC**
LR	87.81%	0.743	0.932	0.827	0.146	368,337	51,093
LDA	99.02%	0.97	0.999	0.985	0.014	415,355	4075
KNN	99.94%	0.999	1	0.999	0.001	419,210	220
DT	99.99%	1	1	1	0	419428	2
GNB	76.95%	0.577	0.981	0.726	0.326	322,792	96,638
SGD	87.66%	0.79	0.824	0.806	0.099	367,706	51,724
K-M	37.67%	0.331	0.982	0.496	0.897	158,004	261,426

Regarding web attack in Table 5, the GNB classifier indicated the lowest accuracy, 20.19%, of all ML techniques, while K-M, SGD, DT, KNN, LDA, and LR had nearly the same AR. However, among them, DT and KNN indicate higher precision values, i.e., 0.962 and 0.904, so there is a lower number of normal packets that are wrongly assigned to the attack class.

Table 5. Web Attack.

Web	**AR**	**P**	**R**	**F1**	**FNR**	**CC**	**IC**
LR	99.95%	0.818	0.049	0.093	0.000005	419,254	176
LDA	99.71%	0.067	0.421	0.116	0.003	418,253	1177
KNN	99.99%	0.904	0.88	0.892	0.00004	419,391	39
DT	99.99%	0.962	0.978	0.97	0.00002	419,419	11
GNB	20.19%	0.001	0.973	0.001	0.798	84,722	334,708
SGD	99.77%	0	0	0	0.002	418,500	930
K-M	99.95%	0	0	0	0.000002	419,246	184

Table 6 illustrates Infiltration attack, where GNB has the maximum FNR value of 0.922 and can be interpreted as large number of attack packets misclassified as normal packets. In contrast to GNB, the LR classifier obtained the lowest FNR, i.e., 0, but at the same time LR also had a lower accuracy rate of 82.76%. That means it had a high FP rate, meaning a huge number of normal packet were interpreted as attack packets. However, LR classifier did not reach the maximum value of the accuracy rate, and it achieved the lowest FN rate and it needed low time demand for creating the training model.

Table 6. Infiltration Attack.

Infiltration	AR	P	R	F1	FNR	CC	IC
LR	82.76%	0.759	0.003	0.005	0	156,292	32,543
LDA	82.86%	0.583	0.026	0.05	0.004	156,475	32,360
KNN	80.29%	0.431	0.443	0.437	0.122	151,630	37,205
DT	86.57%	0.616	0.59	0.603	0.077	163,490	25,345
GNB	22.67%	0.175	0.939	0.295	0.922	42,822	146,013
SGD	82.53%	0.231	0.005	0.01	0.003	155,860	32,975
K-M	74.76%	0.148	0.097	0.118	0.117	141,176	47,659

Table 7 presents Brute Force Attack, where we can see the highest accuracy rate for DT, KNN, and LDA classifier at 99.99%, 99.96%, and 99.77%, respectively. However, among all of them, GNB shows the highest FNR value, 0.64, and has the lowest accuracy rate of 59.26%. Moreover, the LR algorithm has a higher FNR of 0.155, which means that LR can incorrectly classify attack packets as normal packets.

Table 7. Brute Force Attack.

BruteForce	AR	P	R	F1	FNR	CC	IC
LR	81.18%	0.736	0.754	0.745	0.155	170,252	39,463
LDA	99.77%	0.994	1	0.997	0.003	209,252	463
KNN	99.96%	1	1	1	0	209,696	19
DT	99.99%	1	1	1	0	209,715	1
GNB	59.26%	0.472	1	0.641	0.64	124,283	85,432
SGD	88.34%	0.912	0.752	0.824	0.042	185,274	24,441
K-M	63.61%	0	0	0	0	133,413	76,302

For the DDoS attack give in Table 8, all ML algorithms showed excellent performance based on accuracy rate, precision, recall, and F1 score. The K- M classifier performed tbe worst in detecting DDoS attacks based on different metrics, and it was the worst classifier to be used to detect a DDoS attack. It obtained the lowest accuracy rate of 34.39% with an FNR value of 0.005, which can be interpreted as showing that there were a number attacks classified as having a normal value.

Table 8. DDOS Attack.

DDOS	AR	P	R	F1	FNR	CC	IC
LR	99.99%	1	1	1	0.00001	209,714	1
LDA	99.99%	1	1	1	0	209,706	9
KNN	99.99%	1	1	1	0.00003	209,712	3
DT	99.99%	1	1	1	0	209,715	1
GNB	99.99%	1	1	1	0.00006	209,710	5
SGD	99.98%	1	1	1	0	209,680	35
K-M	34.39%	0.488	0.003	0.005	0.005	72,136	137,579

6. Discussion

In this section, we summarize our experimental results to determine the efficiency of ML algorithms to detect various types of attacks based on selected critical IDS parameters such as accuracy, precision, FNR, and F1 score. We present the following results to analyze the best ML algorithm that can be used for IDS.

