Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777649

A STUDY OF UAVs DETECTION AND TRACKING MODE TOWARDS MACHINE LEARNING-BASED WI-FI TRAFFIC

Akshada Kulkarni and Dr. Atul Dattatrye Newase
Department of Computer Application, Dr. A. P. J. Abdul Kalam University, Indore (M.P.), India 452010, Corresponding Author Email : akshada.ma@gmail.com

Abstract :
Recent years have seen considerable growth for consumer unmanned aerial vehicles (UAVs). Consumer UAVs, while their enormous economic development potential, offer significant security and privacy concerns due to the fact that they enable a wide range of applications. Both for invasion detection and forensics, it is essential that intruding UAVs be quickly detected and identified in order to reduce the dangers. We propose a machine learning-based framework for rapid UAV identification via encrypted Wi-Fi data to supplement the current physical detection methods. Because many consumer UAVs utilise Wi-Fi connections for video streaming and control, the project was inspired by this. Using just the packet size and inter-arrival time of encrypted Wi-Fi data, the proposed system can detect and identify UAVs and their operating modes with ease. Our approach uses a re-weighted '1- norm regularisation to speed up online identification by taking into account the amount of samples and computation costs associated with various characteristics. As a result of this approach, feature selection and prediction performance are simultaneously optimised for a single goal. We use the maximum likelihood estimation (MLE) technique to estimate the packet inter-arrival time in order to deal with packet inter-arrival time uncertainty while optimising the trade-off between detection accuracy and latency. Using real-world Wi-Fi data traffic from eight different kinds of consumer UAVs, we gather a significant amount of data for thorough testing of our new approach.

Keywords: Unmanned aerial vehicle (UAV) detection, machine learning, encrypted Wi-Fi traffic classification.

1. INTRODUCTION

The consumer unmanned aerial vehicle (UAV) industry has grown rapidly in recent years for personal leisure. Consumer UAVs have the potential to significantly boost economic development, but their widespread use poses a number of concerns about airspace management, public safety, and individual privacy [1]. In September of last year, a drone operating illegally over a residential area accidentally hit an Army helicopter [2]. Through Massachusetts, a UAV was seen peering in a teen's window in April of 2016 [3]. A tiny unmanned aerial vehicle (UAV) crashed on the White House grounds in January 2015, raising concerns about security [4]. UAV registration procedures established by the Federal Aviation Administration (FAA) have been pushed globally to cope with these concerns and assist law enforcement authorities manage UAVs and their owner information [5]. Protect critical locations from hostile UAV invasion by establishing UAV-restricted zones and geo-fencing in places like airports, nuclear plants, and data centres. Enforcing rules is difficult in reality, though. Many unmanned aerial vehicles (UAVs) remain unregistered, and geo-fencing is often absent or readily disabled on many UAVs. An intruder UAV in a restricted region must be promptly detected or the forensics investigation must be assisted to identify its appearance and operating mode. An ideal detection method would warn us as soon as an unwelcome UAV enters the restricted region. UAV countermeasures may then be implemented, and the UAV's owner can be traced or found if they are detected. Detecting consumer UAVs effectively is thus critical.

Identifying the operating mode of a UAV will be very beneficial for forensics in addition to detecting UAVs. If investigators can determine how invader UAVs operate, that information may be utilised in court as evidence in a legal proceeding and by law enforcement authorities to better prepare for future UAV occurrences. The detection of unmanned aerial vehicles has been suggested using a variety of physical methods, including radar [6], [7], acoustic [8, 9] and vision. It is possible that in certain real situations, such as a busy metropolitan area, these detection techniques will be less successful when utilising just one of these sensors. In a civilian setting, walls, buildings, and other barriers may interfere with radar signals. In non-line-of-sight situations and in the dark, the vision detection method fails to identify the UAV. Environmental sounds may overpower the relatively modest sound generated by small rotor-craft or gliding fixed-wing UAVs, interfering with acoustic detection.

The purpose of this research is to investigate machine learning-based Wi-Fi traffic identification methods to quickly detect and identify the operating mode of unmanned aerial vehicles (UAVs). Many current consumer UAVs have Wi-Fi interfaces and connect with a user portable device (e.g., smartphone) for command control or video streaming, which prompted the idea.

