Les projets ALASKA et UNCOVER: d'intéressants mélanges d'avancées théoriques et de sciences ouvertes

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

***Abstract*—Cet article propose un bref bilan du projet ALASKA (ANR-18-ASTR-0009 : alaska.utt.fr) qui se poursuit au travers du projet Européen Horizon 2020 UNCOVER (No 101021687 : www.uncoverproject.eu). L'intérêt de présenter ce projet réside principalement dans les deux aspects très différents, mais bien complémentaires, qui le composait visant notamment une application des méthodes scientifiques en sécurités des médias "hors des conditions académiques".**

I. Introduction et contexte

La stéganalyse regroupe l'ensemble des techniques permettant de dissimuler un message dans un média numérique. L'opération d'insertion du message ne doit pas changer les propriétés du média (généralement une image numérique) afin que la dissimulation des données, et donc la communication confidentielle, demeure la plus furtive possible. A contrario, la stéganalyse concerne les techniques permettant de détecter la présence d'informations cachées dans des médias.

La stéganographie et la stéganalyse demeurent essentiellement étudiées dans le cadre de laboratoires académiques et donc dans des conditions très restreintes, voire non représentatives de celles pratiques que d'aucuns pourraient trouver en inspectant des médias sur l'Internet ou dans un *firewall*. Le but du projet ALASKA ("*Application on LArge and heterogeneous images database of Steganalysis techniK for Advances into the wild*") était de sortir la stéganographie et la stéganalyse de son contexte académique et cela à travers deux approches différentes, mais bien complémentaires. Précisons que la principale difficulté à la mise en œuvre opérationnelle des méthodes actuelles de stéganographie et de stéganalyse réside dans l'incroyable diversité des images numériques (pour ne parler que de ce type de média) en fonction de leur provenance (appareil photographique) des conditions d'acquisition des images et des traitements post-acquisition.

Les travaux présentés dans cet article ont été financés par l'Agence Nationale de la Recherche (projet "ALASKA" : alaska.utt.fr, ANR-18-ASTR-0009) et par le programme de recherche et d'innovation Horizon 2020 de l'Union Européenne (projet "UNCOVER" : www.uncoverproject.eu n°101021687).

II. Avancées théoriques pour usages pratiques

Nous avons envisagé deux pistes de travail très différentes, mais complémentaires afin de rapprocher les travaux académiques aux cas d'utilisations pratiques. La première de ces approches, essentiellement statistique et méthotodologique, visait à caractériser qualitativement et quantitativement les facteurs engendrant de la variabilité dans les méthodes de stéganalyse exploitant l'intelligence artificielle. Or les méthodes à base d'IA, bien que très performante dans le domaine de la détection d'informations cachées, voient leurs performances être très largement conditionnées par l'utilisation d'une base d'entraînement qui doit être la plus proche possible des conditions réelles. Il est particulièrement notable que dans le domaine de la stéganalyse l'objectif soit de détecter un signal "très faible" dissimulé dans un environnement qui le masque de façon très importante (le contenu du média). Avant nos travaux, il n'y avait pratiquement aucune étude analysant quels facteurs permettent de définir des sources cohérentes d'images (*i.e.* sur lesquelles les performances d'une méthode de stéganalyse sont similaire). Il est clair que sans savoir comment définir une "*source*" d'images semblables il est inenvisageable de pouvoir entraîner une méthode de détection d'informations cachées qui soit adaptée aux caractéristiques de cette "*source*". Notre premier travail a été à visée expérimentale [1], [2] : en utilisant des images provenant de nombreux appareils photographiques distincts et en simulant une chaîne de traitement des images, nous avons pu montrer que c'est essentiellement les traitements post-acquisition d'une image numérique qui définissent le plus "comment détecter des informations cachées". Nous avons alors émis l'hypothèse que la corrélation entre pixels voisins, qui est la caractéristique fondamentale permettant de détecter des informations cachées, est majoritairement impactée par ces traitements. Comme souvent dans ce type d'études expérimentales, nos résultats doivent être confirmés et étendus afin de mieux comprendre la relation entre la corrélation des pixels et la capacité de transfert d'apprentissage de la stéganalyse.

Cependant, nous avons pu poursuivre ce travail dans

une approche de stéganographie : notre étude confirmant combien la stéganalyse à base d'IA vise essentiellement à détecter des modifications quant à la corrélation des pixels voisins nous avons (1) démontrer statistiquement que le signal stéganographique optimal est celui qui a la même matrice de covariance que les pixels originaux [3] et (2) développer une méthode d'estimation de la corrélation pixels en fonction de la chaîne de traitement post-acquisition afin de proposer une nouvelle méthode de stéganographie. Ces travaux ont montré que cette approche, bien que relativement méthodologique et difficilement applicable dans des conditions réelles, se compare très favorablement avec l'état de l'art, pour peu que la stéganographie dispose d'images RAW permettant d'estimer la chaîne de traitements [4]–[6].

