International Journal of Applied Engineering & Technology

A STUDY ON MACHINE LEARNING-BASED WI-FI TRAFFIC FOR UAV DETECTION AND TRACKING

Akshada Pandurang Kulkarni* and Dr. Atul Dattatrya Newase
Department of Computer Application, Dr. A. P. J. Abdul Kalam University, Indore (M.P.), India 452016
akshada.ma@gmail.com and dr.atulnewase@gmail.com

ABSTRACT

In recent years, consumer unmanned aerial vehicles (UAVs) have experienced significant growth. Despite their substantial economic potential, consumer UAVs raise serious security and privacy concerns due to their diverse applications. Quick detection and identification of intruding UAVs are crucial for both invasion detection and forensics to mitigate these risks. We propose a machine learning-based framework for rapid UAV identification via encrypted Wi-Fi data, complementing existing physical detection methods. This project was inspired by the common use of Wi-Fi connections for video streaming and control by many consumer UAVs. Our system can easily detect and identify UAVs and their operating modes using only the packet size and inter-arrival time of encrypted Wi-Fi data. By employing a re-weighted l1-norm regularization, we enhance online identification speed by balancing the number of samples and computation costs of various features. This approach simultaneously optimizes feature selection and prediction performance. To address packet inter-arrival time uncertainty and optimize the trade-off between detection accuracy and latency, we utilize the maximum likelihood estimation (MLE) technique. We collected extensive real-world Wi-Fi data traffic from eight different types of consumer UAVs to rigorously test 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, particularly for personal leisure activities. While consumer UAVs hold significant potential for economic development, their widespread use raises numerous concerns related to airspace management, public safety, and individual privacy. For example, in September of last year, an illegally operated drone collided with an Army helicopter over a residential area. In another instance, a UAV was spotted peering into a teenager's window in Massachusetts in April 2016. Additionally, a small UAV crash-landed on the White House grounds in January 2015, highlighting severe security concerns. To address these issues, the Federal Aviation Administration (FAA) has implemented UAV registration procedures globally, assisting law enforcement authorities in managing UAVs and their owner information.

Protecting critical locations from hostile UAV invasions involves establishing UAV-restricted zones and implementing geo-fencing around sensitive areas such as airports, nuclear plants, and data centers. However, enforcing these rules presents significant challenges. Many UAVs remain unregistered, and geo-fencing can often be disabled or is entirely absent on numerous UAV models. Therefore, it is crucial to promptly detect intruder UAVs in restricted regions or aid forensic investigations to determine their presence and operating mode. An effective detection method should alert authorities as soon as an unauthorized UAV enters a restricted area, allowing for the implementation of countermeasures and tracing the UAV's owner.

Detecting consumer UAVs effectively is of paramount importance. In addition to detecting UAVs, identifying their operating modes can greatly benefit forensic investigations. Understanding how intruding UAVs operate can provide critical evidence in legal proceedings and help law enforcement authorities better prepare for future UAV-related incidents. Various physical detection methods, including radar, acoustic, and vision-based techniques, have been proposed for UAV detection. However, these methods may be less effective in certain real-world scenarios, such as busy urban areas. Radar signals can be obstructed by walls, buildings, and other barriers, while vision-based detection methods fail in non-line-of-sight situations and in the dark. Acoustic detection can be compromised by environmental noise that overpowers the sound generated by small rotor-craft or gliding fixed-wing UAVs.

This research investigates machine learning-based Wi-Fi traffic identification methods to rapidly detect and identify the operating modes of UAVs. The inspiration for this approach comes from the fact that many current consumer UAVs use Wi-Fi interfaces to connect with user devices, such as smartphones, for command control or video streaming. Utilizing wireless traffic identification to detect UAVs offers several advantages. First, Wi-Fi signal sensing and packet capture are less affected by barriers, other flying objects, acoustic noise, and light conditions that might impair physical detection methods. Additionally, using Wi-Fi data traffic for forensic investigations provides valuable information about the type of UAV and its operating mode, enhancing the ability to respond effectively to UAV-related incidents.

