Research of Video Steganalysis Algorithm Based on Machine Learning

**Shay Peretz Gilad Livshitz**

**Abstract:**

In recent times, the occurrence of malware attacks are increasing at an unprecedented rate. Particularly, the image-based malware attacks are spreading worldwide and many people get harmful malware-based images through the technique called steganography. In the existing system, only open malware and files from the internet can be identified. However, the image-based malware cannot be identified and detected. As a result, so many phishers make use of this technique and exploit the target. Social media platforms would be totally harmful to the users. To avoid these difficulties, Machine learning can be implemented to find the steganographic malware images (contents). The proposed methodology performs an automatic detection of malware and steganographic content by using Machine Learning. Steganography is used to hide messages from apparently innocuous media (e.g., images), and steganalysis is the approach used for detecting this malware. This research work proposes a machine learning (ML) approach to perform steganalysis for video files. In the existing system, only open malware and files from the internet are identified but in the recent times many people get harmful malware-based images through the technique called steganography. Social media platforms would be totally harmful to the users. To avoid these difficulties, the proposed Machine learning has been developed to appropriately detect the steganographic malware images (contents). Father, the steganalysis method using machine learning has been developed for performing logistic classification. By using this, the users can avoid sharing the malware images in social media platforms like WhatsApp, Facebook without downloading it. It can be also used in all the photo-sharing sites such as google photos.

This article explores the use of machine learning for steganography malware detection in the least significant bit (LSB) of digital image

**Introduction:**

By definition, steganography is a technique or art of concealing a type of data within a different type of data. The word steganography derives from the Greek words stegano (sealed) and graph (writing), thus meaning "writing a sealed message. The technique was historically used by governments to hide sensitive information. One interesting form of steganography sends and receives the secret messages publicly. There is no way to discover the hidden message without taking help from the sender and receiver. Since the secret message is embedded in the cover file, anyone observing it as an ordinary file does not notice that the cover file contains secret information, with this action they are making steganography even more secure. The person who knows whether the cover file contains secret Video steganography means normal videos embedded with an audio track. Audio Steganography means normal audio file embedded with an audio track. A picture file is used for hiding the secret message. The picture folder serves as the cover for the secret message, and the hidden information gets embedded in the picture file. The resulting image gets shared through any communication passage, and the details will be viewed by others. Video Steganography is a method of hiding a secret message in a video. A video is a collection of frames. One way to hide a message is to use a video file as a cover file. In Audio steganography, the cover file is a sound file containing the secret message, which is embedded within the audio file. Given that the video is big and has high redundancy, it will be suitable for steganography. Steganography involves concealing a message behind a cover file, usually a draft file of another layout, and the secret message is linked in a character sequence. Machine learning and logistic regression algorithm is mainly used to classify the Grayscale images packed with unsecured contents. It can also be used for destroying the contents, which are affected by steganography. The main feature of this proposed system is to identify the image-based malware using the technique called steganography. High stego contents can be separated into multiple stages and it can be detected by using this process. This process is more complicated because of using image bits. To simply the process regression algorithm in machine learning is used.

Machine learning has emerged as a promising approach for steganography malware detection, as it can be used to identify subtle patterns and anomalies within image data that might not be visible to the human eye. In this article, we provide a comprehensive overview of the latest research on machine learning for steganography malware detection. We discuss the various machine learning techniques that have been applied to steganography malware detection, as well as the features and metrics that have been used to train and evaluate these models.

**Related Work:**

**"Detecting Hiding Data in Video Using Steganalysis**"[[1]](#footnote-1), the authors proposed a method to detect hidden data in video frames. The method uses motion vector analysis to extract features from the video frames. These features are then used to train a decision tree classifier (SVM) that can accurately distinguish between original and stego frames.

**Microsoft researchers work with Intel Labs[[2]](#footnote-2)[[3]](#footnote-3)** proposed a deep learning approach for malware classification. Metadata analysis involves examining the attributes and characteristics of a file, such as file size, file type, and digital signature, without executing the code. This can help identify potential malicious behavior or code patterns that are associated with known malware. However, the researchers also explored the use of image analysis to detect malware. They converted binary executable files into grayscale images, where the textural and structural patterns can be used to classify binaries as either benign or malicious. By using transfer learning from computer vision, they were able to train deep learning models to accurately classify malware images and achieve high detection rates. The method uses convolutional neural networks (DNNs) to extract features from the malware binaries. These features are then used to train a classifier that can accurately distinguish between benign and malicious binaries. The experimental results showed that the proposed method achieved high accuracy in malware classification with low false positives.

**In "A Study on Deepfake Detection Using Convolutional Neural Network and Capsule Network"[[4]](#footnote-4)**, the authors proposed a method for detecting deepfake videos using a convolutional neural network (CNN) and a capsule network. The CNN is used to extract features from the frames, while the Recurrent Neural Network (RNN) - specifically, the Long Short-Term Memory (LSTM) network - is used to model the temporal dependencies between the frames of the video. The LSTM network is trained to predict the probability of each frame being real or fake, based on the information from previous frames.

**Proposed System:**

In our work we use different types of LSB steganography for insert the hidden text into the frames of the video. Regardless to former work we accept the images in RGB format and use a various types of images and not only faces.

For classify the video files, we use in CNN model and gausian method/

**Conclusion**

**ראש הטופס**

**תחתית הטופס**

1. " Research of Video Steganalysis Algorithm Based on H265 Protocol" , Kaicheng Wu, 2015 [↑](#footnote-ref-1)
2. "Deep Transfer Learning for Static Malware Classification" Li Chen Security and Privacy Research, Intel Labs [↑](#footnote-ref-2)
3. https://www.microsoft.com/en-us/security/blog/2020/05/08/microsoft-researchers-work-with-intel-labs-to-explore-new-deep-learning-approaches-for-malware-classification/ [↑](#footnote-ref-3)
4. " Deepfake Video Detection Using Recurrent Neural Networks", David Güera; Edward J. Delp [↑](#footnote-ref-4)