Programiranje v Pythonu

Izdelava programa za steganografijo slik

Matej Marinko

februar 2017

Kazalo

Povzetek

1 Uvod

2 Steganografija

Steganografija je znanost, ki omogoča skrivanje podatkov v navidezno nepomembnem prenosnem mediju. Beseda izhaja iz grščine, in pomeni "zakrito pisanje". Predstavlja skupek metod za skrivanje informacij v neke druge informacije. Za razliko od kriptografije oz. šifriranja, ki želi podatke narediti neberljive, steganografija poskuša prikriti, da ti podatki sploh obstajajo [4].

Šibkost kriptografije je, da je sporočilo sumljivo že na prvi pogled. Šibkost steganografije pa je, da ko nekdo enkrat odkrije, kje je sporočilo skrito, ga lahko enostavno prebere. V praksi se zato pogosto uporabljata obe metodi skupaj, saj druga drugi odpravita slabosti.

2.1 Tehnike

Primeri steganografije so znani že od Antike, različne tehnike pa so se ohranile vse do danes. Z razvojem računalništva se je steganografija razvila tudi na digitalnem področju.

Primeri fizičnih tehnik:

- V Stari Grčiji so sporočila skrivali v voščene tablice. Na les so napisali sporočilo, ki so ga nato prekrili z voskom. Na vosek pa so napisali nedolžno in nepomembno sporočilo.
- Drugi postopek, ki so ga uporabljali v Antični Grčiji je bil približno takšen: Sužnju so pobrili glavo in mu vtetovirali sporočilo. Ko so lasje zrasli sporočilo ni bilo več vidno. Očitna pomankljivost te motode je, da moramo počakati, da osebi lasje zrastejo nazaj.
- Sporočilo, ki je napisano z nevidnim črnilom, na nepopisanem delu pisma.
- Sporočilo je v Morsejevi abecedi, napisani na prejo, ki so jo potem vtkali kurirju v blago.
- Del črk v besedilu je napisan z drugačno pisavo kot druge črke (npr. ležeče). Te črke so tvorile skrivno sporočilo. [3]
- Nemci so med drugo svetovno vojno uporabljali mikropike (microdots). Podatke so skrivali v znakih, ki jih je bilo mogoče prebrati samo pod veliko povečavo. [4]
- Cardanovo rešeto - mreža z odprtinami, s katero prekrijemo besedilo, da se prikaže skrito sporočilo.

- Ameriški pilot Jeremiah Denton, ki ga je ujela vietnamska vojska, je bil med televizijsko konferenco prisiljen pričati, da z njim v ujetništvu ravnajo dobro. Hkrati pa je z mežikanjem v morsejevi abecedi črkoval besedo T-O-R-T-U-R-E (mučenje).[5]

Primeri digitalnih tehnik:

- Skrivanje podatkov v najnižje bite slikovnih datotek, z neopaznimi spremembami barv.
- Skrivanje podatkov v zvočni zapis, s spremembami, neslišnimi za človeško uho.
- Tehnika "pletja in presejanja" (chaffing and winnowing), kjer gre za to da paketkom, ki gredo čez nezavarovano povezavo dodamo lažne paketke, v katere lahko skrijemo sporočilo.
- Dodajanje podatkov v neuporabljene dele datoteke, npr. na konec.
- Skrivanje z uporabo unicode znakov, ki izgledajo enako kot ASCII znaki.
- Nekateri moderni tiskalniki z težko vidnimi svetlo rumenimi pikami na vsakem listu označijo serijsko številko tiskalnika in čas tiska. [3]

2.2 Steganografija digitalnih fotografij

Večja kot je datoteka, v katero nameravamo skriti naše sporočilo, v primerjavi z tem sporočilom, lažje jo je skriti. Zato so slike primerne za steganografijo, saj vsebujejo velike količine podatkov. Tako lahko skrijemo podatke npr. na Internetu. Slika je vsem na očeh, venar se nihče ne zaveda, da so v njej skrite še dodatne informacije. Ni znano, kako pogosta je ta praksa, vendar vemo da obstaja.