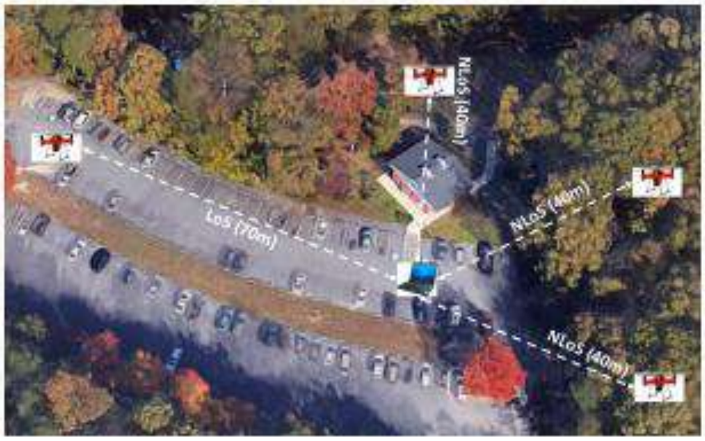

Fig. 1: UAV detection test scenarios

2. CHALLENGES

When using Wi-Fi traffic identification to detect unmanned aerial vehicles (UAVs), several specific challenges distinguish this method from other traffic identification and sensing techniques:

1. **Encryption of UAV Communication**: UAV communication is often encrypted to safeguard it from unauthorized access. Network monitoring and intrusion detection systems that rely on packet header inspection or port filtering are ineffective against encrypted UAV communication. Many Wi-Fi-operated UAVs, such as DJI and Bebop drones, use WPA2 encryption for wireless communication. Although the SSID in the MAC frame may reveal some information about the drone's type or manufacturer, drone control applications make it easy to alter this ID, further complicating identification.

2. **Real-Time Detection Requirements**: Current machine learning techniques are not optimized for the quick detection of UAV traffic. For real-time applications, it is critical to recognize a drone as soon as it enters or approaches a restricted area. Traditional machine learning techniques that aim to minimize detection error are not directly applicable from a learning and classification standpoint. These techniques must consider the detection delay caused by calculations on feature creation, as well as the future packet arrival time.

3. **Irregular Packet Arrival Times**: Traditional time series early detection methods cannot be applied to UAV traffic because UAV data packets arrive at variable intervals. Conventional early detection techniques rely on regular time intervals between packets, making them unsuitable for detecting UAVs with irregular packet arrival times.

To address these challenges, we propose a machine learning-based UAV detection approach that balances detection accuracy and latency with delay awareness. Our method classifies the encrypted data stream as a time series using only packet size and inter-arrival time, extracting statistical characteristics for analysis. We employ a re-weighted l1-norm regularization, which integrates feature selection and performance optimization into a single objective function by considering the computation time for various features. To manage packet inter-arrival time uncertainty, we use the maximum likelihood estimation (MLE) technique to estimate the packet inter-arrival time and evaluate the delay cost function.

Our approach updates misclassification/misdetection and delay costs in real-time as new packets arrive, ensuring that the anticipated total cost function is minimized. This adaptive balance between detection accuracy and latency enhances the effectiveness of UAV detection in real-world scenarios, providing a robust solution for identifying and responding to unauthorized UAV activities.

3. MAIN CONTRIBUTIONS

1. **Machine Learning-Based Framework for UAV Detection and Identification**: We present a novel machine learning-based framework designed to recognize and identify the operating modes of unmanned aerial vehicles (UAVs) using encrypted Wi-Fi data. This framework relies exclusively on packet size and inter-arrival time for feature generation. It is adaptable to other forms of encrypted communication, including cellular traffic or proprietary protocol traffic, as long as packet size and intervals can be monitored.
2. **Integrated Feature Selection and Accuracy Optimization**: Our framework integrates feature selection and accuracy optimization into a single objective function. This approach considers the significance of each feature and the computational time required for different features, thereby reducing model prediction time for rapid UAV detection.
3. **Model-Based MLE Approach for Packet Inter-Arrival Time Estimation**: We propose using a model-based Maximum Likelihood Estimation (MLE) approach to estimate packet inter-arrival times. The accuracy of these estimates is assessed using the mean square error (MSE) metric.
4. **Operating Mode Detection**: Beyond identifying various types of UAVs, our approach also determines their operating modes, such as hovering, flying, or standby.
5. **Extensive Real-World Data Testing**: We collected substantial real-world encrypted Wi-Fi data traffic from both non-UAV sources and eight different types of consumer UAVs. Our framework's effectiveness was rigorously tested against this extensive dataset.