The analysis of the numerical ML algorithms can be itemized in the following items:

- The DT reached the maximum value for accuracy rate compared to all other types of ML classifiers, as indicated by the yellow bar shown in Figure 11. It obtains a 99.99% accuracy rate for Botnet, Brute force, DoS, DDoS, and Web attacks. DT has the lowest false negative rate value of 0.001%, as shown in Figure 11. KNN had second highest accuracy rate in detecting different types of attacks, as indicated by the gray bar. The accuracy rate of KNN is slightly less than that of DT classifier. On the other hand, GNB has the worst accuracy in detecting all types of attacks. The GNB algorithm shows the lowest average accuracy rate of 20.19% with the smallest precision value of 0.001.

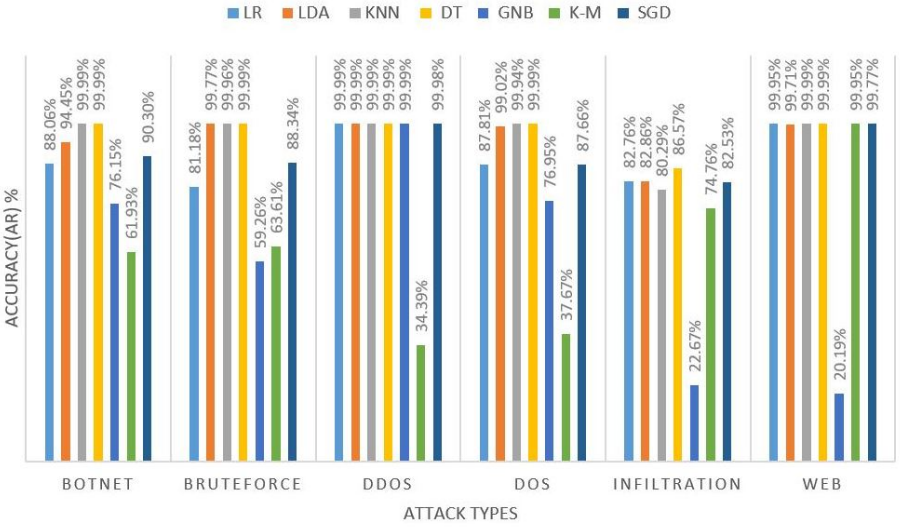

Figure 11. Accuracy of different ML approaches based on various attack types.

- The precision of DT is pretty good in terms of detecting most of the attack types, as can be seen from Figure 12. However, its precision performance is quite lower than that of other ML algorithms in detecting infiltration attack. The GNB classifier precision is lowest among all other ML classifiers. In detecting the DDoS attack, all the machine learning classifiers perform well and have higher precision rates except the GNB classifier.

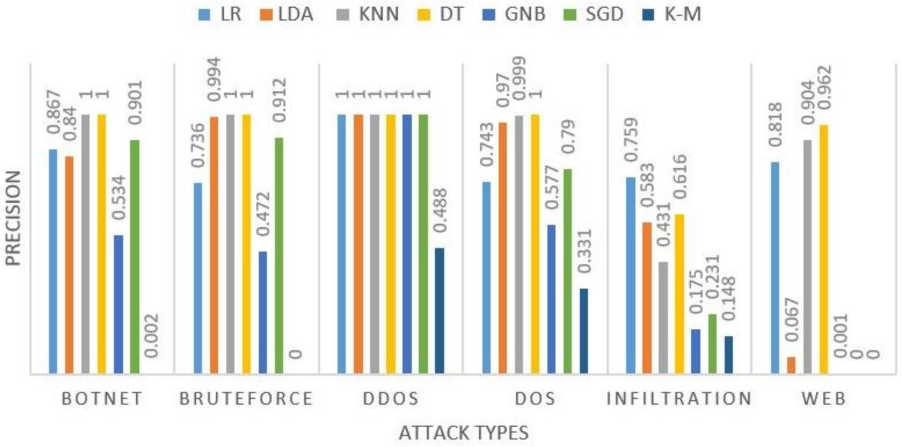

Figure 12. Precision of different ML approaches based on various attack types.

- The FNR is one of the most important parameters in evaluating the IDS. The lower the FNR, the better it is. In Figure 13, the GNB shows the worst FNR performance in detecting various attacks. The LR and K-means algorithm also performs badly in terms of FNR against various attacks. The K-means performs the worst in detecting DoS attacks, with the highest FNR value of 0.897. Most of the ML algorithms have a higher FNR in the case of infiltration attack; however, K-means performs better than any other ML algorithms, with an FNR value of 0.148.

