Le projet ALASKA : Un mélange intéressant d'avancées théoriques avancées théoriques et de sciences ouverte

Rémi Cogranne

The ALASKA project: An interesting mix of theoretical advances and open science

Rémi Cogranne
Université de Technologie de Troyes
LIST3N
Troyes, France
remi.cogranne@utt.fr

***Abstract*—This article provides a brief review of the ALASKA project (ANR-18-ASTR-0009: https://alaska.utt.fr), which is continuing through the European Horizon 2020 project UNCOVER (No 101021687). The interest in presenting this project lies mainly in the two very different, but complementary, aspects of the project which aimed at applying scientific methods in media security "outside of academic conditions".**

I. Introduction and Context

Steganalysis covers all the techniques used to conceal a message in a digital medium. The operation of inserting the message must not change the properties of the media (generally a digital image) so that the concealment of the data, and thus the confidential communication, remains as stealthy as possible. Steganalysis, on the other hand, concerns techniques for detecting the presence of hidden information in media.

Steganography and steganalysis are still mainly studied in the context of academic laboratories and therefore in very restricted conditions, not even representative of the practical ones that some might find by inspecting media on the Internet or in a *firewall*. The goal of the ALASKA project (*Application on LArge and heterogeneous images database of Steganalysis techniK for Advances into the wild*) was to take steganography and steganalysis out of their academic context, through two different but complementary approaches. Let us specify that the main difficulty in the operational implementation of the current steganography and steganalysis methods lies in the incredible diversity of digital images (to speak only of this type of media) according to their origin (camera), the conditions of acquisition of the images and the post-acquisition treatments.

II. Theoretical advances for practical uses

We have considered two very different but complementary approaches in order to bring academic work closer to practical use cases. The first of these approaches, essentially statistical and methotodological, aimed at characterizing qualitatively and quantitatively the factors generating variability in steganalysis methods using artificial intelligence. However, AI-based methods, although very efficient in the field of hidden information detection, see their performances being very largely conditioned by the use of a training base which must be as close as possible to real conditions. It is particularly noteworthy that in the field of steganalysis the objective is to detect a "very weak" signal hidden in an environment that masks it in a very important way (the media content). Before our work, there was practically no study analyzing which factors allow to define coherent sources of images (i.e. on which the performances of a steganalysis method are similar). It is clear that without knowing how to define a "source" of similar images it is inconceivable to train a hidden information detection method that is adapted to the characteristics of this "source".

Our first work was experimental in nature: using images from many different cameras and simulating an image processing chain, we were able to show that it is essentially the post-acquisition processing of a digital image that most defines "how to detect hidden information". We then hypothesized that the correlation between neighboring pixels, which is the fundamental characteristic allowing to detect hidden information, is mostly impacted by these processes. As is often the case in this type of experimental study, our results need to be confirmed and extended in order to better understand the relationship between pixel correlation and the learning transfer capacity of steganalysis.

However, we were able to pursue this work in a steganography approach: our study confirming how AI-based steganalysis is essentially aimed at detecting changes in the correlation of neighboring pixels we have (1) statistically demonstrated that the optimal steganographic signal is the one that has the same covariance matrix as the original pixels and (2) developed a method for estimating the pixel correlation as a function of the post-acquisition processing chain in order to propose a new steganography method. This work has shown that this approach, although relatively methodological and difficult to apply in real conditions, compares very favorably with the state of the art, as long as the steganography has RAW images allowing to estimate the processing chain [1]–[3]

III. Applications to the open science context

The aim of this ALASKA project was to allow the members of the consortium to study the possibilities of steganalysis in real conditions, but also to draw the attention of the whole community to the difficulty that this represents, in particular

The work presented in this paper received funding from French National Research Agency (project "ALASKA" : https://alaska.utt.fr, grant No ANR-18-ASTR-0009) and from the European Union's Horizon 2020 research and innovation programme (project "UNCOVER", grant No 101021687).

because of the simplifying assumptions used in the academic world (i.e. the very vast majority of the works use RAW images coming from the same camera and processed in the same way then largely resized and converted into grey levels). With this objective in mind, we proposed two steganalysis competitions in conditions that we considered closer to the field reality, without however having the great complexity, at the risk of making the competition unsuccessful. For that we have built a base of $80,000$ RAW images coming from more than 50 cameras, almost half of which are mobile phones (the latter having become widely used for taking photographs). We also developed a RAW image development script that simulates post-acquisition processing while maintaining control over the level of diversity of the resulting images. All these data were published on the project website https://alaska.utt.fr. These images were provided in RAW and JPEG formats and will remain available to the community (under CC-BY-ND license). For the contest, images were available with and without hidden information (using different algorithms from the academic community) and an additional $5,000$ worth of test images from the same "source" were provided. Participants were asked to rank these test images from most "likely" steganographic to most "likely" healthy. The first contest that served as a "benchmark test" was held as part of a special session at the ACM Information Hiding and Multimedia Security conference [4]; among other things, one of the participating teams demonstrated a highly original and extremely effective attack on images compressed to the highest JPEG quality [5], [6].