Comprehensive Study Findings

1. **Distinct Traffic Patterns**: UAV traffic exhibits distinct patterns compared to non-UAV traffic. The application of machine learning techniques effectively distinguishes UAV traffic from various other types of data.
2. **Vendor-Specific Traffic Patterns**: Different types of UAVs display unique traffic patterns due to vendor-specific implementations of UAV command control and video streaming protocols. These patterns can be utilized to categorize UAVs from different manufacturers.
3. **Correlation Between Wi-Fi Traffic and UAV Operating Modes**: Different UAV operating modes generate distinct Wi-Fi traffic patterns. This finding suggests a strong correlation between cyber information (data flow) and physical information (UAV operating mode). Leveraging this connection, new cyber-physical protection and forensic methods can be developed.
4. **Extensibility to Other CPS and IoT Applications**: We believe this approach can be extended to other cyber-physical systems (CPS) and Internet of Things (IoT) applications, such as connected vehicles, smart homes, smart healthcare, and industrial control systems, benefiting from the established link between cyber and physical systems.

UAV Detection Mechanisms

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

1. **UAV Detection Through Physical Sensing**: Existing UAV detection systems predominantly rely on physical sensors, such as radar, vision, and sound, to identify threats. Since World War II, radar systems have been a well-established and effective method for detecting aircraft in the sky. To accommodate the detection of small UAVs, X-band radar devices have been suggested. However, radar-based detection may be less effective in urban settings due to the need for a clear line of sight. While video cameras can also identify UAVs, they share the same limitation as radar systems, requiring a direct line of sight to operate effectively. A comprehensive UAV detection system could be built using multiple radars and cameras fused together to cover the desired area, provided cost is not a constraint.

Acoustic signal-based UAV detection techniques can address the out-of-sight issue. However, this approach has its drawbacks. The noise produced by electric-powered rotorcraft with fixed wings can be very loud, and other sources like electric lawn whackers can generate similar acoustic signals. Hybrid methods, combining an acoustic sensor and a video camera, have been suggested to overcome the limitations of individual approaches. Radar sensors can also be part of a hybrid system.

2. **RF-Fingerprinting Based UAV Detection**: A novel RF signal fingerprinting technique has been developed to detect and identify UAV types. This technique proposes using Auxiliary Classifier Wasserstein Generative Adversarial Networks (AC-WGANs) to analyze wireless data gathered from various UAVs. The findings show a 95% success rate in detecting UAVs indoors and an 80% success rate outdoors. Another study by Bisio et al. presented a technique for detecting amateur UAVs based on Wi-Fi statistical fingerprints and current multiclass machine learning methods. In this study, detection latency was not a primary concern; instead, the focus was on developing a machine learning model that can identify an invading UAV using a predetermined and fixed set of statistical characteristics calculated at regular intervals. Our approach, however, suggests an adjustable balance between detection accuracy, latency, and calculation time for the feature.

4. DATA TRAFFIC CLASSIFICATION/IDENTIFICATION

Traditional methods for determining the nature of non-encrypted network data traffic include port-based, payload-based, and deep packet inspection. However, as many applications now encrypt their data traffic for security, our task involves the classification and identification of encrypted data flows. Multiple studies have utilized protocol data fingerprinting to identify encrypted data flows in wired and wireless networks, often using statistical analysis and machine learning techniques.

Our work differs from existing traffic identification studies in several key aspects:

1. **Packet-by-Packet Analysis**: Our model performs packet-by-packet analysis as packets reach the detection system, allowing for quick decision-making.
2. **Adaptive Packet Count Determination**: Our approach adaptively determines the optimal packet count required for high-accuracy identification while considering time costs.
3. **Feature Generation Efficiency**: During model training, we consider feature generation time and select important features, ensuring no unnecessary features are produced during prediction/detection, thereby shortening detection time.

In this article, we determine the detected UAV's operating mode, a departure from our previous work. These modes are identified using multiclass classification machine learning algorithms on a significant quantity of real-world data traffic generated by four UAVs. We expand our delay-aware detection test to include eight commonly used consumer UAVs and evaluate our proposed techniques in detail. Additionally, we use Maximum Likelihood Estimation (MLE) to assess the performance of packet inter-arrival time estimation. The findings show that providing a high volume of packet information to the detection system reduces the Mean Square Error (MSE) of the estimate.

Our detection method applies to UAVs operated by a user's portable device, such as a smartphone. Thus, a communication connection should be established between the UAV and the controller for traffic monitoring and UAV detection using our proposed approach. However, our method will fail if the invading UAV is equipped with sophisticated autonomous technologies like Autonomous Guidance, Navigation, and Control (GN&C), which eliminate the need for a ground control station to provide control.

5. DATA COLLECTION AND PREPARATION

A. UAV Detection Dataset

We collected traffic flows from a variety of consumer UAVs, including the 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). The Wi-Fi network traffic was monitored and collected using a DELL Latitude laptop equipped with an Intel Corporation Wireless 8260 NIC in promiscuous mode. We captured the UAV traffic as each UAV flew and transmitted footage to the controller.

