# DRONE FORENSICS AND DIGITAL FORENSICS APPLIED TO MACHINE LEARNING

COM6016 Cyber Threat Hunting and Digital Forensics - Research Report

DECEMBER 11, 2019
RAMEEZ HUSSAIN

# Table of Contents

# Part A: Drone Forensics

## Drone Forensics and why the area is important

Drone Forensic is a branch of digital forensic which concerns the recovery and examination of drone data. It is also referred to the forensic analysis of drones or Unmanned Air Vehicles (UAVs) even if they have been crashed or damaged [1]. Drone forensic investigators must recover data in a forensically sound manner whether in the drone or in its removable media [2].

The reason why this area is important because drones are used in illegal activities such as spying, drone deployment of weapons, terror attacks and disruption of airport traffic. Drone forensics is helpful in law enforcement and bringing individuals to justice. Forensic departments can extract evidence from drones and monitor suspicious activity and prevent future crime being committed [3].

## Tools that are used for Drone Forensics

Due to the variety of data formats and the number of obtainable evidence, manual extraction is time-consuming. Due to drone forensics being new, very few tools exist for specialists that allow such procedures to take place. Below is a list of some of the drone forensic tools:

**Oxygen Forensic Detective (OFD):** This forensic tool allows experts to extract digital data from the drone internal storage or external SD card, parse and decode data and show it in a readable manner. The drones are controlled via smart devices such as mobile phones through apps like Android or IOS-based smart devices. Some of its data is stored and kept in the user's online account or drone manufacture cloud. OFD can effectively extract information from multiple smart devices providing access to data collected by these apps. OFD can remotely acquire data from multiple cloud services and online. Data that is extracted is placed into a unified dataset. OFD allows the digital forensic investigator to automatically parse GPS locations and flight data and it also decodes data showing drones speed and direction of travel. The investigator can track the meta-data including speed and flight patterns, it also has a built-in mapping tool that allows investigators to see a visual route with points of interest that allows extracting photos and video footage. OFD can extract vital information and data from the cloud, all it needs is the user login credentials. However, if the password is not available the investigators can use authentication token obtained from the computer or cloud [4].

**AcessData technology Forensic Toolkit (FTK):** Digital forensic experts using AcessData technology, data can be extracted from the drone and then using the tool of FTK to perform a detailed investigation. When the drone’s vital data files have been collected, digital forensic investigators can perform their investigation. FTK allows investigators to search for media files that provide detailed information to aircraft operating systems such as internal flash memory such as flight logs, and external SD cards for images and videos taken during flight time [5]. FTK allows investigators to collect and analyse evidence from drones such as video footage, cracking passwords for online accounts and cloud services, recovering information that has been deleted or lost and allows to build forensic reports. Data can be gathered from drone hard drives furthermore, creating a forensic image of digital evidence, even the full drone hard drives allowing to preserve data integrity [6]. FTK is a tool that is made by AcessData, it allows disk imaging. FTK is helpful for imaging a physical drive,

logical drive and partition a USB drive. Using FTK data can be viewed, and deleted files can be recovered based on its grade of being overwritten [7].

## Significant challenges faced when conducting drone forensics and how it is addressed in practice

Drone forensics is a new field and has significant challenges. Some of these challenges are listed below explaining how they are significant and how these challenges are addressed in practice:

**Extraction of Data:** When a drone is captured the first phase that a forensic investigator is to gather data of the drone's holder, flight paths, locations, landing location, photos, and videos allowing the investigator to conduct his investigation. However, gathering data from a drone has been an increasing concern for drone forensic investigators because the investigator must gather data from various components of the drone which include drones, batteries, sensors, remote controllers, ground control stations, cell phones, tablets and the computers used to gather data. Due to the variety of components and technical aspects, it makes the job of extracting data harder. According to *Forensic Focus,* there are no standard practices and methods of storing digital data on drones. Extracting data is a significant challenge because each drone is different, and the extraction process differs from drone to drone. The reason why extraction of data is a significant challenge because its time consuming and complex due to different hardware and operating systems, another reason being some drones allow the extraction of data while being intact, other drones must be dissembled to the chips [8]. Data extraction has proved a lot harder due to encryption in which discretion and data protection can be necessary, data encryption could include encrypting the video link for example. Encryption is the process of hiding information or even converting the information into a secret code that hides the true meaning of the data. In order to decrypt the text, the investigator needs the decipher key to access the data. The reason why this is significant because encryption makes the investigator's job a lot harder as he is unable to read the messages or digital communications making it harder to extract the data for the purpose of evidence [9].