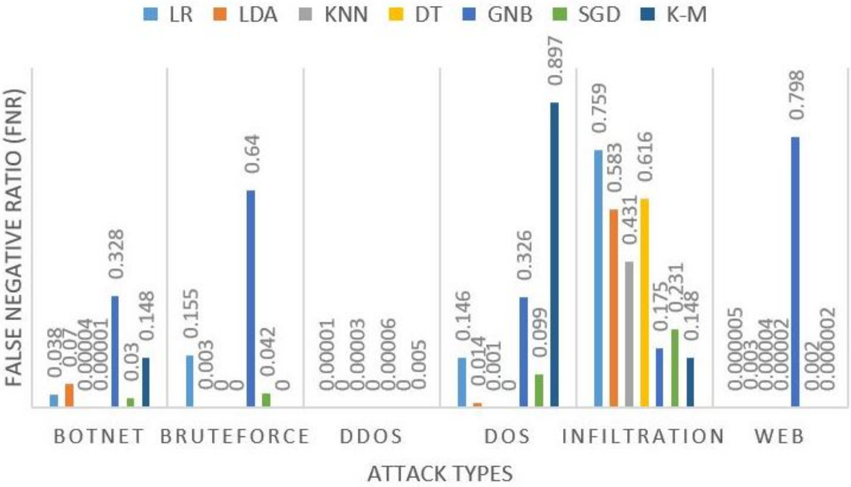

Figure 13. False Negative Rate (FNR) of different ML approaches based on various attack types.

- Similarly, in the case of F1 Score, the higher value of F1 score represents a lower rate of incorrect classified packets; i.e., higher the F1 score, the better it is. F1 is considered the best when its F1 score is 1, whereas the model is a failure when the F1 score is 0. The DT has the highest F1 score value in detecting various types of attacks. KNN also performs well compared to the DT classifier, while LDA performs slightly lower than DT and KNN. However, K-means and GNB has the lowest F1 score value compared to other ML algorithms, as can be seen in Figure 14.

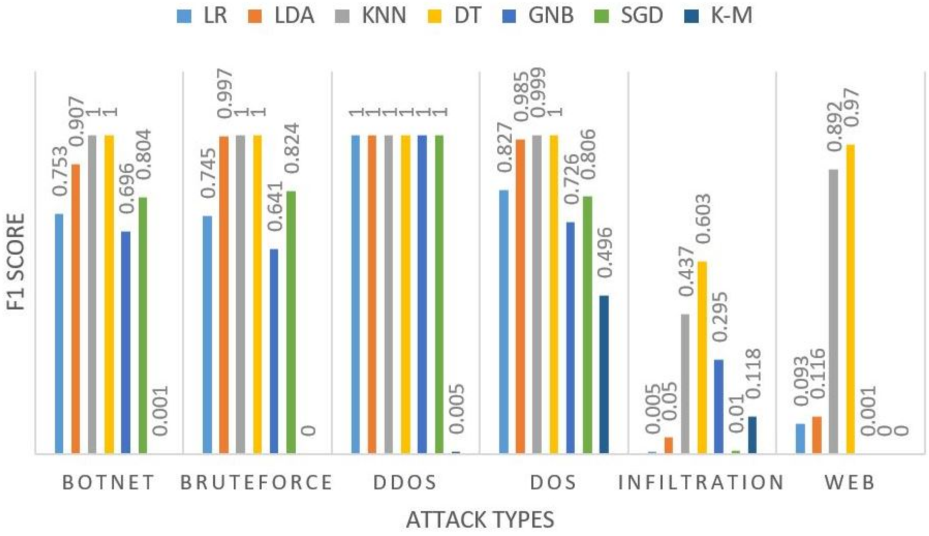

Figure 14. F1 Score of different ML approaches based on various attack types.

Thus, the overall performance of DT classifier is better than any other ML classifier for ML based IDS. It shows the maximum percentage for detecting normal packets correctly followed by KNN. There was no considerable difference between KNN and K-means classifier based on FNR parameters in case of infiltration, Brute force and web attacks. The K-means and GNB performs worst among all other ML algorithms.

We also compared our results with the results of other authors' work on the same benchmark dataset, i.e., CSE CIC-IDS2018. We mainly compared our results based on accuracy obtained using Botnet attacks. Botnet attacks are common attack used by all other authors. All the authors used various ML algorithms to detect specific attack cases. A comparison is given in Table 9. It should be noted that the result for DT is as achieved by the authors in [65], with a 99.99% accuracy rate for the Botnet attack. Moreover, the accuracy rate for KNN is 99.984% in [65], which is same as our result. The result achieved by the authors in [66] is similar to our case, which is about 99.99% for KNN and DT and almost same as our result, but our results are slightly better than [9] in the case of the LDA algorithm.

Table 9. Accuracy comparison based on Botnet attack on the same benchmark dataset.