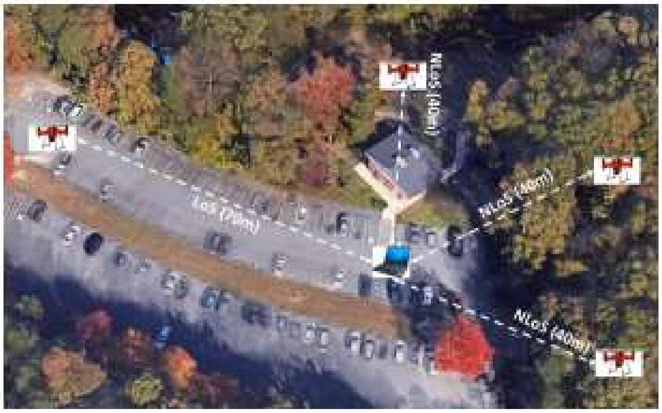

Fig. 1: UAV detection test scenarios

There are many benefits of using wireless traffic identification to detect unmanned aerial vehicles (UAVs). First, barriers, other flying objects, acoustic noise, and light conditions that might influence physical detecting methods have less of an impact on Wi-Fi signal sensing and packet capture. Another benefit of using Wi-Fi data traffic for forensics investigations is that it gives information on the kind of UAV and its operating mode.

2. CHALLENGES

When using Wi-Fi traffic identification to identify an unmanned aerial vehicle (UAV), there are a few specific difficulties that set it apart from other methods of traffic identification and sensing. These include the following:

1) The UAV communication may be encrypted to protect it from prying eyes. Network monitoring and intrusion detection systems that rely on packet header inspection or port

filtering will not function with encrypted UAV communication because of this. Wireless communication is secured by Wi-Fi operated UAVs (such as DJI and Bebop drones) using WPA2. Although the SSID in the MAC frame may disclose information about the drone's kind or manufacturer, drone control applications make it simple to alter the ID.

2) UAV traffic cannot be detected quickly using current machine learning techniques. A drone must be recognised as soon as it enters or approaches a restricted region for real-time applications. Machine learning techniques that simply seek to minimise detection error cannot be directly used from a learning and classification viewpoint [13], [14]. As well as future packet arrival time, detection delay caused by calculations on feature creation should be taken into account.
3) UAV traffic cannot be detected using traditional time series early detection methods (e.g., [16]). Because UAV data has a variable arrival time between packets, the conventional early detection technique that relies on regular time intervals cannot be used directly in this case.

Using machine learning-based UAV detection with delay awareness, we suggest an adjustable balance between UAV detection accuracy and latency in order to solve these issues. Using just packet size and inter-arrival time, we classify the encrypted data stream as a time series and extract statistical characteristics. Our approach uses a re-weighted '1-norm regularisation and combines feature selection and performance optimization into a single objective function by considering the computation time for various features. We utilise the maximum likelihood estimation (MLE) technique to estimate the packet inter-arrival time to deal with the packet inter-arrival time uncertainty while estimating the delay cost function. Misclassification/misdetection, as well as delay costs, are updated in real time as a new packet arrives, so that the anticipated total cost function may be minimised.

Our main contributions are summarized as follows:

- In order to recognise and identify the operating mode of an unmanned aerial vehicle (UAV) via encrypted Wi-Fi data, we present a machine learning-based framework. This framework solely uses information about packet size and inter-arrival time to generate its characteristics. As long as the packet size and interval can be monitored, this framework may be extended to other kinds of encrypted communication, such as cellular traffic or proprietary protocol traffic.
- Feature selection and accuracy optimization are integrated into one objective function, which incorporates the significance of the feature and the difference in computation time between various features, in order to decrease model prediction time for rapid UAV detection.
- To estimate the packet inter-arrival time, we suggest using a model-based MLE approach. The accuracy of the estimate is then assessed using a well-known measure called the mean square error (MSE).
- As well as finding various kinds of UAVs, our approach also determines the mode of operation, such as hovering, flying or being on standby.
- Using real-world encrypted WiFi data traffic from non-UAVs and eight different kinds of consumer UAVs, we gather a significant quantity of data and thoroughly test the effectiveness of the suggested approaches.

