Stégananalyse d'images

sujet thèse 2012-2015

Marc Chaumont
LIRMM (Laboratoire d'Informatique, de Robotique et Microélectronique de Montpellier)
Equipe ICAR
161 rue Ada, 34392 Montpellier cedex 5 - France
Tel : +33 4.67.41.85.14
Fax : +33 4.67.41.85.00
Marc.Chaumont@lirmm.fr

Mots clefs : stéganalyse, stéganographie, classification, multi-classes, fusion de données, vecteur de caractéristiques, filtrage, criminalistique, stéganalyse non aveugle d'algorithmes, stéganalyse quantifitative, multi-utilisateurs (acteurs)…

La stéganographie est l'art de dissimuler un message de manière secrète dans un support anodin. La stéganalyse est l'art de déceler la présence d'un message secret. L'étude de la stéganographie/stéganalyse moderne a réellement débuté au début des années 2000. Actuellement, lorsque l'on effectue une stéganalyse, on définit un « scénario », c'est-à-dire un certain nombre d'hypothèses sur ce que le stéganalyste connait de l'environnement utilisé par le stéganographe.

Le scénario 1 « stéganalyse à clairvoyance » considère que la distribution des sources « couvertures » est connue, c'est-à-dire que l'on a à disposition suffisamment d'images (sans message), pour en déduire une distribution similaire à celle utilisée par un potentiel suspect. On suppose également que l'algorithme de stéganographie utilisé est connu, et que le *payload* (quantité de bits insérée) est connu. Ce scénario ressemble très fortement à celui choisi lors de la compétition BOSS [BOSS 2011] (cependant, la distribution des sources n'était pas totalement connue : « cover-source mismatch »). Pour ce scénario, deux propositions d'attaque très intéressantes ont été proposées dans [Fridrich2011-HOLMES] et [Gul2011].

Le scénario 2 « stéganalyse ciblée » considère que la distribution des sources « couvertures » est connue, l'algorithme de stéganographie utilisé est connu, et le *payload* est inconnu. Pour ce scénario, [Pevný2011] a montré que l'on pouvait obtenir de très bons résultats avec un classifieur ayant appris sur images stéganographiées dont le payload est distribué uniformément.

Le scénario 3 « stéganalyse universelle » considère que l'on ne connait pas l'algorithme de stéganographie utilisé, s'il y a insertion d'un message secret, que l'on ne connait pas le *payload*, mais par contre que l'on connait les sources « couvertures ». Il n'y a pas pour le moment de solution satisfaisante pour ce scénario [Pevný2008, Pevný2011].

Le scénario 4 « stéganalyse universelle avec cover-source mismatch » est un scénario réaliste dans le cadre d'un stéganalyseur automatique de trafic. Ici, l'algorithme de stéganographie utilisé est inconnu, le *payload* est inconnu, la distribution des sources couverture est connue partiellement (certaines sources ne sont pas connues). Le problème de « cover-source mismatch » est un problème ouvert.

Le scénario 5 « stéganalyse universelle avec batch (pool steganography) » est un scénario réaliste d'analyseur de trafic automatique qui analyse ce qui passe sur un réseau. On peut distinguer des acteurs dont certains utilisent certaines fois la stéganographie [Ker2011]. Ce problème est similaire au précédent, mais il existe en plus, une dimension temporelle.

L'objectif de cette thèse est de proposer des solutions pour les scénarios 3, 4 et 5.

Le scénario 3 nécessite de mettre en place des mécanismes de gestion de la nouveauté (algorithmes différents de ceux utilisés lors de l'apprentissage). Il est alors nécessaire de regarder du coté de la classification par fusion de votes, les classifieurs multi-classes, les classifieurs comme le classifieur FLD [Kodovský2011], le classifieur OC-NM [Pevný2008], le classifieur SVM [Chang2011-LibSVM], prendre en main un ensemble d'algorithmes de stéganographie [Pevný2008], et de techniques de stéganalyse [Fridrich2009-BOOK, Böhme2010], prendre en

main des caractéristiques HOLMES [Fridrich2011-HOLMES], mettre en place des protocoles d'évaluation de la sécurité.

Le scénario 4 nécessite de s'intéresser au problème de « cover-source mismatch » et donc de déterminer ce qui caractérise une distribution *cover*. En particulier il serait intéressant de se « libérer » de la dépendance forte à la distribution *cover* lors de l'apprentissage. Friddrich et al. évoquent une approche par « apprentissage sur une base contaminée » [Fridrich2011-Process]. Gul et Kurugollu attaquent le problème de manière détournée en ajoutant à la base d'apprentissage la base de test filtrée [Gul2011]. Il pourrait être intéressant de filtrer la base d'apprentissage pour être moins sensible au « cover-source mismatch ». Par ailleurs, d'autres pistes liées à l'étude de criminalistique d'appareils photo [Fridrich2008, Fridrich2009, Cortiana2011, Kang2011] peuvent également être envisagées.