**Solution:** Due to drone forensics being a relatively new field not many solutions are found to address the challenges. However, recently digital investigators have found some promising techniques to tackle this problem. The best method to acquire a forensic image is through The National Institute of Standards and Technology (NIST) as it provides a repository of forensic images from electronical devices. Furthermore, Computer Forensic Reference Datasets (CFReDs) is a repository that makes available documents sets of digital evidence (forensics images) to work within the case. To tackle the problem of encoded data or encrypted data there are specialist tools aiding the digital investigators such as EPRB by ElconSoft and OFD in which passwords could be recovered and decoding of data. This is a helpful tool because it allows the investigators to extract data that is encrypted in a readable manner. Watson conducted an experiment of extracting data from drones using three methods, keeping the drone intact, disassembling the camera and circuit board and disassembling the chips. Data was extracted from pilot controls and other devices remotely connected to the drones [8]. JTAG method is another method to extract data as it is a non-invasive method of physical acquisition that can extract data from electronical devices that one could not do through software. The process contains connecting to the Test Access Ports (TAPs) on the device and transferring the raw data on connected memory chips. It allows direct access to the drones without endangering it [10]. There are tools for extracting data like XRY and OFD which allows preserving data integrity.

**Cloud Data:** Cloud storage solutions represents a significant challenge for drone forensic investigators. Cloud storage includes the distribution of data over a wider range of servers and computing systems. This growing development of all possible locations where data could be preserved and found poses significant challenges and work for digital forensic investigators because all the information is scattered all over the place. There seems to be an increasing number of data which is useful for digital forensic investigators that are never saved on the drone device itself but rather saved on the cloud storage. Cloud backups provide the ability to recover data by the users or the data being locked, broken or wiped devices. Gathering and collecting this data proves a significant challenge to investigators because of legal constraints and due to security mechanisms such as passwords and two-factor authentication method. Not being able to gather data from the cloud could result in losing data and evidence which could be vital for the investigation purpose. Security mechanisms could include passwords, the usage of Local Data Mode and turning off Wi-Fi so data is not shared and making it difficult for investigators to get hands-on data [11]. Normally DJI drones communicate through the cloud storing drone data in the user's online account or the drone manufacture cloud. This represents challenges for specialist both legally and technically because manual extraction is complicated and hard work or even impossible. The use of CFReDs has been used to recover and extract data, authenticate tools and analyse images from the drone via software tools [12].

**Solution:** OFD is useful as it allows data to be extracted from the cloud. For example, DJI accounts all that is required is accessing the data through the user's credentials. However, these credentials are not always available, but the investigators can use an authentication token taken for the user's computer that was used to access the cloud. There are no two-factor authentications that the DJI drone uses or implements making it easy for the drone forensic investigators in accessing the data stored on the cloud [4].

## Part B: Digital forensics and Machine Learning

TALLÓN-BALLESTEROS, A. AND RIQUELME, J. (2014). Data Mining Methods Applied to a Digital Forensics Task for Supervised Machine Learning. *Studies in Computational Intelligence*, pp.413-428.

https://www.researchgate.net/profile/Jose_Riquelme2/publication/289029388_Data_Mining_Methods_Applied_to_a_Digital_Forensics_Task_for_Supervised_Machine_Learning/links/5734606908ae9f741b27b67a/Data-Mining-Methods-Applied-to-a-Digital-Forensics-Task-for-Supervised-Machine-Learning.pdf

### Digital forensics problems addressed in the research paper and why it is important

The digital forensic problem that is being addressed here is that of data mining methods that were applied to a digital forensic task for the purpose of supervised machine learning. This paper takes a wider range of look at the type of glass-based on its chemical analysis. The dataset was named glass identification. The main aim of the researcher was the classification of glass remains in which three data mining techniques were implemented. This paper assesses the digital forensic task of glass identification for multiclass supervised learning. This particular problem has been around for many years and the researcher showed an observational synopsis of the performance with numerous classifiers from different machine learning approaches.