Through comprehensive study, we obtain the following findings:

- The UAV traffic differs from non-UAV traffic in terms of patterns. The use of machine learning techniques helps distinguish UAV traffic from a broad variety of other types of data.
- Distinct kinds of UAVs have different traffic patterns, which may be used to categorise UAVs from different vendors due to vendor specific implementation of UAV command control and video streaming protocols.
- Depending on the UAV operating mode, various patterns of Wi-Fi traffic appear. This result suggests that cyber information (data flow) and physical information (UAV operating mode) have a high connection or linkage. As a result of this discovery, new cyber-physical protection and forensics methods will be developed that take advantage of the connection between cyber and physical systems. For many Internet of Things (IoT) applications, such as linked vehicles, smart homes, smart healthcare, and industrial control systems, we think this approach can be extended to other cyber-physical systems (CPS).

3. UAV DETECTION MECHANISMS

In this section, we briefly describe the existing UAV detection methods.

1) UAV Detection Through Physical Sensing:

Existing UAV detection systems mostly rely on physical sensors, such as radar, vision, and sound, to identify threats. Since World War II, radar systems have been a well-known and effective method of finding aircraft in the sky. X-band radar devices were suggested to accommodate the detection of tiny unmanned aerial vehicles (UAVs). Because of the need for a clear line of sight in urban settings (such a metropolis), radar-based detection may be less effective. While video cameras may be used to identify unmanned aerial vehicles (UAVs), they have the same drawback as radar-based systems in that they need a direct line of sight to work. A reliable and promising UAV detection system would be built using several radars and cameras fused to cover the desired region if cost is not an issue.

The out-of-sight issue may be solved using an acoustic signal-based UAV detection technique. However, there are disadvantages to using this approach. Due to noise produced by electric-powered rotorcraft with fixed wings, the acoustic signal from an unmanned aerial vehicle (UAV) may be very loud. Second, comparable acoustic signal generators like electric lawn whackers may provide sound signals that are quite similar to UAVs. Hybrid methods combining an acoustic sensor and a video camera have been suggested to address the shortcomings of separate approaches. The radar sensor may also be used as part of a hybrid system.

2) RF-fingerprinting Based UAV Detection:

In order to detect and identify the UAV type, a novel RF signal fingerprinting technique has been developed. Using wireless data gathered from a variety of UAVs, they propose to create Auxiliary Classifier Wasserstein Generative Adversarial Networks (AC-WGANs). According to their findings, this technique has a 95% success rate in detecting unmanned aerial vehicles (UAVs) inside and an 80% success rate outside. In another recent study, Bisio et al. presented a technique for detecting amateur unmanned aerial vehicles (UAVs) based on Wi-Fi statistical fingerprints and current multiclass machine learning methods. Detection latency isn't an issue in this study, therefore the focus is on developing a machine learning model that can identify an invading UAV using a predetermined and fixed set of statistical characteristics that are calculated at regular intervals during the experiment. For detection, we suggest an adjustable balance between detection accuracy and latency as well as calculation time for the feature.

4. DATA TRAFFIC CLASSIFICATION/IDENTIFICATION

The nature of non-encrypted network data traffic may be determined using traditional methods such as port-based, payload-based, and deep packet inspection. Many applications' data traffic is encrypted these days for their own security, therefore our job is inextricably linked to the categorization and identification of encrypted data flow. Multiple studies have utilised protocol data fingerprinting to identify encrypted data flows in wired and wireless networks. These methods often use statistics as well as leanings from machines to identify the encrypted data flow.

However, our work is different from the existing traffic identification works in the following aspects:

1) As packets reach the detection system, our model does packet-by-packet analysis, allowing us to make decisions quickly.
2) While taking time cost into account, our approach adaptively determines the optimum packet count required for optimal identification with high accuracy (or delay).
3) Feature generation time is also taken into account while training the model, and important features are chosen. As a result, no unnecessary features are produced during prediction/detection, and as a result, the detection time is shortened.