III. Applications au contexte sciences ouvertes

Le but de ce projet ALASKA était de permettre aux membres du consortium d'étudier les possibilités de stéganalyse en condition réelles, mais également d'attirer l'attention de l'ensemble de la communauté sur la difficulté que cela représente, notamment en raison des hypothèses simplificatrices utilisées dans le monde académique (*i.e.* la très vaste majorité des travaux utilisent des images RAW provenant du même appareil photo et traitées de la même façon puis largement redimensionnée et converties en niveaux de gris).
Dans cet objectif nous avons proposé deux concours de stéganalyse dans des conditions que nous avons jugé plus proches de la réalité terrain sans toutefois en avoir la grande complexité, au risque de rendre la compétition infructueuse. Pour cela nous avons constitué une base de $80\,000$ images RAW provenant de plus de 50 appareils photographiques dont presque une moitié de téléphone portable (ces derniers s'étant largement répendue pour la prise de photographie). Nous avons également développé un script de développement des images RAW simulant les traitements post-acquisition tout en gardant la maitrise du niveau de diversité des images qui en résultent. L'ensemble de ces données a été publié sur le site Internet du projet https://alaska.utt.fr. Ces images étaient fournies aux formats RAW et JPEG et resteront disponibles pour la communauté (sous license CC-BY-ND). Pour le concours, les images étaient disponibles avec et sans informations cachées (en utilisant différents algorithmes issus de la communauté académique) et un lot additionnel de $5\,000$ images de test provenant de la même "source" était fourni. Les participants devaient classer ces images de test de la plus "vraisemblablement" stéganographiée à la plus "vraisemblablement" saine.
Le premier concours qui a servi de "test de référence" a été organisé dans le cadre d'une session spéciale de la conférence ACM Information Hiding and Multimedia Security [7] ; une des équipes participantes a notamment mis en évidence une attaque très originale et extrêmement efficace pour les images compressées avec le plus haut niveau de qualité JPEG [8], [9].
Le second concours [10] a été ouvert sur la plateforme Kaggle dédiée aux compétitions dans la communauté IA. Nous avons réussi a proposer un prix de 25.000\$ pour les trois meilleures équipes (se répartissant en $12\,000$\$ pour le vainqueur, $8\,000$\$ pour le second et $5\,000$\$ pour le troisième). Ce second challenge a été un succès incontestable avec plus de $1\,000$ équipes et $2\,500$ participants (les participants peuvent former une équipe comportant jusqu'à 5 personnes) et de nombreuses informations partagées dans le fourm de discussion. Cela a permis de proposer des méthodes de stéganalyse dont les performances dépassent sensiblement l'état de l'art de l'époque et qui reste la référence actuellement ; en outre, les principaux enseignements que la communauté a pu tirée de cette compétition sont les suivants : (1) les méthodes d'IA usuelles peuvent être adaptées pour la détection d'informations cachées de façon très performante (2) l'apprentissage sur une base d'entraînement extrêmement vaste et diverse permet aux méthodes d'IA actuelle de contourner en partie seulement l'hétérogénéité des sources (3) l'utilisation de modèles "pré-entrainer" sur des bases très vastes pour la reconnaissance d'objets est extrêmement intéressant sur la plan du temps de calcul. Pour plus de détails, le lecteur peut consulter les articles [10]–[12].
Enfin, nous avons souhaité diminuer l'avantage de la communauté stéganographie en utilisant une méthode d'insertion nouvelle et pour cela nous avons utilisé la méthode présentée dans la section II tout en simplifiant très largement le modèle statistique en négligeant la corrélation des pixels (qui ne peut être estimée en aveugle sur une image donnée). À notre surprise, nous avons constaté que cette méthode se comporte toujours favorablement par rapport à l'état de l'art ce qui montre la marge de progression que la meilleure prise en compte de la corrélation entre pixels voisins pourrait offrir [13], [14].

Notons également que, dans le cadre du projet Européen UNCOVER, nous allons organiser un troisième et dernier challenge. Ce dernier utilisera des images issues de l'Internet, et donc avec une hétérogénéité beaucoup plus importante, mais nous utiliserons des logiciels de stéganographie populaires sur Internet qui sont nettement moins sécurisés que les méthodes utilisées dans la communauté académique. Par ailleurs, nous fournirons une plus large base de test afin de favoriser les propositions permettant d'obtenir un très faible taux de faux-positifs, ce qui est un enjeu déterminant pour l'utilisation de la stéganalyse en pratique, mais qui reste peu étudié dans la littérature académique (et dans les communautés ML et IA en général).

IV. Conclusion

Dans le cadre du projet ALASKA, le principal objectif était de rapprocher les méthodes de travail académiques actuelles en stéganographie et stéganalyse vers les conditions réelles dont la diversité des médias, notamment, est bien trop complexe. Bien que nous ne puissions prétendre avoir résolu toutes les difficultés soulevées par une application dans des conditions opérationnelles, nous avons proposé deux approches différentes, l'une plutôt théorique et statistique pour expliquer et prendre en compte la très grande diversité des images en stéganalyse et la seconde à vocation de science ouverte et participative. Les deux approches nous ont permis d'obtenir des résultats satisfaisants et, surtout, nous ont permis de nous

aventurer sur des sentiers que nous ne connaissions pas et qui demeurent peu habituels en recherche, mais extrêmement intéressants.

REFERENCES

[1] Q. Giboulot, R. Cogranne, and P. Bas, "Steganalysis into the wild: How to define a source?" in *Media Watermarking, Security, and Forensics*, ser. Proc. IS&T, Jan 2018, pp. 318–1 – 318–12.

[2] Q. Giboulot, R. Cogranne, D. Borghys, and P. Bas, "Effects and solutions of cover-source mismatch in image steganalysis," *Signal Processing: Image Communication*, vol. 86, p. 115888, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596520300941

[3] Q. Giboulot, P. Bas, and R. Cogranne, "Multivariate side-informed gaussian embedding minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security (*submitted), pp. 1–16, 2022.

[4] 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

[5] 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.

[6] 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.

[7] 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

[8] 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

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

[10] 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.

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

[12] 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.

[13] 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

[14] ——, "Efficient steganography in jpeg images by minimizing performance of optimal detector," *IEEE Transactions on Information Forensics and Security*, pp. 1–16, 2021.