Ref.	Authors	DT	KNN	LDA	GNB
[65]	Huancayo et al.	99.99	99.98	—	90.24
[66]	Karatas et al.	99.99	99.97	93.34	99.94
[67]	Qusyairi et al.	98.6	—	—	73.6
[68]	LinPeng et al.	96.2	—	—	—
[69]	Khan	97.75	—	—	—
Ours		99.99	99.99	94.45	76.15

The ROC curve is a graphical approach for displaying the trade-off between the true-positive rate and the false-positive rate of a model. The area under the ROC curve (AUC) is its quantitative indication, and it indicates how well the identified model performs. To make the detection effect more clear, the Receiver Operating Characteristic (ROC) curves of the ML models for DDoS are presented in Figure 15. The areas under the ROCs for most of ML technieques are quite good at detecting DDoS, except the K-means algorithm. The ROC curve for BotNet attack is given in Figure 16. The DT performs better in detecting the BotNet, while K-means, GNB, and LR perform worst among other ML techniques. Similarly, Figure 17 shows the ROC for BruteForce attack, the performance of K-means, GNB, and LR is worst compared to other ML techniques. Similarly, Figures 18 and 19 show the AUC curve of web attack and DoS attack, and in both cases, the DT performs best among all other ML techniques.

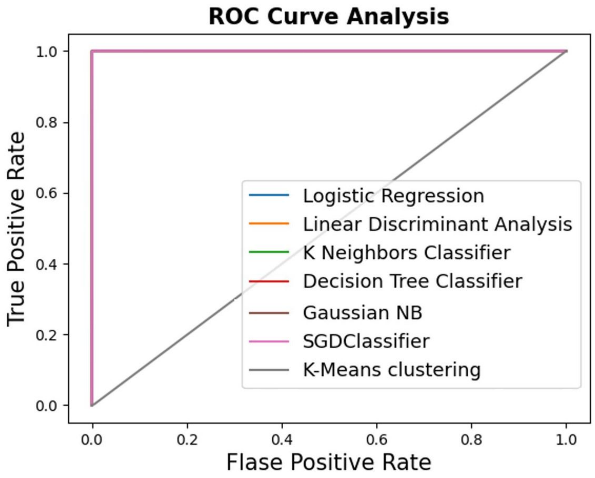

Figure 15. The ROC curves of ML models for DDoS.

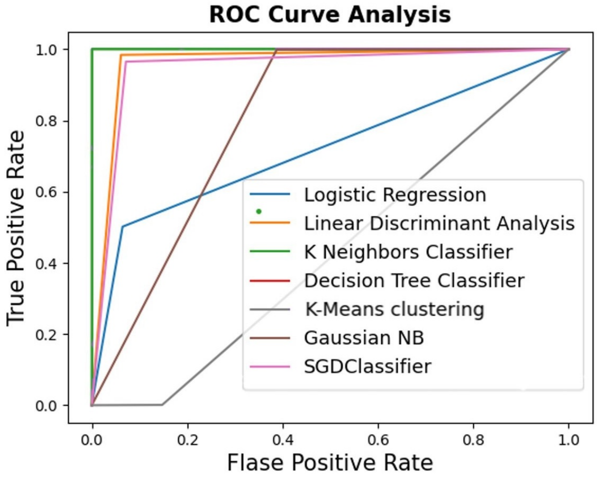

Figure 16. The ROC curves of ML models for Botnet.

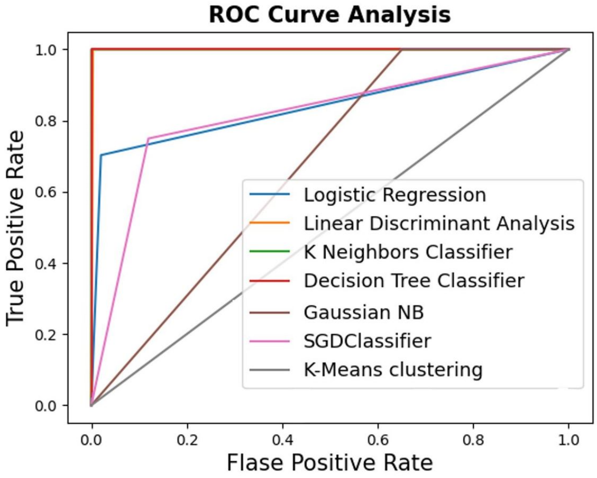

Figure 17. The ROC curves of ML models for Bruteforce.

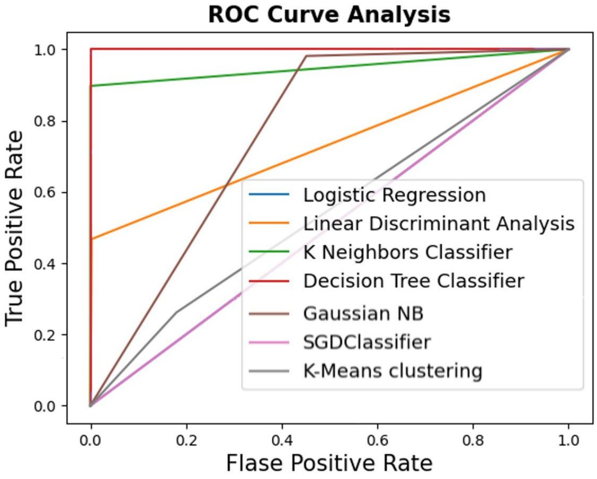

Figure 18. The ROC curves of ML models for.

