



جامعة الموصل
كلية علوم الحاسوب والرياضيات

كشف الاخفاء باعتماد الكشف عن الشذوذ باستخدام تصنيف الفئة الواحدة

محمد عبد الكريم محمد حسين التميمي

رسالة ماجستير
علوم الحاسوب

بإشراف

د. أحمد سامي نوري

أستاذ مساعد

الملخص

الكتابة المغطاة هو علم/فن إخفاء المعلومات بطريقة يجب ألا تجذب الانتباه الى الرسالة المخفية في الوسائط المرسله، فإذا حدث وأثير شك فيصبح الغرض من الاخفاء غير ذي أهمية. ثم ظهر جزء الاخفاء المتمم أو المضاد وهو كشف الاخفاء الذي يهدف الى الشك بالإخفاء وتحليل الوسائط المرسله ليقرر فيما إذا كانت متضمنة لبيانات مخفية او لا، أي أجريت عليها عملية إخفاء او لا.

تم في هذه الرسالة الاعتماد على آلة المتجه الداعم ذات الصنف الواحد (OC-SVM) التي تم تدريبها على صنف واحد من الصور وهو صنف الصور النظيفة ليستطيع المصنف المدرب تصنيف المشاهدات الجديدة الى صنفها الصحيح اي صنف الصور النظيفة (Clean) التي لا تحتوي على بيانات مضمنة او صور تحوي بيانات مضمّنة (stego).

تبين أن تدريب مصنّف على صنف واحد (صنف الصور النظيفة) عمل شاق ويتطلب وقت تنفيذ طويل؛ لأن تقسيم البيانات الى بيانات تدريب وبيانات اختبار بناء على قيمة أولية او قيمة بذرة (seed)، تجربة جميع تراكيب البيانات المحتملة، إعداد معاملات المصنف تقييم النموذج (model) كلها تمت بطريقة يدوية أي تمت برمجة جميع ما سبق وتجربة احتمالات كثيرة من فضاء احتمالات واسع جدا للتوصل الى نتيجة مرضية.

باستعمال لغة البرمجة بايثون ومكتبة (SciKit-Learn) الخاصة بتعليم الآلة كانت نتائج التصنيف جيدة في التقرير إذا كانت الصورة المدخلة خالية من البيانات المضمنة (clean image) او تحتوي على بيانات مضمّنة (stego image)، حيث وصلت نسبة كشف الصور النظيفة في

مجموعة البيانات الشخصية (تم جمعها وترتيبها بمجهود شخصي) والمسماة (my dataset) الى (٧٤.٢٪)، ونسبة كشف الصور المحملة ببيانات مضمّنة الى (٨٨.٨٨٪)، في حين كانت نسبة كشف الصور النظيفة (٧١.٩٪) في مجموعة البيانات الثانية المسماة (CALTECH dataset) وبلغت نسبة كشف الصور التي تم اجراء عملية تضمين عليها (٧٤.٦٦٪)، اما آخر مجموعة بيانات والمسماة (BOSS base dataset) فقد بلغت نسبة الكشف عن الصور التي لا تحمل بيانات مضمّنة داخلها (٩١.٥٥٪)، وكانت نسبة كشف الصور الحاملة لبيانات مخفية في مجموعة البيانات نفسها (٧٣٪).

تم حساب نسب الكشف بالنسبة للصور الخالية من التضمين عن طريق إيجاد النسبة المئوية لبيانات الاختبار الخالية من التضمين التي صنّفها المصنّف بشكل صحيح اما بالنسبة للصور المحدودة العدد الحاوية على التضمين تم حساب النسبة المئوية للصور التي صنّفها المصنّف بشكل صحيح كحاوية لبيانات مضمّنة بكافة احجام التضمين التي احتوتها (1KB, 2KB, 3KB, 25KB, 50KB).

**UNIVERSITY OF MOSUL
COLLEGE OF COMPUTER SCIENCES
AND MATHEMATICS**



Steganalysis Based Anomaly Detection Using One-Class Classification

Muhammad Abdul-Karim Muhammad AL-Tamimi

M.Sc. Thesis

Computer Science

Supervised by

**Dr. Ahmed Sami Nori
Assistant Professor**

Abstract

Steganography is the science/art of hiding information in a way that must not draw attention to the message hidden in the transmitted media, if a suspicion is raised then the purpose of steganography is defeated. then appeared its counterpart, Steganalysis, which aims to suspect and analyze the transmitted media to decide wither it contain an embedded data or not.

Here in this thesis we depend on a **one-class support vector machines (OC-SVM)** which has been trained on only one class of images which is the clean images class with the aim that the trained classifier can classify new reviews to their correct class i.e. clean or stego class.

Training a classifier on only one class (clean images class) turned to be hard work and required long execution time since classifier parameters tuning, data separation to train and test data based on seed value and model evaluation needed to be done manually.

Using **python** as a programing language and **SciKit-Learn** as a machine learning library the classification results have been very good in managing to decide whether an input image is clean or stego image, The results for the personal dataset in detection of clean images was (74.2%) and for the stego images it was (88.88%), as for the (CALTECH Dataset) the detection percentage of images without embedded data was (71.9%) and it was (74.66%) for images with embedding, finally the (BOSS Base dataset) detection of clean images was (91.55%) and the detection of stego images was (73%).

The detection ratio for the clean images was calculated by finding the percentage of the clean test data that was classified correctly as clean by the classifier, as for the limited number of the stego test data the percentage of the images that was correctly classified as stego was calculated for all embedding sizes (1KB, 2KB, 3KB, 25KB,50KB).