

Des millions d'images pour la stéganalyse : inutile !

Pasquet Jérôme^{2,3}, Sandra Bringay^{2,3,4} et Marc Chaumont^{1,2,3}

¹ UNIVERSITE DE NIMES, F-30021 Nîmes Cedex 1, France

² UNIVERSITE MONTPELLIER 2, UMR5506-LIRMM, F-34095 Montpellier Cedex 5, France

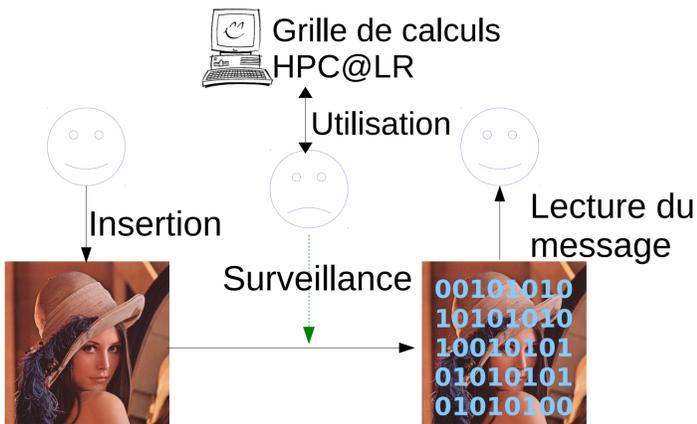
³ CNRS, UMR5506-LIRMM, F-34392 Montpellier Cedex 5, France

⁴ AMIS, UNIVERSITE MONTPELLIER 3, Route de Mende 34199 Montpellier Cedex 5, France

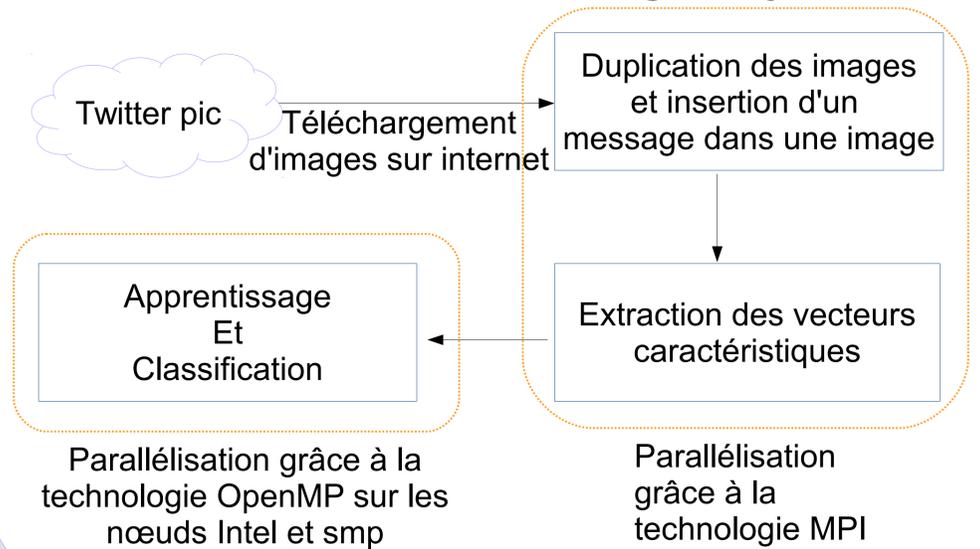
{jerome.pasquet, sandra.bringay, marc.chaumont}@lirmm.fr

Définition

La stéganalyse est l'art qui cherche à déceler la présence d'un message secret dans un support.