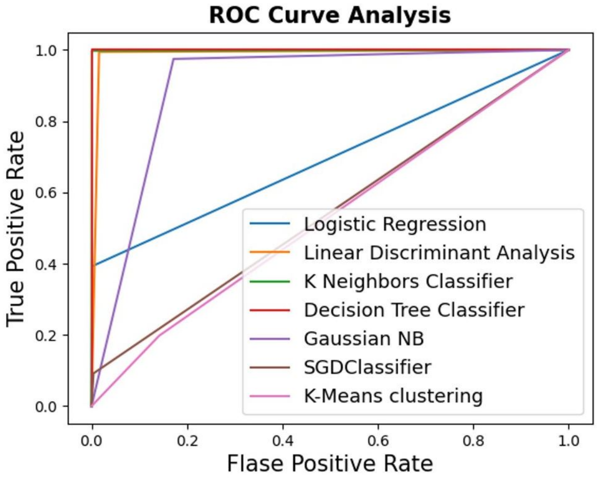

Figure 19. The ROC curves of ML models for DoS.

7. Future Works

In this paper, we have discussed the IDS based on ML techniques for 5G satellite-connected UAV networks to provide secure communication. However, in the future, the sky will be filled with massive numbers of UAVs and other flying objects such as flying taxis and air cargo vehicles at different airspace levels. Thus, for massive UAV connections, and for providing seamless connectivity, communication beyond 5G and 6G will be used. These new communication technologies can help in UAV traffic management systems in urban scenarios. However, security will still be an important issue. The 5G Ultra-Reliable Low-Latency Communications (URLLC) applications provide temporal and short packet transmission, achieving 99.999% reliability with 1ms latency. However, UAVs require higher reliability and very low latency for communication and control of the aircraft for real-time applications such as mission-critical applications. To overcome these issues, new adaptation and learning capability in machine learning methods (e.g., artificial neural networks) along with communication beyond 5G and 6G will be required. At the same time, more advanced software and artificial-intelligence-defined security algorithms will be needed that can identify the attacks and counter them in an optimum way. Moreover, current SDN and NFV ideas must be enhanced with embedded intelligence for robustness to meet the objectives beyond 5G and 6G [70,71]. In this context, the security mechanism in containerized Virtual Network Function (VNF) boxes in gateways will observe 6G traffic based on new ML techniques that will help to detect threats and mitigate attacks. Thus, in the future, we will study how technology beyond 5G and 6G will incorporate the concept of NFV, SDN, and ML to provide essential service for UAVs and how they provide efficient end-to-end network security based on IDS.

8. Conclusions

We designed a model for a 5G software-defined security system to show the benefits of machine learning in a satellite and UAV network for threat detection. We used various types of ML algorithms in networked based intrusion detection to detect new types of intrusion in the UAV networks. The efficiency and performance of various ML algorithms has been verified based on different parameters. The results demonstrate that there is no unique machine-learning algorithm that succeeds in preventing all types of attacks. However, the decision tree obtained the minimum value of false negative rate of 0% with a maximum accuracy of 99.99% for all types of tested attacks, except infiltration, which had 86.57% accuracy. Among all ML classifiers, Gaussian Naive Bayes reached the lowest accuracy rate and the maximum false negative rate. Furthermore, to detect the intrusion in the network, the FN rate is very significant to provide availability and the confidentiality in addition to precision, recall and the accuracy rate parameters. These results show a promising capability for the application of ML in network threat detection for cellular-based UAVs and satellite networks.

Author Contributions: Conceptualization, R.S., A.O., S.A.R. and R.A.; writing—original draft preparation, R.S., A.O. and S.A.R.; investigation, R.S., R.A. and S.K.; supervision, R.A. and S.K.; writing—review and editing, R.S., R.A. and S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Korean government (MSIT) (No. 2021-0-01352) through Institute for Information & Communications Technology Planning & Evaluation (IITP) for development of technology to validate autonomous driving service and the APC was funded by the same funding source.

Acknowledgments: This work was supported by Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korean government (MSIT) (No. 2021-0-01352, Development of technology for validating the autonomous driving services in perspective of laws and regulations).