V RGBA zapisu uporabimo 32 bitov za vsak piksel, to pomeni 8 bitov za vsako komponento. Samo rdeča barva ima 2^8 različnih intenzivnosti. (Glej poglavje: Zapis slik v računalniku) Razliko med $10111111_{(2)}$ in $10111110_{(2)}$ v inteziteti barve človeško oko zelo težko prepozna. Zato lahko najnižje bite (least significant bits) spremenimo in v njih skrijemo svoje podatke. Če v vsakem pikslu v vseh štirih komponentah spremenimo 2 najmanj pomembna bita, lahko v en piksel skrijemo 1 bajt (8 bitov) podatkov. V celotni sliki lahko, če je dovolj velika, skrijemo tudi več MB podatkov.

Pri tem je pomembno, da slika, v katero želimo skriti podatke ni enobarvna, oziroma sestavljena iz večjih enobarvnih ploskev. Če je slika takšna,

obstaja večja možnost, da kdo opazi različne odtenke sosednjih pikslov na sliki. Zato so slike, ki se uporabljajo za steganografijo pogosto slike narave, živali, ipd.

2.3 Steganaliza

Steganaliza je veda, ki se ukvarja z zaznavanjem sporočil, skritih s pomčjo steganografije. Pogosto se povezuje s kriptoanalizo. Namen je torej najti sumljiva sporočila, ter ugotoviti ali se v njih skriva skrito sporočilo in, če je možno, prebrati to sporočilo.

Problema se ponavadi lotimo z statistično analizo. Analiziramo recimo slike, ki so bile posnete z enakim fotoaparatom, ali zvočne posnetke in ugotavljamo skupne značilnosti. Zaradi pogoste izgubne kompresije je predvidljivo, kakšni naj bi bili podatki. V JPEG sliki, naprimer, lahko precej dobro sklepamo, katere barve je piksel, če poznamo vse sosednje piksle. Ker so takšne razporeditve predvidljive, bodo slike, v katerih je skrito stenografsko sporočilo hitreje opazne.

Najlažje pa je odkriti skrito sporočilo, če imamo na voljo originalno sliko, v kateri ni skritih podatkov, saj bomo hitro opazili razliko in bomo posledično lažje izluščili podatke.

Še naprednejše metode predpostavijo, da so podatki, ki so skriti poleg tega še kriptirani. Posledica sodobne enkripcije je, da so podatki videti naključni. Večina metod skriva podatke v najmanj pomembne bite (least-significant bits). Če bo razporeditev 1 in 0 v najmanj pomembnih bitih skoraj popolnoma naključna, je velika verjetnost, da je v datoteki skrito kriptirano sporočilo.

3 Zapis slik v računalniku

Slika je v računalniku zapisana tako, da je razdeljena na drobne kvadratke, imenovane piksli. Vsak piksel ima podatke kako močno in v katerih barvah žari. Za zapis se uporablja več različnih formatov, med katerimi so najbolj znani JPEG, PNG, BMP, GIF... Med seboj se razlikujejo v različnih lastnostih, kot so način zgoščevanja, število možnih barv posameznega piksla...

Osnovna zgradba vsakega formata sestoji iz:

glave, ki vsebuje glavne podatke, kot so velikost, barvna globina in kompresijska tehnika.

slikovnih podatkov oziroma niza pikslov. Podatki so lahko kompresirani ali nekompresirani.

polj za metapodatke (metadata), kot so datum posnetka, avtor slike...

3.1 Format PNG

PNG (Portable Network Graphics) je razmeroma nov slikovni format, ki je popularen predvsem na spletu.

Format PNG uporablja brezizgubno kompresijo, kar nujno potrebujemo pri steganografiji, saj se zanašamo da bomo podatke skrili tako, da jih človeško oko ne bo zaznalo. Izgubna kompresija pa izpušča podatke, ki jih človeško oko ne more zaznati, ter tako onemogoča pridobivanje skritega sporočila nazaj iz slike.

Ena izmed prednosti formata PNG pred drugimi slikovnimi formati je podpora več različnih barvnih tabel. Poleg standardnega RGB (Red Blue Green) zapisa podpira tudi RGBA (Red Green Blue Alpha) zapis, katerega bistvena razlika je dodaten alpha kanal. Le-ta se navadno uporablja za prosojnost slik. Če je vrednost kanala 0% je piksel popolnoma prosojen, če pa je vrednost 100% pa je piksel enak običajnim pikslom.

