BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2024
http://excel.fit.vutbr.cz

Digitální stegoanalýza

Petr Dufek*

Abstrakt

V práci je prezentována nově navrhovaná metoda určená pro detekci LSB steganografie v obrázcích s vlastnostmi charakteristicky podobnými digitálním ikonám. Navrhovaná metoda SameColor je založena na měření a porovnávání délek sekvencí pixelů v řadě beze změny hodnoty bitu využívaného pro uložení skrytých dat. V části obrázku s obsahem steganografie se nacházejí násobně kratší sekvence v porovnání s částí obrázku, která nebyla pro uložení dat využita. V porovnání s metodou Chi-Square dosahuje nově navržená metoda vyšší úspěšnosti při detekci LSB steganografie. Zcela zásadně lepších výsledků dosahuje zejména v případě nízké míry zaplněnosti krycího obrázku. Při testování implementovaná metoda správně klasifikovala více než 99 % steganogramů a 81,5 % krycích obrázků. S využitím této metody je tedy možné poměrně spolehlivě určit, zda digitální ikona nebo jiný obrázek podobných charakteristik přenáší potenciálně nebezpečná data skrytá metodou LSB, či nikoli.

*xdufek17@vut.cz, *Fakulta informačních technologií, Vysoké učení technické v Brně*

1. Úvod

Steganografie nabízí možnosti jak ukrýt data, která mají být nepozorovaně přenesena k jejich příjemci bez povšimnutí případného pozorovatele sledujícího probíhající komunikaci. Digitální steganografie se zabývá ukrýváním dat do obrázků, zvuků, spustitelných souborů, video souborů a dalšího digitálního obsahu. Ukrytá data mohou být potenciálně nebezpečná a proto je důležitý rozvoj stegoanalýzy, která může nabídnout schopnost detekce přítomnosti těchto dat.

Ve své práci se zabývám několika existujícími metodami digitální obrazové stegoanalýzy, jejich implementací do stegoanalytického nástroje a návrhem metody nové. Nově navržená metoda by měla být schopna spolehlivě detekovat vloženou steganografii do digitálního obrázku.

Mezi existující steganografické metody je možné zařadit například metodu Chi-Square, nebo metodu PDH. Každá z nich se zaměřuje na jinou metodu steganografie a nejsou při použití vzájemně zaměnitelné. Metoda Chi-Square se zaměřuje na steganografii typu LSB a její spolehlivost není v případě nízké míry zaplněnosti krycího obrázku vloženými daty příliš vysoká. V těchto případech není schopna spolehlivě odlišit steganogram od krycího obrázku. Metoda PDH se specializuje na PVD steganografii.

Všudypřítomnými obrázky v digitálním prostoru jsou ikony. Ať už se jedná o loga webových stránek, sociálních sítí nebo ikony programového vybavení počítačů, je možné tyto obrázky využít pro utajenou komunikaci. Datová sada s obsahem LSB steganografie, která byla použita v rámci této práce, tuto skutečnost reflektuje a je složena z několika tisíců takovýchto obrázků [1].

Vyvinutá metoda SameColor je schopna poměrně spolehlivě analyzovat obrázky s charakteristikou digitální ikony a s pomocí měření délek sekvencí po sobě jdoucích pixelů beze změny bitů používaných pro uložení skrytých dat rozhodnout o jejich ne/přítomnosti.

Při testování na 8000 obrázcích z již zmiňované datové sady metoda dosáhla 95,2 % správně klasifikovaných souborů a tím překonala metodu Chi-Square o 37,26 %.

2. LSB steganografie

Steganografická metoda LSB a její modifikace jsou nejrozšířenějšími typy steganografie zejména pro její jednoduchost. Jak znázorňuje obrázek 2, metoda ukládá bity skrytých dat přímo do pixelů krycího obrázku. Vložením těchto dat jsou nahrazeny nejméně významné bity každého z barevných kanálů pixelu. V případě šedotónových obrázků je obsažen pouze je-

den barevný kanál. Touto záměnou může také dojít ke změně výsledného barevného odstínu změněného pixelu.

3. Detekce LSB steganografie

3.1 Metoda Chi-Square

Steganografii vloženou do obrázku metodou LSB je možné detekovat například metodou Chi-Square. Tato metoda prochází sekvenčně analyzovaným obrázkem a zjišťuje počet výskytů každého barevného odstínu v celé ploše obrázku. Rozhodujícím faktorem prozrazujícím použitou steganografii při použití této stegoanalytické metody je sobě se blížící počet jednoho barevného odstínu a odstínu po něm následujícího. Takový barevný odstín je na binární úrovni odlišný pouze v jednom bitu. Pokud je obrázek barevný, je nutné jej procházet odděleně po jednotlivých barevných složkách.

Jak naznačuje obrázek č. 3, metoda označí obrázek jako steganogram na základě výsledků Chí-kvadrátového testu. Pokud jsou ukrytá data do obrázku vložena sekvenčně shora a analyzovaná oblast bude taktéž shora postupně rozšiřována, po překročení hranice konce ukrytých dat dojde ke změně hodnot výstupů statistické funkce. K tomuto poklesu však dojde i v případě, že obrázek steganograficky ukrytá data neobsahuje. Avšak v tomto případě dojde k poklesu těchto hodnot v průběhu analýzy dříve než v případě, kdy byla stegangrafie na obrázek použita. Je nezbytné určit správně polohu této hranice pro konkrétní charakteristiky analyzovaných obrázků pro zajištění spolehlivé analýzy [2].

3.2 Metoda SameColor

Metoda SameColor analyzuje část obrazových dat z horní a z dolní oblasti analyzovaného obrázku. Zaměřuje se na jednotlivé pixely a měří délky sekvencí po sobě jdoucích pixelů beze změny LSB bitu, který je používán pro uložení steganografie stejnojmennou metodou. Tato metoda předpokládá, že ukrývaná data jsou do obrázku vkládána sekvenčně od jeho počátku, jak popisuje základní metoda LSB. Měření délky sekvencí probíhá ve dvou oknech znázorněných na obrázku č. 4. Pokud byl obrázek typu digitální ikony, (charakteristické většími plochami jedné barvy a jednoduchými obrazci) pozměněn LSB steganografií, je možné v zaznamenaných délkách sekvencí mezi oknem z počátku obrazových dat a oknem z jejich konce zjistit několika (typicky více než 2) násobný rozdíl. Sekvence nacházející se v části s vloženou steganografií jsou kratší než v části druhé. Jestliže nebyl obrázek modifikován, délky sekvencí se mezi analyzovanými okny podobají.

4. Porovnání metod

Graf č. 1 zobrazuje porovnání úspěšnosti metody Chi-Square a metody SameColor implementované v rámci této práce. Metoda SameColor dosahuje vyšší úspěšnosti při analýze obrázků z použité datové sady než metoda Chi-Square a to zejména v oblasti dat s velmi krátkými vloženými daty.

Literatura

[1] Cassavia, N., Caviglione, L., Guarascio, M., Manco, G. a Zuppelli, M. Detection of Steganographic Threats Targeting Digital Images in Heterogeneous Ecosystems Through Machine Learning. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. Září 2022, sv. 13, č. 3. DOI: 10.22667/JOWUA.2022.09.30.050.

[2] Fridrich, J. a Goljan, M. Practical Steganalysis of Digital Images - State of the Art. Proceedings of SPIE - The International Society for Optical Engineering. Červen 2002, sv. 4675. DOI: 10.1117/12.465263.