14408 AYOUBA Anrezki

Analyse des invariants de dissimulation : détection universelle de la stéganographie

Depuis mes années au lycée, la stéganographie a toujours suscité ma curiosité. J'avais alors conçu un site permettant de cacher des informations dans des images ou des textes grâce à différentes méthodes de dissimulation, que j'implémentais progressivement. Avec une amie, nous nous amusions à échanger des messages cachés, mais cette pratique a rapidement soulevé un défi : comment détecter si un message dissimulait ou non une information, sans devoir tester manuellement chaque méthode ? Pour y répondre, j'ai développé un bot capable de tester toutes les méthodes disponibles dans ma base de données et de décoder en fonction de chacune.

Cependant, à mesure que le nombre de méthodes de dissimulation augmentait, ce bot devenait de plus en plus inefficace, la complexité de ses tests ralentissant considérablement le processus. Ce constat a fait émerger une question plus fondamentale, qui constitue aujourd'hui la problématique de mon TIPE : existe-t-il un invariant commun à toutes les données issues de processus stéganographiques, indépendamment de la méthode utilisée ?

Positionnement thématique (phase 2)

INFORMATIQUE (Informatique pratique), INFORMATIQUE (Informatique théorique), INFORMATIQUE (Apprentissage Automatique et Statistiques), MATHÉMATIQUES (Théorie de l'Information), MATHÉMATIQUES (Analyse et Traitement des Signaux), MATHÉMATIQUES (Algèbre Linéaire et Tenseur)

Mots-clés (phase 2)

Mots-Clés (en français)	**Mots-Clés** (en anglais)
Stéganographie	Steganography
Stéganalyse	Steganalysis
Invariant de dissimulation	Hiding invariant
Entropie	Entropy
Transformée de Fourier	Fourier Transform

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 notre étude se concentre sur les invariants potentiels permettant d'identifier des informations dissimulées, même en cas de diversité des techniques de dissimulation. 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. Ce livre 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 est un domaine de recherche intensif, notamment dans le contexte des attaques. Niels Provos [**5**] offre un aperçu des techniques de détection utilisées pour analyser les images et les fichiers audio. L'objectif est de mettre en évidence des anomalies statistiques créées par la dissimulation d'informations. Andrew Ker [**6**] approfondit cette idée en se concentrant sur l'analyse des bits de moindre poids dans des images numériques. Cette étude propose des algorithmes robustes pour détecter des modifications infimes dans les données, ce qui est pertinent pour identifier des invariants exploités par différentes méthodes de dissimulation.

Enfin, des outils pratiques comme StegExpose [**7**], un logiciel open source de stéganalyse, permettent de tester et de valider les hypothèses sur des fichiers réels. Cet outil offre une base expérimentale pour appliquer les concepts théoriques et évaluer leur efficacité face à des données réelles. De manière complémentaire, les simulateurs et frameworks comme OpenStego [**8**] fournissent des environnements contrôlés pour expérimenter avec divers algorithmes de stéganographie. Ces outils permettent de modéliser les comportements des systèmes et d'étudier leurs points faibles, renforçant ainsi notre compréhension des invariants.

En combinant des approches théoriques, des études expérimentales et des outils pratiques, cette bibliographie met en lumière les défis liés à la détection des informations dissimulées. Elle constitue une base solide pour explorer les invariants, en mettant l'accent sur les relations entre les techniques de dissimulation et les signatures qu'elles laissent dans les données.

Problématique retenue

Est-il possible d'identifier un invariant de dissimulation, 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 ?

Objectifs du TIPE du condidat

1. Définir formellement la notion d'information dans le cadre de la dissimulation.

2. Sélectionner un ensemble représentatif de méthodes de stéganographie.
3. Étudier en détail le fonctionnement des méthodes de dissimulation sélectionnées.
4. Analyser les données issues de la dissimulation à l'aide de techniques statistiques.
5. Créer une base de données combinant données de couverture et résultats d'analyse statistique.
6. Implémenter un modèle de deep learning pour identifier des caractéristiques communes aux données.
7. Développer un algorithme de reconnaissance (déterministe et non déterministe) pour détecter des informations dissimulées.
8. Évaluer la performance du modèle de deep learning dans la détection de la stéganographie.
9. Comparer l'efficacité de l'algorithme développé avec d'autres techniques existantes.

Abstract

Steganography, the art of hiding information within digital media, plays a crucial role in secure communication. This study focuses on identifying invariants that persist across various steganographic techniques, enabling robust detection of hidden data. By leveraging theoretical principles from information theory and practical tools for steganalysis, we aim to characterize unique statistical signatures left by different methods of data embedding. This approach seeks to bridge the gap between the diversity of concealment methods and their detection, contributing to the development of more effective analytical frameworks for digital security.

Référence bibliographiques (phase 2)

[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] Provos, Niels, and Honeyman, Peter. Hide and Seek: An Introduction to Steganography, IEEE Security & Privacy, 2003.

[6] Ker, Andrew. Steganalysis of Embedding in Two Least-Significant Bits, IEEE Transactions on Information Forensics and Security, 2005.

[7] StegExpose, outil de stéganalyse open source, disponible sur GitHub.

[8] OpenStego, framework de stéganographie open source, disponible sur opestego.sourceforge.net.

DOT

[1] *Février 2024 : définition d'une problématique et début de la recherche documentaire*

[2] *Février 2024 : choix de trois méthodes de dissimulations (LSB, DCT, Masquage psychoacoustique dans l'audio)*

[3] *Mars 2024 : implémentation des méthodes de dissimulations choisi*

[4] *Mars 2024 : Mise en place d'une base de données d'étude*

[5] *Mars 2024 : implémentation en python d'une méthode de détection de la méthode LSB en utilisant le test du khi carré et l'apprentissage supérvisé*

[6] *Mars 2024 : implémentatation en python des algorithmes pour le traitement et la récolte des données*

[7] *Avril 2024/ septembre 2024 : implémentation d'un agorithme de deep learning pour l'analyse des données*

[8] *Septembre 2024 / décembre 2024 : Étude des resulmtats pour la mise en évidence d'un invariant de dissimulation*

[9] *Janvier 2025 : implémentatation de la méthode de détection (apprentissage supervisé et algorithme déterministe)*

[10] *Janvier 2025 : Rencontres avec des doctorants pour présenté mes résultats et une tentative délargissement de la méthodes de détection*

[11] *Février 2025 : Tentative d'élargissment du model avec de nouvelles méthodes dissimulations*

[12] *Avril 2025/ Juin 2025 : Étude des donnée et conclusion*