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

Rémi Cogranne

▶ **To cite this version:**

Rémi Cogranne. The ALASKA project: An interesting mix of theoretical advances and open science. ERI Network & Telcom - EUt+ Workshop, 2022. hal-03625585

HAL Id: hal-03625585

https://hal-utt.archives-ouvertes.fr/hal-03625585

Submitted on 31 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The ALASKA project: An interesting mix of theoretical advances and open science[1]

Rémi COGRANNE

Université de Technologie de Troyes
LIST3N

Topic: Cybersecurity in Telecommunication and Networks ; Signal processing applied to telecommunications
Keywords: Statistical detection, Security, Secret communication, Image processing, AI.

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".

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 (in the vast majority of cases, a digital image) so that the embedding of secret data and, thus, the confidential communication remains as stealthy as possible. Steganalysis, on the other hand, referred to all 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, often not representative of the practical context that one may find by inspecting media on the Internet or in a *firewall*.
The goal of the ALASKA project (***A**pplication on **LA**rge and heterogeneous images database of **S**teganalysis techni**K** for **A**dvances "into the wild"*) was to take steganography and steganalysis out of their academic context, through two different but complementary approaches. Let us state first that the main difficulties when shifting the current steganography and steganalysis methods into an operational context lies in the extremely large diversity of digital images (to focus only of this type of media) according to their origin (camera), the acquisition settings and the post-acquisition processing.

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 one is based essentially on statistical methods and aims at characterizing qualitatively and quantitatively the factors generating discrepancies in steganalysis results using artificial intelligence. While AI-based methods, although very efficient in the field of hidden information detection, their performances largely depends on the training base which must be as close as possible to real conditions. This particularly holds true in the field of steganalysis whose objective is to detect a "*very weak*" signal hidden in a complex media content that can change radically. Before our works, there was nearly no study analyzing which factors allow to define consistent *sources of images* (i.e. on which the training leads roughly to the same results). It is clear that without knowing how to define a "*source*" of similar images it is out of reach to train a steganalysis method that is adapted to the properties of this "*source*".
Our first work was mostly of 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

[1]The work presented in this paper was funded by National Research Agency (ANR-18-ASTR-0009, https://alaska.utt.fr) and by European Union's Horizon 2020 research and innovation program (project "UNCOVER", grant No 101021687).

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 often for this type of experimental studies, our results need to be confirmed and extended in order to understand better the relationship between pixel correlation and the transfer capacity of machine learning based steganalysis.

However, we were able to continue this work in an adversarial approach for steganography: our studies 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 pixels correlations as a function of the post-acquisition processing chain in order to propose a new data hiding method. This method outperforms significantly the current state of the art; however, it remains hardly possible to apply in practice, in real conditions, as it requires RAW images to estimate the processing pipeline and its impact on pixels correlations [1,2,3].

Applications to the open science context:

The aim of this ALASKA project was to allow the members of the consortium to study how to apply steganalysis in real-life contexts, but also to draw the attention of the whole community to the difficulty that this represents, in particular because of the simplifying assumptions used in the academic world (*i.e.* a 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 gray levels…).

With this objective in mind, we proposed two steganalysis challenges under conditions that we considered closer to the real-life without, however, having its great complexity, at the risk of making the challenge unsuccessful. For this purpose, we have built a base of 80,000 RAW images coming from more than 50 cameras, among which half are smartphones (in order to catch up with their recent massive use in photography). We also developed a RAW image development script that simulates post-acquisition processing while keep controlling the level of diversity of resulting images. All these materials, JPEG as well as RAW images development script and embedding simulators, were provided on the project website (alaska.utt.fr) 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 set 5,000 of images, from the same "*source*", was used as the evaluation testing set. Participants were asked to rank these test images from most "likely" steganographic to most "likely" covers.

The first contest that served as a benchmark 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 proposed a highly original and extremely effective attack on images compressed with the highest JPEG quality factor [5,6].

The second contest [7] was opened on the platform Kaggle (www.kaggle.com) 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 unprecedented success with more than 1,000 teams and over 2,000 participants (who can gather into teams of up to 5 competitors) and a lot of information shared in the dedicated forum. This challenge leads to the development of steganalysis methods whose performance significantly outperform the state of the art and remains the current reference; in addition, the main lessons the community learned from this challenge 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 CNNs models "pre-trained" on very large sets of images for pattern recognition is extremely efficient ; For more details, the reader is referred to the papers [7,8,9].

Eventually, we wished to decrease the advantage of the steganography community by using a new embedding method and, to this aim, we used the method presented in the previous section while simplifying very much the statistical model by neglecting pixel correlations (which cannot be estimated blindly on a given image). Surprisingly, we have found that this method still outperforms

the state of the art which shows the margin of progress that better considering the correlation between neighboring pixels could offer [10,11].

Let us also conclude by announcing 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 images test set in order to favor proposals that reach high detection accuracy with a very low false-positive rate (another 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).

Conclusion

Within the framework of the ALASKA project, the main objective was to bring the academic community in steganography and steganalysis closer to realistic conditions where the diversity of media, in particular, is far too complex. Although we cannot claim to have solved all the difficulties raised such 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 towards open and participative science. Both approaches allowed us to obtain satisfactory results and, above all, to discover paths that we were not familiar with and that remains 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 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Barcelona, Spain, May 2020.

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

[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 Proc. ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec), 2020, pp. 125-137. 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 Proc. ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec), 2019, pp. 138-149.

[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 IEEE Information Forensics and Security (WIFS), December 2020.

[8] Y. Yousfi, J. Butora, E. Khvedchenya, and J. Fridrich, "Imagenet pretrained CNNs for jpeg steganalysis," in IEEE Information Forensics and Security (WIFS), December 2020.

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

[10] R. Cogranne, Q. Giboulot, and P. Bas, "Steganography by minimizing statistical detectability: The cases of JPEG and color images," in Proc. ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec), 2020, pp. 161-167.

[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.

Additional information

Contact details: Rémi COGRANNE : remi.cogranne@utt.fr

Researcher profile on the web: ORCID ; google Scholar ; UTT webpage

Member of labs / working groups / institutes: LIST3N lab ; UT Troyes

Topics of research: Detection methods for network security ; malicious behavior detection

Interest in the institute: Find collaboration within my topic of research (or perhaps explore novel research opportuinities)