



## Stage : “Stéganalyse par deep learning via une approche atomistique pour lutter contre le cover source mismatch”

Marc CHAUMONT, Frederic COMBY, Mehdi YEDROUDJ, Ahmed OULAD AMARA  
LIRMM (Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier)  
Equipe ICAR, 161 rue Ada, 34392 Montpellier cedex 5 - France

Tel : +33 4.67.14.97.59, [Marc.Chaumont@lirmm.fr](mailto:Marc.Chaumont@lirmm.fr)

Mots clefs : Traitement d’images, Stéganographie, Stéganalyse, Machine Learning, Deep Learning.

La stéganographie / stéganalyse peut être expliquée comme un jeu à trois participants. Les stéganographes classiquement appelés Alice et Bob, souhaitent envoyer un message, dont l'existence même n'est connu que d'eux seul. La stéganalyste, généralement appelé Eve, observe les échanges d'images qui ont lieu entre Alice et Bob et cherche à déterminer si Alice et Bob communiquent [Simmons83]. La stéganographie est donc l’art de dissimuler un message dans un support pour le transmettre de manière secrète, et la stéganalyse est l’art de déceler la présence de ce message. Cette discipline dans sa version moderne, c'est-à-dire numérique, a débuté au début des années 2000.

La stéganalyse est effectué le plus souvent en reprenant les principes de [Kerckhoffs 1883] utilisés en cryptographie. Le stéganalyste a une connaissance de tous les paramètres publics ainsi qu'une bonne approximation de la distribution cible. Or, dans la "vrai vie", Eve, la stéganalyste, n'a pas accès à tous les paramètres et en particulier a une connaissance très mauvaise de la distributions des images utilisés par Alice et Bob. Dans le cadre de ce stage, nous souhaitons aborder cette stéganalyse dite "real life" / "real world" / "into the wild" [Ker et al. 2013 - Real World].

Nous souhaitons donc traiter les cas où la distribution d'apprentissage et celle de test sont différentes. Lorsque ce phénomène est présent, on parle de cover-source mismatch (CSM). Nous souhaitons en particulier aborder le problème par des approches **atomistique** (= approche partitionné, microscopique, analytique, de type divide and conquer, ou individualisé) qui considèrent que l'espace des images, l'espace de caractéristiques, l'espace des sources, etc, peut être partitionné pour ensuite y associer un classifieur pour chaque cellule de la partition.

L'objectif du stage donc de réduire l'influence du Cover-Source mismatch (CSM) par une approche atomistique dont l'idée serait de partitionner l'espace de caractéristique, d'une manière qui reste à déterminer, éventuellement de la même façon que ce que nous avons proposé dans [Pasquet et al. 2014 - Islet] et cela en utilisant des approches par Deep learning.

Une thèse sur le CSM est en cours, la stéganalyse par Deep learning est développée au sein de l'équipe [Yedroudj et al 2018 - Yedroudj-Net] [Chaumont 2018 - Deep Learning infancy], et nous avons de nombreuses pistes pour ce qui concerne le problème associé à ce stage. La compétition Alaska (<https://alaska.utt.fr/>) pourra également servir de base pour la réflexion.

Pour mener à bien ce sujet, il est préférable d'avoir certaines connaissances : en traitement des images, et/ou en classification/fouille de données, et/ou en architecture des machines/installation d'OS. Il est également intéressant d'avoir de bonnes bases en programmation et en math.

**Profil recherché** : Master (M2) ou Ecole d'Ingénieur (3ème année) ayant une bonne maîtrise de la programmation (C++, Python...), des connaissances en fouille de données / indexation / classification, traitement des images, sécurité.

**Encadrement** : Marc CHAUMONT (Enseignant Chercheur), Frederic COMBY (Enseignant Chercheur), Mehdi YEDROUDJ (doctorant), Ahmed OULAD AMARA (doctorant).

**Modalité de candidature** : Envoyez un CV, une lettre de motivation ainsi que votre relevé de notes de M1 le plus tôt possible. Après pré-sélection des candidatures, des entretiens téléphoniques ou en personne seront planifiés.

**Contacts** : Marc Chaumont ([marc.chaumont@lirmm.fr](mailto:marc.chaumont@lirmm.fr))

**Lieu du stage** : LIRMM, équipe ICAR.

**Période du stage** : 1er semestre 2019 (5-6 mois).

**Gratification de stage** : environ 550€ mois.

### **Bibliographie:**

[Simmons83] G. J. Simmons, "The prisoners problem and the subliminal channel," in *Advances in Cryptography, CRYPTO*, Aug. 1983, pp. 51–67.

[Kerckhoffs 1883] Kerckhoffs, A. (1883). *La cryptographie militaire*. *Journal des sciences militaires*, IX(3):5–83. Cité page 39.

[Ker et al. 2013 - Real World] A. D. Ker, P. Bas, R. Böhme, R. Cogramne, S. Craver, T. Filler, J. Fridrich, and T. Pevny. Moving steganography and steganalysis from the laboratory into the real world. In *Proc. 1st ACM workshop on Inf. hiding and multimedia security (IH&MMSec)*, Montpellier, France, pages 45–58, June 17-19, 2013.

[Pasquet et al. 2014 - Islet] J. Pasquet, S. Bringay, M. Chaumont, "Steganalysis with Cover-Source Mismatch and a Small Learning Database", *EUSIPCO 2014, 22nd European Signal Processing Conference 2014*, Lisbon, Portugal, sep., 2014, 5 pages.

[Chaumont 2018 - Deep Learning infancy] Marc Chaumont, " The emergence of Deep Learning in steganography and steganalysis ". Tutorial given at a research day in France, the 16th of January 2018, Poitiers, France, Journée " Stéganalyse : Enjeux et Méthodes", labelisée par le GDR ISIS et le pré-GDR sécurité, January 2018 DOI 10.13140/RG.2.2.35080.32005 01/2018, DOI: 10.13140/RG.2.2.35080.32005. This tutorial is about the infancy of CNNs in steganography / steganalysis in the period 2015 - 2017.

[Yedroudj et al 2018 - Yedroudj-Net] Mehdi Yedroudj, Frédéric Comby, and Marc Chaumont, " Yedrouj-Net: An efficient CNN for spatial steganalysis ", *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'2018*, 15–20 April 2018, Calgary, Alberta, Canada, 5 pages.