V primeru steganografije dodaten kanal veliko pripomore, saj lahko v sliko skrijemo kar $\frac{1}{4}$ več informacij, kot v zapisu RGB (Glej poglavje: Steganografija digitalnih fotografij).

4 AES kriptiranje

AES (Advanced Encryption Standard) je eden najbolj uporabljenih standardov za simetrično enkripcijo. Simetrična enkripcija pomeni, da imata pošiljatelj in prejemnik isti ključ, s katerim kriptirata oziroma dekriptirata sporočilo.

Algoritem sta razvila belgijska kriptografa Joan Daemen in Vincent Rijmen, ter ga poimenovala Rijndael. Pozneje je standard z manjšimi spremembami prevzela ameriška vlada kot naslednji standard po uporabi DES (Data Encryption Standard), saj so ključi DES postali prekratki in jih je bilo mogoče razbiti z močnejšimi računalniki.

AES je "substitution-permutation network", kar pomeni, da glavnino operacij, ki jih opravlja predstavljajo različne zamenjave in permutacije bitov.

Poleg višje varnosti je glavna prednost standarda AES hitrost. Operacije, ki jih opravlja so nezahtevne, v nasprotju z asimetričnimi kriptiranji. AES pa je v nasprotju z njimi možno zapisati v vezje v procesorju, tako da lahko današnji običajni namizni računalniki kriptirajo AES tudi z več TB/s.

4.1 Algoritem

Algoritem poenostavljeno sestoji iz štirih korakov, ki se naprej delijo na manjše korake.

1. `KeyExpansions` - razširitve ključa. Ključ, ki je vnaprej določene dolžine se razširi na več ključev, saj AES za vsak krog zahteva svoj ključ.

2. `InitialRound` - dodajanje ključa. Nastavjo se začetna stanja s pomočjo posameznih delov ključa.

3. `Rounds` - del, ki se večkrat ponovi, vsakokrat z novim ključem, ki smo ga razširili iz originalnega ključa. Če je ključ 128-biten se ponovi 10-krat, 12-krat pri 192-bitnih ključih in 14-krat pri 256-bitnih ključih.

   (a) `SubBytes` - preprosta substitucija znakov z uporabo tabele. Pri tem je pomembno, da ime tabela določene lastnosti, ki naredijo to preslikavo nelinearno, kar zelo oteži razbijanje šifre.

   (b) `ShiftRows` - operacija na vrsticah trenutnega stanja. Vsak bit v neki vrstici se ciklično zamakne za neko število. Biti iz konca se premaknejo na začetek.

   (c) `MixColumns` - korak, kjer se stolpci zamenjajo z drugimi stolpci. Vsa stanja v novem stolpcu so neposredno odvisna od vsakega posameznega stanja v prvotnem stolpcu. Če spremenimo en znak, se popolnoma spremeni celoten novi stolpec.

   (d) `AddRoundKey` - korak skoraj enak postopku v InitialRound. Stanju se doda nov ključ, ki ustreza trenutnemu krogu, z operacijo `XOR` (ekskluzivni ali).

4. `Final Round` - še zadnja ponovitev, ki je skoraj enaka vsem ostalim krogom, edina razlika je, da ne vsebuje koraka `MixColums`.[2]

5 Python

Python je sodobno programski jezik, ki je primeren za razvoj najrazličnejših programov, od preprostih skript do numerično zahtevnih simulacij in sodobnih spletnih aplikacij. Zaradi svoje enostavnosti je postal eden najpriljubljenejših programskih jezikov vseh časov, ter je primeren za učenje programiranja. Python je tolmačen jezik, to pomeni, da se sproti med izvajanjem pretvarja v strojno kodo.[1] Zato je razmeroma počasnejši od prevajanih jezikov, kot so C++, Java in C#. V praksi se z uporabo različnih knjižnjic, kot je numpy, njegova hitrost izvajanja lahko približa hitrosti teh jezikov.

Prednost (in hkrati tudi slabosti Pythona) je uporaba dinamičnih tipov, kar pomeni, da imamo lahko v istu spremenljivki ob različnih časih različne podatkovne tipe. Sintaksa Pythona omogoča, da razvijalci prišejo kodo hitreje kot v drugih programskih jezikih, saj potrebujejo manj vrstic kode kot v konkurenčnih programskih jezikih.