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

References

1. Ippolito, L.J. Introduction to Satellite Communications. In *Satellite Communications Systems Engineering: Atmospheric Effects, Satellite Link Design and System Performance*; Wiley: Hoboken, NJ, USA, 2017; pp. 1–16. [CrossRef]
2. Seeber, G. *Satellite Geodesy: Foundations, Methods, and Applications*; Walter de Gruyter: Berlin, Germany, 2008.
3. Chaisatien, W. *The 6 Golden Rules for Digital Transformation Success: Strategies and Insights from Industry Leader*; Technical Report; Ericsson: Stockholm, Sweden, 2021.
4. De Sanctis, M.; Cianca, E.; Araniti, G.; Bisio, I.; Prasad, R. Satellite communications supporting internet of remote things. *IEEE Internet Things J.* **2016**, *3*, 113–123. [CrossRef]
5. Qu, Z.; Zhang, G.; Cao, H.; Xie, J. LEO satellite constellation for internet of things. *IEEE Access* **2017**, *5*, 18391–18401. [CrossRef]
6. Siris, V.A.; Thomas, Y.; Polyzos, G.C. Supporting the iot over integrated satellite-terrestrial networks using information-centric networking. In Proceedings of the 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Larnaca, Cyprus, 21–23 November 2016; pp. 1–5.
7. Chien, W.C.; Lai, C.F.; Hossain, M.S.; Muhammad, G. Heterogeneous Space and Terrestrial Integrated Networks for IoT: Architecture and Challenges. *IEEE Netw.* **2018**, *33*, 15–21. [CrossRef]
8. Chelle, H.; Crosnier, M.; Dhaou, R.; Beylot, A.L. Adaptive load control for IoT based on satellite communications. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–7.
9. Mukherjee, J.; Ramamurthy, B. Communication technologies and architectures for space network and interplanetary internet. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 881–897. [CrossRef]
10. Giambene, G.; Kota, S.; Pillai, P. Satellite-5G Integration: A Network Perspective. *IEEE Netw.* **2018**, *32*, 25–31. [CrossRef]
11. Boero, L.; Bruschi, R.; Davoli, F.; Marchese, M.; Patrone, F. Satellite Networking Integration in the 5G Ecosystem: Research Trends and Open Challenges. *IEEE Netw.* **2018**, *32*, 9–15. [CrossRef]
12. Curry, T.; Abbas, R. 5G Coverage, Prediction, and Trial Measurements. *arXiv* **2020**, arXiv:2003.09574.
13. Wang, X.; Du, J.; Wang, J.; Zhang, Z.; Jiang, C.; Ren, Y. Key issues of security in space-based information network review. In Proceedings of the International Conference on Cyberspace Technology (CCT 2014), Beijing, China, 8–10 November 2014; pp. 1–6.
14. He, D.; Li, X.; Chan, S.; Gao, J.; Guizani, M. Security Analysis of a Space-Based Wireless Network. *IEEE Netw.* **2018**, *33*, 36–43. [CrossRef]
15. Mamdouh, M.; Elrukhsi, M.A.; Khattab, A. Securing the Internet of Things and Wireless Sensor Networks via Machine Learning: A Survey. In Proceedings of the 2018 International Conference on Computer and Applications (ICCA), Beirut, Lebanon, 25–26 August 2018; pp. 215–218.
16. Lam, J.; Abbas, R. Machine Learning based Anomaly Detection for 5G Networks. *arXiv* **2020**, arXiv:2003.03474v1.
17. Singh, M.; Kim, S. Chapter Four—Blockchain technology for decentralized autonomous organizations. In *Role of Blockchain Technology in IoT Applications*; Kim, S., Deka, G.C., Zhang, P., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; Volume 115, pp. 115–140. [CrossRef]
18. Singh, M.; Kim, S. Branch based blockchain technology in intelligent vehicle. *Comput. Netw.* **2018**, *145*, 219–231. [CrossRef]
19. Shrestha, R.; Nam, S.Y.; Bajracharya, R.; Kim, S. Evolution of V2X Communication and Integration of Blockchain for Security Enhancements. *Electronics* **2020**, *9*, 1338. [CrossRef]
20. Shrestha, R.; Nam, S.Y. Regional Blockchain for Vehicular Networks to Prevent 51. *IEEE Access* **2019**, *7*, 95033–95045. [CrossRef]
21. Usman, M.; Ahmed, I.; Aslam, M.I.; Khan, S.; Shah, U.A. SIT: A lightweight encryption algorithm for secure internet of things. *arXiv* **2017**, arXiv:1704.08688.

