14408 AYOUBA Anrezki

Analyse des invariants de dissimulation : détection de la stéganographie dans les images JPG

Depuis le lycée, ma curiosité pour la stéganographie m'a poussé à explorer les méthodes de dissimulation d'informations, et face à la difficulté de détecter automatiquement ces messages cachés, j'ai voulu approfondir la question en recherchant un invariant permettant d'identifier la présence de stéganographie dans les images JPG.

Mon étude s'inscrit dans le thème "Transition, transformation, conversion", car j'explore la transition entre une image "clean" et une image "stégo", ainsi que la transformation des données pour détecter des invariants. Cela inclut la conversion des caractéristiques visuelles en indices exploitables par des algorithmes de détection.

Positionnement thématique (ÉTAPE 1) :

- *INFORMATIQUE (Informatique pratique)*
- *INFORMATIQUE (Informatique Théorique)*
- *MATHEMATIQUES (Analyse)*

Mots-clés (ÉTAPE 1) :

Mots-clés (en français)	**Mots-clés (en anglais)**
Stéganographie	*Steganography*
Stéganalyse	*Steganalysis*
Invariant de dissimulation	*Hiding invariant*
Image JPG	*JPG image*
Apprentissage automatique	*Machine learning*

Bibliographie commentée

La stéganographie est une technique ancienne et moderne permettant de dissimuler des informations au sein de supports numériques. Elle repose sur des concepts fondamentaux de la théorie de l'information et s'appuie sur des méthodes variées et sophistiquées. Le sujet de mon étude se concentre sur les invariants potentiels permettant d'identifier des informations dissimulées dans des images JPG, indépendamment de la méthode de dissimulation utilisée. Cette recherche implique une compréhension approfondie des principes théoriques, des algorithmes, ainsi que des attaques possibles contre ces méthodes.

La théorie de l'information, développée par Claude Shannon, offre une base essentielle pour comprendre comment l'information peut être cachée dans un signal tout en minimisant la perturbation. L'article fondateur *[1]* de Claude Shannon expose les bases de la quantification de l'information, qui s'applique également à la stéganographie. Cette référence est incontournable pour établir les concepts fondamentaux liés à la capacité et à la sécurité des canaux de communication. De manière plus spécifique, Christian Cachin *[2]* propose une modélisation mathématique pour évaluer la sécurité des systèmes de stéganographie, en introduisant des notions comme l'entropie conditionnelle et la divergence de Kullback-Leibler. Ces mesures permettent de quantifier l'efficacité des algorithmes et de déterminer si une dissimulation est détectable.

Christian Cachin *[3]* fournit une description exhaustive des techniques modernes de stéganographie, en particulier celles qui opèrent sur des images numériques. Les principes des méthodes par substitution des bits de moindre poids (LSB) sont détaillés, ainsi que leurs limites en termes de robustesse face à la stéganalyse. Fridrich introduit également des stratégies visant à réduire les artefacts détectables dans les données modifiées, une problématique directement liée à l'étude des invariants. Dans un contexte plus large, Wayner Peter *[4]* explore les liens entre la stéganographie et la cryptographie. Dans son livre, il met en lumière l'utilisation conjointe de ces deux domaines pour sécuriser l'information, tout en présentant des algorithmes pratiques comme ceux basés sur la manipulation des pixels ou des fréquences. Ces informations sont essentielles pour comprendre les techniques existantes et les défis qu'elles posent à la détection.

La détection des informations dissimulées dans les images JPG est un domaine de recherche intensif. Le développement d'outils d'analyse comme StegExpose *[5]* permet d'automatiser cette détection en analysant les anomalies statistiques générées par la dissimulation. Ces outils peuvent constituer un point de départ pour identifier des invariants spécifiques aux fichiers JPG.

Des approches telles que celle de l'apprentissage supervisé *[7][8]* peuvent aussi être utilisées pour l'analyse des invariants, mais aussi pour l'identification d'invariants pertinents, notamment grâce à des modèles comme les CNN (Convolutional Neural Networks). Ces techniques permettent de détecter des motifs complexes et d'affiner les méthodes de stéganalyse, ouvrant la voie à des outils de plus en plus performants pour l'identification des informations dissimulées.

Problématique retenue

Est-il possible d'identifier un invariant de dissimulation dans des fichiers JPG, c'est-à-dire une caractéristique commune à toutes les données issues d'un processus de stéganographie, indépendamment de l'algorithme utilisé ou du type de données, permettant ainsi de détecter la présence d'information cachée dans ces fichiers ?

Objectifs du TIPE du candidat

Définir la notion d'information dans la dissimulation des images JPG.

Sélectionner plusieurs méthodes de stéganographie appliquées aux fichiers JPG.

Étudier le fonctionnement de ces méthodes de dissimulation.

Analyser les données issues de la dissimulation à l'aide de techniques statistiques.

Créer une base de données avec des images "clean" et "stego".

Implémenter un modèle d'apprentissage supervisé (CNN) pour identifier les caractéristiques des images dissimulées.

Développer un algorithme de détection des informations dissimulées.

Évaluer la performance du modèle dans la détection.

Comparer l'algorithme avec d'autres techniques existantes.

Étendre le modèle aux nouvelles méthodes de dissimulation et d'autres types d'images.

Références bibliographiques (ÉTAPE 1)

[1] SHANNON CLAUDE : A Mathematical Theory of Communication *: Bell System Technical Journal, 1948.*

[2] CACHIN CHRISTIAN : An Information-Theoretic Model for Steganography *: Information Hiding Workshop, 1998.*

[3] FRIDRICH JESSICA : Steganography in Digital Media: Principles, Algorithms, and Applications *: Cambridge University Press, 2009.*

[4] WAYNER PETER : Disappearing Cryptography: Information Hiding: Steganography & Watermarking *: Morgan Kaufmann, 2009.*

[5] STEFAN M : StegExpose, outil de stéganalyse open source *: disponible sur GitHub.*

[6] SUDHAKAR P : OpenStego, framework de stéganographie open source *: disponible sur opestego.sourceforge.net.*

[7] VINCENT BARRA, ANTOINE CORNUÉJOLS ET LAURENT MICLET : Apprentissage artificiel - 4e édition : Concepts et algorithmes *: 2021 par les Éditions Eyrolles*

[8] RÉMI COGRANNE, PATRICK BAS, MARC CHAUMONT : Stéganalyse : détection d' information cachée dans des contenus multimédias. *: Sécurité Multimédia - Partie 1 : Authentification et Insertion de Données Cachées, pages 261-303, publié par ISTE Editions.*