5.1 Verzije Pythona

Python trenutno obstaja v dveh glavnih verzijah. To sta 2.x in 3.x (v nadaljevanju tudi Python2 in Python3). Verzija Python3 je novejša, bolj optimizirana, v njej so tudi popravili nekaj pomembnih "napak", ki so bile v Python2. Pomembnejše izboljšave v Python3 so:

- Podpora Unicode znakov. Unicode znaki so lahko vključeni v nizih in tudi v imenih spremenljivk.
- Pravilnejša implementacija nekaterih delov osrednjih jezika - `print` in `exec` nista več stavka (statements), temveč funkciji. Deljenje dveh celih števil vrne racionalno število.
- Optimizacije delovnega spomina - različne funkcije (`range()`, `map()`, `filter()` ...) vrnejo iteracijske objekte, namesto da bi ustvarile celotne sezname.
- Besedi `True` in `False` sta rezervirani in jih programer ne more več po nesreči spreminjati.

Python3 ima veliko prednosti in skoraj nobene slabosti v primerjavi z Python2. Ena od slabosti je, da je za majhna števila malo počasnejši kot Python2, saj ne uporablja tipov `int` ampak tipe `long`, ki zahtevajo več spomina in novejše procesorje.

5.2 Knjižnjice

Python-ova standardna knjižnjica je razmeroma velika, že vsebuje orodja za veliko različnih nalog. Vsebuje že knjižnjice za izdelavo preprostih internetnih aplikacij, grafičnih vmesnikov in tudi knjižnjici za poganjanje testov.

Veliko programov pa ni vključenih v Pythonovo standardno knjižnjico, vendar jih je veliko vključenih v PyPI (Python Package Index), kjer je trenutno (februarja 2017) 99610 različnih paketov. Z različnimi moduli si poenostavimo pisanje programa, saj nam ni treba ponovno implementirati celotnih funkcij, ki bi jih potrevali. Nekateri moduli pa nam omogočijo stvari, ki jih

v "čistem Pythonu" sploh ne moremo napisati, oziroma so napisani deloma v drugem programskem jeziku zato, da se naš program izvaja hitreje.

Pomembnejše knjižnjice za izdelavo programa steganografije slik:

Pillow: Knjžnjica za delo z različnimi formati slik. Povzeta po knjižnjici `PIL` (Python Imaging Library), ki je napisana za Python verzije 2.x, medtem ko je Pillow namenjena za verzije 3.x.

PyCrypto Zbirka varnih hash funkcij in različnih kriptirnih algoritemov (RSA, AES, DES...).

Tkinter Pythonova standardna knjižnjica za izdelavo grafičnega uporabniškega vmesnika (GUI).

6 Izdelava programa

6.1 Priprava okolja

Sam sem pri programiranju uporabljal operacijski sistem Ubuntu, popularno distribucijo Linuxa, ker je po mojem mnenju programiranje na Linuxu veliko lažje in bolj praktično kot na Windowsih. Tudi programski jezik Python je že prednaložen na večini Linux distribucijah, tako verzija 2.x kot 3.x. Program je napisan v Pythonu verzije 3.5.

Najprej naložimo knjižnjici, ki jih potrebujemo:

```
$ pip3 install pillow
$ pip3 install pycrypto
```

7 Zaključek

Literatura

[1] Andrej Brodnik, Luka Fürst, Alenka Krapež in sod. *Računalništvo in Informatika 1*. 2015. URL: `lusy.fri.uni-lj.si/ucbenik/`.

[2] Wikipedija The Free Encyclopedia. *Advanced Encryption Standard*. URL: `en.wikipedia.org/wiki/Advanced_Encryption_Standard` (pridobljeno 2017).

[3] Wikipedija The Free Encyclopedia. *Steganography*. URL: `en.wikipedia.org/wiki/Steganography` (pridobljeno 2017).

[4] Marko Hölbl. “Skrivanje podatkov - steganografija”. V: *Monitor* (2008). URL: www.monitor.si/clanek/skrivanje-podatkov-steganografija.

[5] Mojca Kumerdej. “Doma sem nikjer in hkrati povsod. Moja identiteta je steganografska”. V: *DELO* (feb. 2017).