22. Raza, S.; Wallgren, L.; Voigt, T. SVELTE: Real-time intrusion detection in the Internet of Things. *Ad Hoc Netw.* **2013**, *11*, 2661–2674. [CrossRef]
23. Shrestha, R.; Han, K.H.; Choi, D.Y.; Han, S.J. A Novel Cross Layer Intrusion Detection System in MANET. In Proceedings of the 2010 24th IEEE International Conference on Advanced Information Networking and Applications, Perth, WA, Australia, 20–23 April 2010; pp. 647–654. [CrossRef]
24. Sedjelmaci, H.; Senouci, S.M.; Feham, M. An efficient intrusion detection framework in cluster-based wireless sensor networks. *Secur. Commun. Netw.* **2013**, *6*, 1211–1224. [CrossRef]
25. Faisal, M.A.; Aung, Z.; Williams, J.R.; Sanchez, A. Data-stream-based intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study. *IEEE Syst. J.* **2015**, *9*, 31–44. [CrossRef]
26. Rajasegarar, S.; Leckie, C.; Palaniswami, M. Anomaly detection in wireless sensor networks. *IEEE Wirel. Commun.* **2008**, *15*, 34–40. [CrossRef]
27. Doshi, R.; Apthorpe, N.; Feamster, N. Machine learning ddos detection for consumer internet of things devices. In Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 24 May 2018; pp. 29–35.
28. Alsheikh, M.A.; Lin, S.; Niyato, D.; Tan, H.P. Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1996–2018. [CrossRef]
29. Sharafaldin, I.; Habibi Lashkari, A.; Ghorbani, A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy—ICISSP, INSTICC, SciTePress, Funchal, Portugal, 22–24 January 2018; pp. 108–116. [CrossRef]
30. 3GPP. *Study on Architecture Aspects for Using Satellite Access in 5G*; Technical Specification (TS) 23.737; Release 16; 3rd Generation Partnership Project (3GPP): Nice, France, 2018.
31. 3GPP. *Technical Specification Group Services and System Aspects; Study on Architecture Aspects for Using Satellite Access in 5G [Rel. 17]*; Technical Report; 3rd Generation Partnership Project (3GPP): Nice, France, 2019.
32. Bae, J.; Choi, Y.S.; Kim, J.S.; Chung, M.Y. Architecture and performance evaluation of MmWave based 5G mobile communication system. In Proceedings of the 2014 International Conference on Information and Communication Technology Convergence (ICTC), Busan, Korea, 22–24 October 2014; pp. 847–851.
33. Felita, C.; Suryanegara, M. 5G key technologies: Identifying innovation opportunity. In Proceedings of the 2013 International Conference on QiR, Yogyakarta, Indonesia, 25–28 June 2013; pp. 235–238.
34. Hossain, E.; Hasan, M. 5G cellular: Key enabling technologies and research challenges. *arXiv* **2015**, arXiv:1503.00674.
35. Zeng, Y.; Wu, Q.; Zhang, R. Accessing from the Sky: A Tutorial on UAV Communications for 5G and Beyond. *arXiv* **2019**, arXiv:1903.05289.
36. 3GPP. Unmanned Aerial Systems over 5G. In *The Mobile Broadband Standard*; Technical Report; 3rd Generation Partnership Project (3GPP): Nice, France, 2019.
37. Shrestha, R.; Bajracharya, R.; Kim, S. 6G Enabled Unmanned Aerial Vehicle Traffic Management: A Perspective. *IEEE Access* **2021**. [CrossRef]
38. Leevy, J.L.; Khoshgoftaar, T.M. A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data. *J. Big Data* **2020**, *7*, 104. [CrossRef]
39. Ferrag, M.A.; Maglaras, L. DeliveryCoin: An IDS and Blockchain-Based Delivery Framework for Drone-Delivered Services. *Computers* **2019**, *8*, 58. [CrossRef]
40. Ferrag, M.A.; Maglaras, L.; Moschoyiannis, S.; Janicke, H. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *J. Inf. Secur. Appl.* **2020**, *50*, 102419. [CrossRef]
41. Basnet, R.B.; Shash, R.; Johnson, C.; Walgren, L.; Doleck, T. Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks. *J. Internet Serv. Inf. Secur. (JISIS)* **2019**, *9*, 1–17.
42. D'hooge, L.; Wauters, T.; Volckaert, B.; De Turck, F. Inter-dataset generalization strength of supervised machine learning methods for intrusion detection. *J. Inf. Secur. Appl.* **2020**, *54*, 102564. [CrossRef]
43. KDD. *KDD Cup*; Technical Report; KDD: Washington, DC, USA, 1999.
44. Tavallaee, M.; Bagheri, E.; Lu, W.; Ghorbani, A.A. A detailed analysis of the KDD CUP 99 data set. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, 8–10 July 2009; pp. 1–6. [CrossRef]
45. Kanimozhi, V.; Jacob, T.P. Artificial Intelligence based Network Intrusion Detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. *ICT Express* **2019**, *5*, 211–214. [CrossRef]
46. Kim, J.; Kim, J.; Kim, H.; Shim, M.; Choi, E. CNN-Based Network Intrusion Detection against Denial-of-Service Attacks. *Electronics* **2020**, *9*, 916. [CrossRef]
47. Gamage, S.; Samarabandu, J. Deep learning methods in network intrusion detection: A survey and an objective comparison. *J. Netw. Comput. Appl.* **2020**, *169*, 102767. [CrossRef]
48. Zhang, H.; Li, J.L.; Liu, X.M.; Dong, C. Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection. *Future Gener. Comput. Syst.* **2021**, *122*, 130–143. [CrossRef]
49. Damaševičius, R.; Venčkauskas, A.; Toldinas, J.; Grigaliūnas, Š. Ensemble-Based Classification Using Neural Networks and Machine Learning Models for Windows PE Malware Detection. *Electronics* **2021**, *10*, 485. [CrossRef]