In this article, the detected UAV's operating mode is determined, which differs from our previous work. These modes are identified using multiclass classification machine learning algorithms on a significant quantity of real-world operations data traffic generated by four unmanned aerial vehicles (UAVs). We also increase the number of frequently used consumer UAVs in our delay-aware detection test from four to eight, and evaluate the suggested techniques in great detail. In addition, we use MLE to evaluate the performance of packet inter-arrival time estimate. The findings show that when a high number of packets' information is provided in the detection system, the MSE of estimate is decreased. Also, keep in mind that our detection method is applicable to UAVs operated by a user's portable device that are sneaking up on us (e.g., smartphone). So, a communication connection should be established between the UAV and controller so that traffic may be monitored and the UAV detected using the suggested approach. On the other side, our approach will fail if the invading UAV is equipped with sophisticated autonomous technologies like Autonomous Guidance, Navigation and Control (GN&C), which eliminates the need for a ground control station to provide control.

5. DATA COLLECTION AND PREPARATION

A. UAV Detection Dataset

DBPower UDI U842 Predator FPV (UDI), DBPOWER Discovery FPV (Discovery), DJI Tello (Tello), Tenergy TDR Phoenix Mini RC Quadcopter Drone (TDR), and Wingsland Mini Racing Drone (Wingsland) are among the consumer UAVs from which we gather traffic flows (Wingsland). The Wi-Fi network traffic is monitored and collected using a DELL Latitude laptop equipped with an Intel Corporation Wireless 8260 NIC in promiscuous mode. We gather the UAV traffic from each kind of UAV as it flies and provide footage to the controller. To accomplish this, we used Wireshark version 2.4.11 to collect Wi-Fi traffic data while setting the monitoring sensor's channel frequency to the same channel as the UAV's operational channel. There are 3,000 traffic traces in each UAV dataset, with n = 200 packets in each trace. We clean the data and prepare it for training and testing after gathering it and detecting the UAV traffic patterns. We delete all broadcast packets (e.g., 802.11 beacon frames), broken packets, and packets with just a receiving address during the data cleaning phase of the transmission. (e.g., 802.11 ACK frames). Only a few packets remain. These include things like video, UAV status updates like velocity and altitude (together with GPS data), and UAV control instructions (including UAV reaction to control orders).

B. Non-UAV Dataset

So that our non-UAV dataset would be varied, we created a dataset divided into the following two major sub-datasets: To begin, we use the CRAWDAD database's Wi-Fi data flow. The following are the reasons why we selected this particular collection of data.

1) To create this dataset, we used popular apps like Google Hangouts, ooVoo, and Skype to collect video streaming traffic. We then used that traffic to create the TED and YouTube videos you see today.
2) A smartphone app tracks the user's various movement habits and collects the traffic statistics. The second thing that we've done is collect encrypted Wi-Fi data from a university campus Wi-Fi network, which typically has a mix of different kinds of traffic including video streaming and social network apps as well as VoIP and email. Assuming that the campus has a UAV identification system in place, our system should be able to distinguish UAV traffic from non-UAV traffic using the information we collect. For the non-UAV dataset, there are additional 3,000 traffic trace records with n = 200 consecutive packets (e.g. Google Hangouts, OoVoo, Skype, TED, and Campus traffic).

C. UAV Operation Mode Dataset

The following steps are taken for an operation mode data traffic collection of a specific UAV type:

1) Wi-Fi connection is established between the UAV and controller.

2) A specific operation mode command (e.g., “Forward”) is given via controller to the UAV and is held.

3) Wi-Fi medium monitoring sensor is activated to monitor the wireless channel traffic.

4) Wireshark is run on the promiscuous mode to capture the packets.

5) Before releasing the command in the controller, first, Wireshark is stopped, and then the collected traffic is saved and labeled according to the commanded operation.