The main reason why this digital forensic problem is important is that multiclass is complicated due to it is very difficult to standardize as this problem has been looked into before however, this was for one issue only. This paper takes a wide range of looks at the performance with many classifiers from a variety of machine learning approaches.

### How machine learning is applied to the problem

Machine learning has been applied to this area as the researcher focused on different supervised machine learning algorithms. The experiment was based on a digital forensic task for multi-class classification with numerous algorithms such as decision trees, Bayes classifier, based on rules, artificial neural networks and based on the nearest neighbors. The classifiers have been tested with two performance measures which are accuracy and Cohen's Kappa. A four-cross validation with thirty repetitions has been implemented for non-deterministic algorithms for the purpose of finding reliable results from 120 runs. A statistical analysis has been implemented by relating algorithms by t-test implementing accuracy and Cohen's Kappa metrics.

The experiment strategy used was a stratified four-cross validation, the motive for this is to split the full dataset into four equivalent parts by testing it with test data and the rest used for a training set. Stochastic algorithms have been tested thirty times across the four folds and the results were averaged by 120 runs to acquire reliable results. For the experiment algorithms which are incorporated in Waikato Environment for Knowledge Analysis (WEKA) without using Classification and Regression Trees (CART) and Radial Basis Function (RBF) as version 3.5.7 was used instead. These methods were tested against supervised machine learning algorithms such as decision trees, Bayes classifiers, based on rules, artificial neural networks and based on the nearest neighbors.

Certain representative methods are applied on these supervised machine learning algorithms for example, in decision trees representative methods of C4.5 and CART was used, Bayes classifier Bayesian Network was used, Rule-based classifier the classifier type was Repeated Incremental Pruning to Produce Error Reduction (RIPPER), artificial neural networks Multi-Layer Perceptron (MLP) approach was used and classifiers based on nearest neighbors 1-NN variants such as Manhattan and Euclidean was implemented. Nine further algorithms were used which are C4.5, CART, BayesNet, back-propagation methods, Manhattan 1-NN and Euclidean. The first experiment conducted the algorithms have been tested with default values and the second experiment used fine-tuned parameters.

A statistical analysis has been carried out to determine the fundamental differences between the results attained by the stochastic algorithms, however the non-stochastic algorithms it was not feasible to investigate the analysis due to only having one result per fold and degrees would be too low for it.

The results for the first experiment performed which used the default parameter values that were set by their own respective authors in relation to the deterministic algorithms, Manhattan 1-NN had the greatest performance for accuracy and Cohen's Kappa metrics with an accuracy over 70%. The performance of this was better than Euclidean 1-NN with the difference of 0.5 for accuracy and 0.01 for Cohen's Kappa metrics, this has proven that the nearest neighbors classifiers can be applied successfully. The stochastic algorithms the best performers are CART and RIPPER, it was not feasible to draw a direct comparison between the non-stochastic and stochastic algorithms due to the different number of iterations for different methods. The statistical test results for accuracy we can assume that CART has the best performance and in second place is RIPPER, however, the differences are with their competitors so therefore, the best classifier is the decision tree method. Examining the neural networks there is no big difference even though MLP is better than RBF.

The second experiment was the fine-tuned parameter values that were established for the fine set by the way of a grid search with the training set of each fold. The best deterministic algorithm is the C4.5 or 1-NN  Manhattan in terms of their performance valuation measures. The best non-deterministic algorithm is CART with both measures taken into consideration. The fine-tuning of the parameter of the non-stochastic algorithm has improved the results and has exceeded the top as 73.5% of accuracy and reached a Cohen's Kappa close to 0.595.