Le scénario 5 tel qu'il a été étudié par A. Ker [Ker2011] se pose la question de l'efficacité/complexité d'un stéganalyseur image par image par rapport à une analyse par acteur. Pour le moment les études sont insuffisantes pour décider de ce qui est le plus avantageux et le plus efficace. Des avancées dans les scénarios 3 et 4 pourraient donner des éléments supplémentaires de compréhension.

La thèse a pour objectif de mieux comprendre et faire avancer la stéganalyse moderne. La définition de nouvelles *features* (comme celles de Holmes [Fridrich2011-HOLMES] ou de [Gul2011]), la définition de nouvelles approches de classification (comme le classifieur par ensemble de [Kodovský2011]), le filtrage de base de données, la stéganalyse quantitative, etc., sont des pistes que l'on creusera tout au long de la thèse.

La thèse se déroulera dans un premier temps par l'étude de scénario 3 « stéganalyse universelle ». Dans un second temps, le scénario 4 « stéganalyse universelle avec cover-source mismatch » sera abordé. Enfin, l'étude de ces deux scénarios devrait permettre d'obtenir des avancées sur le scénario 5.

Références :

[BOSS 2011] BOSS (Break Our Steganography System) is the first challenge on Steganalysis. The challenge started the September 9th 2010 and ended the 10th of January 2011. The goal of the player was to figure out, which images contain a hidden message and which images don't. http://www.agents.cz/boss/BOSSFinal/
The steganographic algorithm was HUGO: "Using High-Dimensional Image Models to Perform Highly Undetectable Steganography", T. Pevny, T. Filler and P. Bas, 12th Information Hiding Conference, June 28 - 30, 2010, Calgary, Alberta, Canada.

[Fridrich2011-HOLMES] J. Fridrich, J. Kodovský, V. Holub, and M. Goljan, "Steganalysis of Content-Adaptive Steganography in Spatial Domain", 13th Information Hiding Conference, IH'2011, Prague, Czech Republic, May 18–20, 2011, to appear in LNCS, Springer-Verlag.

[Gul2011] G. Gul and F. Kurugollu, "A New Methodology in Steganalysis: Breaking Highly Undetectable Steganographpy (HUGO)", 13th Information Hiding Conference, IH'2011, Prague, Czech Republic, May 18–20, 2011, to appear in LNCS, Springer-Verlag.

[Pevný2011] T. Pevný, « Detecting messages of unknown length », Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE 21th Annual Symposium on Electronic Imaging, SPIE'2011, Volume 7880, San Francisco, California, USA, Feb 2011.

[Pevný2008] T. Pevný and J. Fridrich, « Novelty Detection in Blind Steganalysis », ACM Multimedia and Security Workshop, MM&Sec2008, Oxford, UK, September 22-23, pp. 167-176, 2008.

[Ker2011] A. D. Ker and T. Pevný, "A New Paradigm for Steganalysis via Clustering", Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE 21th Annual Symposium on Electronic Imaging, SPIE'2011, Volume 7880, San Francisco, California, USA, Feb 2011.

[Kodovský2011] J. Kodovský and J. Fridrich, "Steganalysis in high dimensions: fusing classifiers built on random subspaces", Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE 21th Annual Symposium on Electronic Imaging, SPIE'2011, Volume 7880, San Francisco, California, USA, Feb 2011.

[Chang2011-LibSVM] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[Fridrich2009-BOOK] J. Fridrich, Steganography in Digital Media: Principles, Algorithms, and Applications, 1st Edition. Cambridge University Press, New York, NY, USA, 2009.

[Böhme2010] Rainer Böhme, Advanced Statistical Steganalysis, *Information Security and Cryptography Texts and Monographs Information Security and Cryptography*, Springer, 285 pages, 2010.

[Fridrich2011-Process] J. Fridrich, J. Kodovský, V. Holub, M. Goljan "Breaking HUGO: the process discovery", 13th Information Hiding Conference, IH'2011, Prague, Czech Republic, May 18–20, 2011, to appear in LNCS, Springer-Verlag.

[Fridrich2008] J. Fridrich, M. Chen, M. Goljan, and J. Lukas, "Determining Image Origin and Integrity Using Sensor Noise", IEEE Transactions on Information Security and Forensics, 3(1), pp. 74-90, March 2008.

[Fridrich2009] J. Fridrich, "Digital Image Forensic Using Sensor Noise", IEEE Signal Processing Magazine, vol. 26, no. 2, March 2009, pp. 26-37.

[Cortiana2011] A. Cortiana, V. Conotter, G. Boato, F. G. B. De Natale, "Performance comparison of denoising filters for source camera identification", Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE 21th Annual Symposium on Electronic Imaging, SPIE'2011, Volume 7880, Number 05, San Francisco, California, USA, Feb 2011.

[Kang2011] X. Kang, Y. Li, Z. Qu, J. Huang, "Enhancing ROC performance of trustworthy camera source identification", Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE 21th Annual Symposium on Electronic Imaging, SPIE'2011, Volume 7880, Number 07, San Francisco, California, USA, Feb 2011.