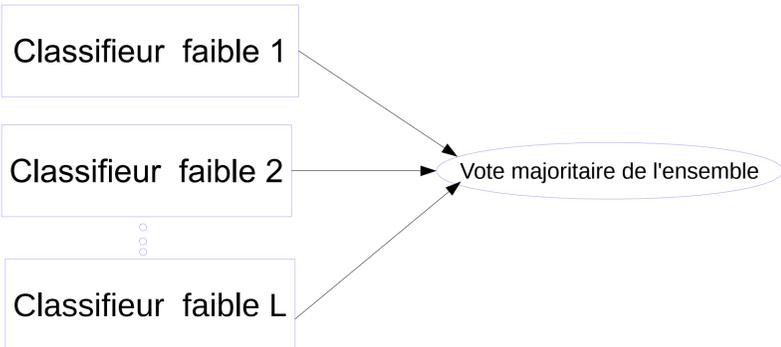


Processus de Stéganalyse



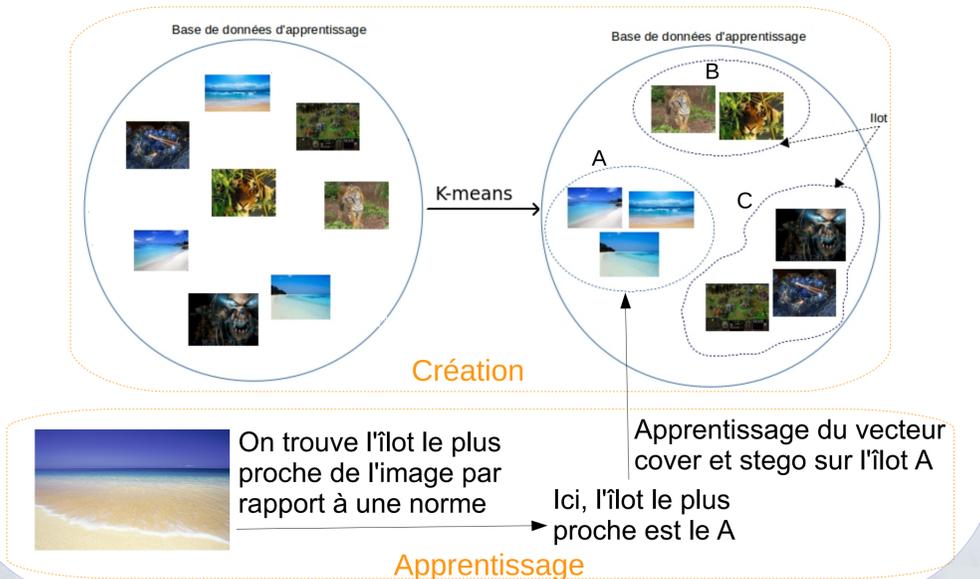
Méthode d'apprentissage automatique

Méthode utilisant un ensemble de classifieurs 'faibles', apprenant sur un sous-échantillon du vecteur caractéristique d'entrée.



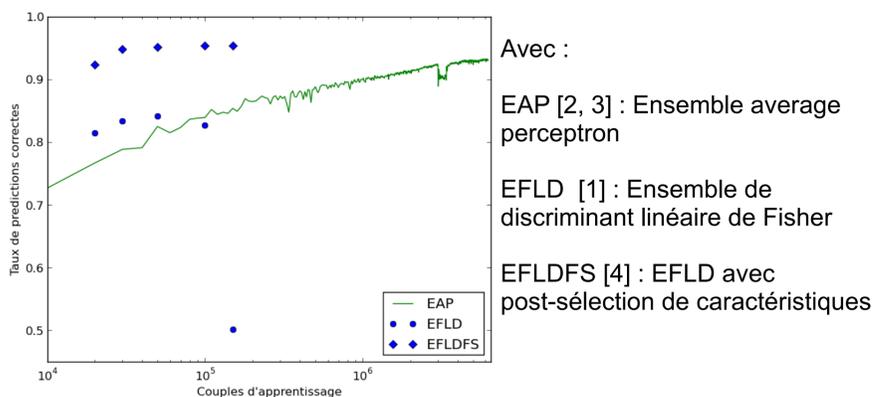
Concept d'îlots

Partitionnement de la base d'apprentissage en îlots, permettant de générer des sous-bases homogénéisées.



Expérimentations

Utilisation d'une grande base d'apprentissage très hétérogène d'environ un million d'images provenant de twitter images.



Utiliser de grandes bases de données n'est pas l'unique solution pour lutter contre le cover-source mismatch !

Tableau représentant les performances et le gain obtenu par l'approche avec des îlots de EFLDFS

Nb d'îlots	Nb d'images par îlot	Taux de prédictions correctes / gain
1	150 000	0.9539625
2	75 000	0.9581125 (+0.41%)
3	50 000	0.9583375 (+0.43%)
4	37 500	0.958275 (+0.43%)
5	30 000	0.9588625 (+0.49%)
6	25 000	0.960625 (+0.57%)
7	21 428	0.9572625 (+0.33%)
10	15 000	0.9573375 (+0.33%)

Les îlots permettent de lutter contre la diversité des images et on obtient un gain 0.5 % plus performant.

Conclusions

- L'hypothèse de Lubenko et Ker [1], selon laquelle il faut utiliser de grandes bases de données d'images pour traiter du cover source mismatch, **n'est pas la seule solution**.
- Mise en avant de l'EFLDFS [4] qui permettent de lutter très efficacement contre la **diversité des images**.
- Les **îlots** sont une contribution prometteuse afin de lutter contre le cover-source mismatch en **homogénéisant** la base de données.

Perspectives

- Utiliser plus d'images d'apprentissage et augmenter le nombre d'îlots
- Définir les îlots avec les histogrammes des images
- Définir les îlots contenant toutes les données aberrantes
- Des îlots de SVM

[1] I. Lubenko et A. D. Ker, "Steganalysis with mismatched covers: do simple classifiers help?", MM&Sec '12 (2012)

[2] I. Lubenko et A. D. Ker, "Going from small to large data in steganalysis", Proc. SPIE (2012)

[3] J. Kodovský et J. Fridrich, "Steganalysis in high dimensions: fusing classifiers built on random subspaces", Proc. SPIE (2011)

[4] M. Chaumont et S. Kouider, "Steganalysis by ensemble classifiers with boosting by regression, and post-selection of features", ICIP (2012)