50. Yong, B.; Wei, W.; Li, K.C.; Shen, J.; Zhou, Q.; Wozniak, M.; Połap, D.; Damaševičius, R. Ensemble machine learning approaches for webshell detection in Internet of things environments. *Trans. Emerg. Telecommun. Technol.* **2020**, e4085. Available online: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ett.4085 (accessed on 9 April 2021). [CrossRef]
51. Hemalatha, J.; Roseline, S.A.; Geetha, S.; Kadry, S.; Damaševičius, R. An Efficient DenseNet-Based Deep Learning Model for Malware Detection. *Entropy* **2021**, *23*, 344. [CrossRef]
52. Erhan, D.; Anarım, E. Boğaziçi University distributed denial of service dataset. *Data Brief* **2020**, *32*, 106187. [CrossRef]
53. Damasevicius, R.; Venckauskas, A.; Grigaliunas, S.; Toldinas, J.; Morkevicius, N.; Aleliunas, T.; Smuikys, P. LITNET-2020: An Annotated Real-World Network Flow Dataset for Network Intrusion Detection. *Electronics* **2020**, *9*, 800. [CrossRef]
54. Hodo, E.; Bellekens, X.; Hamilton, A.; Dubouilh, P.L.; Iorkyase, E.; Tachtatzis, C.; Atkinson, R. Threat analysis of IoT networks using artificial neural network intrusion detection system. In Proceedings of the 2016 International Symposium on Networks, Computers and Communications (ISNCC), Yasmine Hammamet, Tunisia, 11–13 May 2016; pp. 1–6.
55. Shrestha, R.; Oh, I.; Kim, S. A Survey on Operation Concept, Advancements, and Challenging Issues of Urban Air Traffic Management. *Front. Future Transp. Syst. Model.* **2021**, 1–27. [CrossRef]
56. Ferrag, M.A.; Maglaras, L.; Argyriou, A.; Kosmanos, D.; Janicke, H. Security for 4G and 5G cellular networks: A survey of existing authentication and privacy-preserving schemes. *J. Netw. Comput. Appl.* **2018**, *101*, 55–82. [CrossRef]
57. Conti, M.; Dragoni, N.; Lesyk, V. A survey of man in the middle attacks. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2027–2051. [CrossRef]
58. Shrestha, R.; Djuraev, S.; Nam, S.Y. Sybil attack detection in vehicular network based on received signal strength. In Proceedings of the 2014 International Conference on Connected Vehicles and Expo (ICCVE), Vienna, Austria, 3–7 November 2014; pp. 745–746. [CrossRef]
59. 3GPP. *Study on Evolution of Cellular IoT Security for the 5G System*; Technical Specification (TS) 33.861; Release 16; 3rd Generation Partnership Project (3GPP): Nice, France, 2018.
60. Xiao, L.; Wan, X.; Lu, X.; Zhang, Y.; Wu, D. IoT security techniques based on machine learning. *arXiv* **2018**, arXiv:1801.06275.
61. Banerjee, N.; Giannetsos, T.; Panaousis, E.; Took, C.C. Unsupervised Learning for Trustworthy IoT. In Proceedings of the 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
62. Tjur, T. Coefficients of determination in logistic regression models—A new proposal: The coefficient of discrimination. *Am. Stat.* **2009**, *63*, 366–372. [CrossRef]
63. Ye, J. Least squares linear discriminant analysis. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 1087–1093.
64. Tan, P.N.; Steinbach, M.; Kumar, V. Classification: Alternative techniques. In *Introduction to Data Mining*; Pearson Addison-Wesley: Boston, MA, USA, 2005; pp. 207–315.
65. Huancayo Ramos, K.S.; Sotelo Monge, M.A.; Maestre Vidal, J. Benchmark-Based Reference Model for Evaluating Botnet Detection Tools Driven by Traffic-Flow Analytics. *Sensors* **2020**, *20*, 4501. [CrossRef] [PubMed]
66. Karatas, G.; Demir, O.; Sahingoz, O.K. Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset. *IEEE Access* **2020**, *8*, 32150–32162. [CrossRef]
67. Fitni, Q.R.S.; Ramli, K. Implementation of Ensemble Learning and Feature Selection for Performance Improvements in Anomaly-Based Intrusion Detection Systems. In Proceedings of the 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia, 7–8 July 2020; pp. 118–124. [CrossRef]
68. Lin, P.; Ye, K.; Xu, C.Z. Dynamic Network Anomaly Detection System by Using Deep Learning Techniques. In *Cloud Computing—CLOUD 2019*; Da Silva, D., Wang, Q., Zhang, L.J., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 161–176.
69. Khan, M.A. HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System. *Processes* **2021**, *9*, 834. [CrossRef]
70. Ylianttila, M.; Kantola, R.; Gurtov, A.; Mucchi, L.; Oppermann, I.; Yan, Z.; Nguyen, T.H.; Liu, F.; Hewa, T.; Liyanage, M.; et al. 6G White paper: Research challenges for Trust, Security and Privacy. *arXiv* **2020**, arXiv:2004.11665.
71. Ali, S.; Saad, W.; Rajatheva, N.; Chang, K.; Steinbach, D.; Sliwa, B.; Wietfeld, C.; Mei, K.; Shiri, H.; Zepernick, H.J.; et al. 6G White Paper on Machine Learning in Wireless Communication Networks. *arXiv* **2020**, arXiv:2004.13875.