To collect Wi-Fi traffic data, we used Wireshark version 2.4.11, setting the monitoring sensor's channel frequency to match the UAV's operational channel. Each UAV dataset contains 3,000 traffic traces, with n = 200 packets per trace. After gathering the data and identifying the UAV traffic patterns, we cleaned and prepared it for training and testing. During data cleaning, we removed all broadcast packets (e.g., 802.11 beacon frames), broken packets, and packets with only a receiving address (e.g., 802.11 ACK frames). The remaining packets included video, UAV status updates like velocity and altitude (with GPS data), and UAV control instructions (including UAV responses to control orders).

B. Non-UAV Dataset

To ensure diversity, our non-UAV dataset was divided into two main sub-datasets:

1. **CRAWDAD Database Wi-Fi Data Flow**: This dataset was selected for several reasons. First, we collected video streaming traffic from popular apps like Google Hangouts, ooVoo, and Skype, which was then used to create TED and YouTube videos. Second, a smartphone app tracked the user's various movement habits and collected the traffic statistics.
2. **University Campus Wi-Fi Data**: We collected encrypted Wi-Fi data from a university campus network, which typically includes a mix of different types of traffic such as video streaming, social network apps, VoIP, and email. This dataset is designed to simulate an environment where a UAV identification system could distinguish UAV traffic from non-UAV traffic. The non-UAV dataset also contains 3,000 traffic trace records, with n = 200 consecutive packets each (e.g., Google Hangouts, ooVoo, Skype, TED, and campus traffic).

C. UAV Operation Mode Dataset

To collect data for specific UAV operation modes, we followed these steps:

1. Establish a Wi-Fi connection between the UAV and the controller.
2. Issue a specific operation mode command (e.g., "Forward") via the controller and maintain it.
3. Activate the Wi-Fi medium monitoring sensor to monitor the wireless channel traffic.
4. Use Wireshark in promiscuous mode to capture the packets.
5. Before releasing the command on the controller, stop Wireshark and save the collected traffic, labeling it according to the commanded operation.

By meticulously collecting and preparing these datasets, we ensured that our model could effectively distinguish between UAV and non-UAV traffic, as well as identify different UAV operation modes.

6. CONCLUSION

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 Wi-Fi UAV data to augment current physical detection methods. During the model training phase, we used re-weighted l1-norm regularization, which accounts for computation time among different variables, to extract features from packet size and inter-arrival time. This approach combines feature selection and performance optimization into a single goal.

We utilized the model-based Maximum Likelihood Estimation (MLE) approach to estimate packet inter-arrival times, addressing the uncertainty of packet inter-arrival time while estimating the cost function. We collected and thoroughly analyzed the encrypted Wi-Fi communication data generated by eight different kinds of consumer UAVs to evaluate the effectiveness of our proposed techniques. Experiments indicate that using our suggested techniques, UAVs can be detected and identified with an accuracy of 85.7% to 95.2% within a timeframe of 0.15 to 0.35 seconds. In line-of-sight (LoS) situations, UAV detection ranges are up to 70 meters, while in non-line-of-sight (NLoS) situations, detection ranges are up to 40 meters. Additionally, UAV operating modes can be recognized with an accuracy of 88.5% to 98.2%.

Identifying UAVs' operating modes reveals the cyber-physical connection feature, allowing us to deduce their physical state (operating mode) from their cyber data (Wi-Fi traffic). Our proposed machine learning-based detection framework and techniques are general enough to be extended to various cyber-physical and IoT systems utilizing other wireless communication protocols, such as Bluetooth and cellular. We believe our research will shed light on co-detection or co-defense mechanisms for cyber-physical attacks across many additional CPS/IoT systems.

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. [Online]. Available: http://doi.acm.org/10.1145/3001836

[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] RT, “Peeping drone: UAV hovers outside of massachusetts teen’s bedroom window,” RT, Apr. 2016. [Online]. Available: https: //www.rt.com/usa/341404-drone-privacy-teenager-window/

[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] A. Moses, M. J. Rutherford, and K. P. Valavanis, “Radar-based detection and identification for miniature air vehicles,” in Proc. IEEE International Conference on Control Applications (CCA), Sep. 2011, pp. 933–940.

[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] A. Rozantsev, V. Lepetit, and P. Fua, "Detecting flying objects using a single moving camera," in Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 5, pp. 879–892, May 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] A. Rozantsev, V. Lepetit, and P. Fua, "Flying objects detection from a single moving camera," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015, pp. 4128–4136.

[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.