6. CONCLUSION

In order to enforce regulations, conduct forensic investigations, ensure public safety, and safeguard individuals' personal privacy it is essential to find and identify consumer UAVs. We developed a machine learning-based UAV detection and operating mode identification framework using delay-aware machine learning over encrypted WiFi UAV data to augment current physical detection methods. During the model training phase, a re-weighted '1-norm regularisation with account of computation time among different variables is used to extract features from packet size and inter-arrival time. As a result, the selection of features and the optimization of performance are combined into a single goal. We used the model-based MLE approach to estimate the incoming flow's packet inter-arrival times in order to cope with packet inter-arrival time uncertainty while estimating the cost function. The quantity of encrypted Wi-Fi communication generated by eight different kinds of consumer UAVs was gathered and thoroughly analysed to see how effective our suggested techniques were. Experiments indicate that using the suggested techniques, UAVs may be detected and identified with an accuracy of 85.7% to 95.2% within a timeframe of 0.150.35s. Line-of-sight (LoS) and non-line-of-sight (NLoS) situations have UAV detection ranges of 70m and 40m, respectively. UAV modes of operation may also be recognised with an accuracy of 88.598.2 percent. Uncovering UAVs' cyber-physical connection feature is revealed by identifying their operating mode. Given information on the cyber portion of UAVs and this connection, we may deduce information on their physical state (operating mode) (Wi-Fi traffic data). The suggested machine learning based detection framework and technique are generic enough to be extended to various cyber-physical/IoT systems utilising other wireless communication protocols, even though this study makes use of Wi-Fi data to identify consumer UAVs (e.g., Bluetooth and cellular). For many additional CPS/IoT systems, we believe our research will provide light on cyber physical assault co-detection or co-defense.

REFERENCES

[1] R. Altawy and A. M. Youssef, "Security, privacy, and safety aspects of civilian drones: A survey," ACM Trans. Cyber-Phys. Syst., vol. 1, no. 2, pp. 7:1–7:25, Nov. 2016.

[2] D. Furfaro, L. Celona, and N. Musumeci, "Civilian drone crashes into army helicopter," RT, 2017. [Online]. Available: http://nypost.com/ 2017/09/22/army-helicopter-hit-by-drone/

[3] U Butkar, M Waghmare. An Intelligent System Design for Emotion Recognition and Rectification Using Machine Learning. Computer Integrated Manufacturing Systems, 29(2), 32–42. Retrieved from http://cims-journal.com/index.php/CN/article/view/783 .

[4] M. Shear and M. Schmidt, "White House drone crash described as a US workers drunken lark," New York Times, Jan. 2015. [Online]. Available: https://www.nytimes.com/2015/01/28/us/white-house-drone.html

[5] F. A. Administration, "UAS registration," FAA website: https://www.faa.gov/uas/getting started/registration/, FAA, 2015.

[6] Umakant Butkar, " Review On- Efficient Data Transfer for Mobile devices By Using Ad-Hoc Network", International Journal of Engineering and Computer Science, vol 5, Issue 3, 2016

[7] D. H. Shin, D. H. Jung, D. C. Kim, J. W. Ham, and S. O. Park, "A distributed fmcw radar system based on fiber-optic links for small drone detection," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 2, pp. 340–347, Feb 2017.

[8] A. M. Zelnio, E. E. Case, and B. D. Rigling, "A low-cost acoustic array for detecting and tracking small rc aircraft," in Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th, Jan 2009, pp. 121–125.

[9] P. Marmaroli, X. Falourd, and H. Lissek, "A UAV motor denoising technique to improve localization of surrounding noisy aircrafts: proof of concept for anti-collision systems," in Acoustics, April 2012, pp. 23–27.

[10] Butkar Uamakant, “A Formation of Cloud Data Sharing With Integrity and User Revocation”, International Journal Of Engineering And Computer Science, Vol 6, Issue 5, 2017

[11] P. A. Prates, R. Mendona, A. Loureno, F. Marques, J. P. Matos-Carvalho, and J. Barata, “Vision-based UAV detection and tracking using motion signatures,” in Proc. IEEE Industrial Cyber-Physical Systems (ICPS), May 2018, pp. 482–487.

[12] Umakant Butkar, “A Fuzzy Filtering Rule Based Median Filter For Artifacts Reduction of Compressed Images”,IJIFR, Vol 1, Issue 11, 2014

[13] N. Jing, M. Yang, S. Cheng, Q. Dong, and H. Xiong, “An efficient SVMbased method for multi-class network traffic classification,” in Proc. 30th IEEE International Performance Computing and Communications Conference, Nov 2011, pp. 1–8.

[14] A. McGregor, M. Hall, P. Lorier, and J. Brunskill, “Flow clustering using machine learning techniques,” in Passive and Active Network Measurement, C. Barakat and I. Pratt, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 205–214.

[15] R. Bar Yanai, M. Langberg, D. Peleg, and L. Roditty, “Realtime classification for encrypted traffic,” in Experimental Algorithms. Springer, 2010, pp. 373–385.