The second contest [7] was opened on the platform Kaggle dedicated to competitions in the AI community. We managed to offer a $\$25,000$ prize for the top three teams (split into $\$12,000$ for the winner, $\$8,000$ for the runner-up and $\$5,000$ for the third place). This second challenge was an unqualified success with more than $1,000$ teams and $2,500$ participants (participants can form a team of up to 5 people) and a lot of information shared in the discussion forum. This resulted in steganalysis methods whose performance significantly exceeded the state of the art at the time and remains the benchmark today; in addition, the main lessons the community learned from this competition are the following : (1) the usual AI methods can be adapted to detect hidden information in a very efficient way (2) learning on an extremely large and diverse training base allows current AI methods to circumvent only partially the heterogeneity of the sources (3) the use of "pre-training" models on very large bases for object recognition is extremely interesting in terms of computation time For more details, the reader can consult the articles [7]–[9]. Finally, we wished to decrease the steganography community advantage by using a new insertion method and for that we used the method presented in section II while simplifying very largely the statistical model by neglecting the pixel correlation (which cannot be estimated blindly on a given image). To our surprise, we found that this method still performs favorably compared to the state of the art which shows the margin of progress that the better consideration of the correlation between neighboring pixels could offer [10], [11].

Let us also note that, in the framework of the European project UNCOVER, we will organize a third and last challenge. This last one will use images from the Internet, and thus with a much higher heterogeneity, but we will use steganography software popular on the Internet which are much less secure than the methods used in the academic community. Furthermore, we will provide a larger test base in order to favor proposals that allow to obtain a very low false-positive rate, which is a crucial issue for the use of steganalysis in practice, but which remains little studied in the academic literature (and in the ML and AI communities in general).

IV. Conclusion

Within the framework of the ALASKA project, the main objective was to bring the current academic working methods in steganography and steganalysis closer to real conditions where the diversity of media, in particular, is far too complex. Although we cannot claim to have solved all the difficulties raised by an application in operational conditions, we proposed two different approaches, one rather theoretical and statistical to explain and take into account the very large diversity of images in steganalysis and the second one with an open and participative science vocation. Both approaches allowed us to obtain satisfactory results and, above all, to venture on paths that we were not familiar with and that remain unusual in research, but extremely interesting.

References

[1] Q. Giboulot, R. Cogranne, and P. Bas, "JPEG Steganography with side Information from the Processing Pipeline," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, ser. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), IEEE, Ed. Barcelone, Spain: IEEE, May 2020. [Online]. Available: https://hal-utt.archives-ouvertes.fr/hal-02470179

[2] Q. Giboulot, P. Bas, and R. Cogranne, "Synchronization minimizing statistical detectability for side-informed jpeg steganography," in *Information Forensics and Security (WIFS), IEEE 12th International Workshop on*, December 2020, p. 4.

[3] Q. Giboulot, R. Cogranne, and P. Bas, "Detectability-based JPEG steganography modeling the processing pipeline: The noise-content trade-off," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2202–2217, 2021.

[4] R. Cogranne, Q. Giboulot, and P. Bas, "The alaska steganalysis challenge: A first step towards steganalysis," in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, ser. IH&MMSec'19. New York, NY, USA: ACM, 2019, pp. 125–137. [Online]. Available: https://alaska.utt.fr

[5] Y. Yousfi, J. Butora, J. Fridrich, and Q. Giboulot, "Breaking alaska: Color separation for steganalysis in jpeg domain," in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, ser. IH&MMSec'19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 138–149. [Online]. Available: https://doi.org/10.1145/3335203.3335727

[6] J. Butora and J. Fridrich, "Reverse jpeg compatibility attack," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1444–1454, 2020.

[7] R. Cogranne, Q. Giboulot, and P. Bas, "Alaskav2: Challenging academic research on steganalysis with realistic images," in *Information Forensics and Security (WIFS), IEEE 12th International Workshop on*, December 2020, p. 4.

[8] Y. Yousfi, J. Butora, E. Khvedchenya, and J. Fridrich, "Imagenet pre-trained cnns for jpeg steganalysis," in *Information Forensics and Security (WIFS), IEEE 12th International Workshop on*, December 2020, p. 4.

[9] K. Chubachi, "An ensemble model using cnns on different domains for alaska2 image steganalysis," in *Information Forensics and Security (WIFS), IEEE 12th International Workshop on*, December 2020, p. 4.

[10] R. Cogranne, Q. Giboulot, and P. Bas, "Steganography by minimizing statistical detectability: The cases of jpeg and color images," in *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security*, ser. IH&MMSec'20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 161–167. [Online]. Available: https://doi.org/10.1145/3369412.3395075

[11] R. Cogranne, Q. Giboulot, and P. Bas, "Efficient steganography in jpeg images by minimizing performance of optimal detector," *IEEE Transactions on Information Forensics and Security*, pp. 1–16, 2021.