High-speed detection technique for encrypted steganography using GPU*

WonSeok Choi[1†], Kyung-Mo Sung[1], Laihyuk Park[2], Woongsoo Na[3]
[1] Telecommunications Technology Association, Korea
[2] Dept. of Computer Science and Engineering, Seoul National University of Science &Technology, Korea
[3] Department of Software, Kongju National University, Korea.
{wschoi, skm}@tta.or.kr, lhpark@seoultech.ac.kr, wsna@kongju.ac.kr

Abstract

This paper demonstrates that by applying (1) classification and channel multiplexing of the training dataset, (2) a three-layer filter, and (3) a training parameter tuning algorithm, it is possible to detect steganographic images with higher accuracy, improving upon the limitations of the existing SRNet algorithm. The experimental results showed a detection accuracy of 50.1%, an improvement over SRNet's 37.12% accuracy.

1 Introduction

Steganography is a technique of inserting data into other data, widely used to conceal information. It is a technology for hiding specific information in digital media (music files, image files, video files) by encrypting it.The word "steganography" originates from the Greek steganographia, which combines steganós (σ τ ε γ α ν ός), meaning "covered or hidden," and -graphia (γ ρ α φ ή), meaning "writing." Steganography is extremely difficult to detect using common methods, boasting such high security that it was even reportedly used during the 9/11 attacks. Leveraging this advantage, it is widely used for digital watermarking techniques and is also utilized for various purposes such as confidential communication and data integrity verification. Recently, however, its use for malicious purposes has been an increasing trend. These advantages are exploited by hackers to hide malware in ordinary image or document files, bypassing firewall or antivirus detection. It is also used to secretly exfiltrate confidential information by hiding it in media like image or audio files.

This paper proposes a high-speed technique utilizing machine learning to determine whether steganography has been applied to an image. Although this method is currently limited to images, it is anticipated that the algorithm can be applied similarly to various other digital media.

* Proceedings of the 9th International Conference on Mobile Internet Security (MobiSec'25), Article No. P-50, December 16-18, 2025, Sapporo, Japan.
† Corresponding author: Intelligent Network Dept., Telecommunications Technology Association, Bundang-gu, Seongnam-city, Gyeonggi-do, 13591, Korea, Email: wschoi@tta.or.kr

2 Algorithm and Test Result

This paper proposes a high-speed method utilizing GPUs to determine the presence of steganography in images. Conventional methods involved manually checking LSB (Least Significant Bit) modulation values or using various forensic tools to check for modifications in headers or specific data areas; these methods were very time-consuming and had a low detection probability. Recently, algorithms for steganalysis (detecting steganography) based on differences between neighboring pixels have been actively researched, and various detection models have been proposed. The most representative of these is SRNet(Steganalysis Residual Network), proposed in the paper "Deep Residual Network for Steganalysis of Digital Images." Before the emergence of SRNet, steganalysis was dominated by methods using manually engineered 'features' (e.g., Rich Models), which suffered from a very high false positive rate. Since secret messages are hidden in the subtle 'noise' domain of an image rather than its 'content' (like people or landscapes), capturing this noise is crucial. SRNet successfully demonstrated that steganography can be detected by eliminating manual feature engineering and adopting an 'End-to-End' deep learning approach.

This paper proposes a technique that improves the detection rate by addressing several shortcomings of SRNet. First, we train a CNN model using a GPU, with clean (non-steganographic) and steganographic images labeled accordingly. While conventional CNN models typically use pooling layers, steganalysis requires the detection of very subtle noise in the image. For this reason, if a standard CNN model is used, the pooling layers—which retain major features while discarding minor information—can eliminate the steganographic information entirely. To solve this problem, we constructed the model with a fixed Stride of 2, similar to the method researched in the original SRNet. Furthermore, the original SRNet used a dataset composed of pairs of original images and their corresponding steganographic versions. This setup can cause the model to learn incorrect associations based on the inherent features of the images themselves (content) rather than the steganographic artifacts. Therefore, we modified the training process to use randomly shuffled original and steganographic images, training on (image, label) tuples rather than image pairs. Additionally, in this model, we prevented the loss of subtle noise during preprocessing by scaling using float32 tensors, rather than simple scaling to the [0, 1] range. We also added a non-trainable high-pass filter layer to the existing filters and increased the input channels to 3 for training. For hyperparameter tuning, Weight Decay was added to prevent overfitting during training. For training, we utilized the BOSSBASE dataset, which is well-known for testing steganography algorithms. A total of 18,000 images were used for training (9,000 original images and 9,000 steganographic images), and testing was conducted on a test set of 2,000 images (1,000 original and 1,000 steganographic). Compared to the existing SRNet's detection accuracy of 37.12%, the proposed algorithm detected various steganographic images with an accuracy of approximately 50.1%. Furthermore, regarding training iterations, SRNet required $5x10^5$ iterations. In contrast, the proposed algorithm achieved higher accuracy than SRNet while training for only about 500 iterations and using approximately 5GB of GPU memory. For future work, we plan to construct and validate datasets specific to various techniques, such as LSB and DCT-based steganography. We expect this will lead to the development of a more sophisticated algorithm.

Acknowledgments

This research was supported by the MSIT (Ministry of Science and ICT), Korea, and supported by the IITP (Institute of Information & communications Technology Planning & Evaluation). (No.2022-0-00979, Development of technology and test criteria for evaluating the security of self-driving vehicle data and V2X communication network.).

References

[1] Mehdi Boroumand, Mo Chen, Jessica Fridrich., “Deep Residual Network for Steganalysis of Digital Images”, 2019, IEEE Transactions on Information Forensics and Security, Volume: 14, Issue: 5