## How successful was the application of machine learning to the problem?

The application of machine learning has been applied successfully particularly for the fine-tuning parameters experiment as the non-stochastic algorithm has an accuracy of 73.5% with a C4.5 classifier and relatively close to 0.595 for Cohen's Kappa metric with 1-NN Manhattan. The performance of the stochastic algorithm got 68.22 and 0.5576 for accuracy and Cohen's Kappa metrics. These results are promising and prove that the application of machine learning has been applied successfully. Furthermore, it was proven that the nearest neighbors classifier can be applied successfully due to the performance in the first experiment where deterministic algorithms Manhattan 1-NN had the best performance and proves its better than Euclidean 1-NN.

## Possible further research to be conducted

Possible further research could be conducted in considering pre-processing data mining techniques so that it could act on features, instances or values of the attribute. The way that this could be applied is by applying standardisation of data meaning data received will be converted into a common format. This also could be implemented by splitting the data using K nearest neighbors which a non-parametric method is used for classification and regression. This can be used for shaping the accuracy of the training set which gives better results [13].

# References


[1] QCC Global. (2019). *Drone Forensics - QCC Global*. [online] Available at: https://www.qccglobal.com/drone-forensics/ [Accessed 10 Nov. 2019].

[2] Fruhlinger, J. (2019). *What is digital forensics? And how to land a job in this hot field*. [online] CSO Online. Available at: https://www.csoonline.com/article/3334396/what-is-digital-forensics-and-how-to-land-a-job-in-this-hot-field.html [Accessed 10 Nov. 2019].

[3] Data Forensics Simplified — Software Tools for Digital Forensic Analysis. (2019). *Drone Forensics - Detailed Analysis Done & Explained*. [online] Available at: https://www.dataforensics.org/drone-forensics/ [Accessed 10 Nov. 2019].

[4] Courcier, S. (2019). *Oxygen Drone Forensics – How To Deal With A New Threat*. [online] Forensic Focus - Articles. Available at: https://articles.forensicfocus.com/2018/03/06/oxygen-drone-forensics/ [Accessed 10 Nov. 2019].

[5] AccessData. (2019). *The Emerging World of Drone Forensics: Extracting Data from an…*. [online] Available at: https://accessdata.com/blog/the-emerging-world-of-drone-forensics-extracting-data-from-an-unmanned-aeri [Accessed 11 Nov. 2019].

[6] Infosec Resources. (2019). *Commercial Computer Forensics Tools*. [online] Available at: https://resources.infosecinstitute.com/category/computerforensics/introduction/commercial-computer-forensics-tools/ [Accessed 11 Nov. 2019].

[7] AccessData. (2019). *Forensic Toolkit (FTK)®*. [online] Available at: https://accessdata.com/products-services/forensic-toolkit-ftk [Accessed 11 Nov. 2019].

[8] Medium. (2019). *Digital Forensics is ready for its latest challenge: Drones*. [online] Available at: https://medium.com/@haniahshafi/digital-forensics-is-ready-for-its-latest-challenge-drones-936c1418e928 [Accessed 12 Nov. 2019].

[9] Courcier, S. (2019). *An Introduction To Challenges In Digital Forensics*. [online] Forensic Focus - Articles. Available at: https://articles.forensicfocus.com/2017/06/29/an-introduction-to-challenges-in-digital-forensics/ [Accessed 12 Nov. 2019].

[10] Infosec Resources. (2019). *The Mobile Forensics Process: Steps & Types*. [online] Available at: https://resources.infosecinstitute.com/category/computerforensics/introduction/mobile-forensics/the-mobile-forensics-process-steps-types/#gref [Accessed 12 Nov. 2019].

[11] T3k-forensics.com. (2019). *10 challenges in Mobile Forensics | T3K-Forensics GmbH*. [online] Available at: http://www.t3k-forensics.com/allgemein-en/10-main-challenges-in-mobile-forensics2/ [Accessed 12 Nov. 2019].

[12] Heliguy.com. (2019). *DJI Launch Local Data Made for Private Operations | Heliguy*. [online] Available at: https://www.heliguy.com/blog/2017/10/04/djis-privacy-mode-launched/ [Accessed 12 Nov. 2019].

[13] Tallón-Ballesteros, A. and Riquelme, J. (2014). Data Mining Methods Applied to a Digital Forensics Task for Supervised Machine Learning. *Studies in Computational Intelligence*, pp.413-428.