



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

DEPARTMENT OF COMPUTER SYSTEMS

DIGITÁLNÍ TEXTOVÁ STEGANOGRRAFIE

DIGITAL TEXT STEGANOGRAPHY

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

PETR POUČ

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. JOSEF STRNADEL, Ph.D.

BRNO 2021

Abstrakt

Tato bakalářská práce se zaměřuje na rozbor vědní disciplíny zvaná steganografie, konkrétně digitální textovou steganografií.

Cílem této práce bylo se seznámit s tímto oborem a podrobně prozkoumat metody z oblasti digitální steganografie. Součástí bakalářské práce bylo naimplementovat vhodné metody pro skrývání informací v textových datech.

Metody jež jsem si zvolil a naimplementoval jsem na závěr podrobil řadě automatizovaných testů a podrobné analýze. Provedeným výzkumem jsme zjistili, že.... Vytvořené řešení bylo zpracováno do přehledné tabulky, ve které lze vidět, že úspěšnost některých metod překonala hranici až X procent...TODO

Výsledky této práce umožňují zvolit vhodnou metodu pro nejspolehlivější skrytí informace do textového souboru a zamezit tak nechtěnému dešifrování skryté zprávy.

Abstract

This bachelor's thesis aims to identify the characteristics of a field of science called steganography, especially digital text steganography.

The purpose of this study is to become acquainted with this field and analyze some of the methods of digital text steganography in a detail. Another objective of this bachelor's thesis was to implement suitable methods for information hiding in text data.

The chosen methods that were implemented were finally verified by automated tests and analyzed in a detail. Based on the results of this research, it can be concluded that.... The final statistics were processed into a transparent table, in which we can see, that the success rate of some methods even gets over X percent...TODO

Based on the results of this study, we can choose a suitable method for the most reliable information hiding in the text file, so we can prevent unwanted decryption of a hidden message.

Klíčová slova

Steganografie, skrývání informací, změna textového souboru, bezpečnost, lingvistika, tajná komunikace, důvěrná data

Keywords

Steganography, information hiding, text file modification, security, linguistics, secret communication, confidential data

Citace

POUČ, Petr. *Digitální textová steganografie*. Brno, 2021. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Josef Strnadel, Ph.D.

Digitální textová steganografie

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana doktora Strnadela. Další informace mi poskytl... Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

.....

Petr Pouč
31. března 2022

Poděkování

V této sekci je možno uvést poděkování vedoucímu práce a těm, kteří poskytli odbornou pomoc (externí zadavatel, konzultant apod.).

Obsah

1	Úvod	2
2	Rešerše k tématu bakalářské práce	3
2.1	Základní pojmy	3
2.2	Historie steganografie a její vývoj	4
2.3	Současný stav steganografie	9
2.4	Digitální textová steganografie	10
2.5	Porovnání klíčových vlastností steganografických metod	14
3	Návrh a rozbor řešení	16
3.1	Formát souboru a typ skrývané informace	16
3.2	Realizační prostředky	16
3.3	Aplikace	16
3.4	Návrh řešení	17
4	Implementační část práce	21
4.1	Způsob implementace	21
5	Testování a analýza vybraných metod	28
5.1	Testovací data	28
5.2	Úspěšnost implementovaných metod	28
5.3	Výsledky vlastních metod	32
5.4	Zhodnocení výsledků a porovnání metod	34
5.5	Možné modifikace metod	35
6	Závěr	36
	Literatura	37

Kapitola 1

Úvod

Steganografie byla po dlouhou dobu oblíbeným způsobem utajené komunikace nejen pro armádu, ale i pro politické vůdce. V dnešní moderní době je naprostá většina informací přenášena přes internet a proto jsou možnosti využití steganografie větší než kdy dříve.

Skrývání různých typů informací, kombinování různých metod a nástrojů nám zajišťuje vysokou rozmanitost využití. Nyní máme možnost pomocí steganografie nepozorovaně doručit téměř jakoukoliv informaci. Můžeme schovat do obrázku plán na sestavení letadla, nebo do videa nepozorovaně vložit seznam hesel. Steganografie je také často využívána v oblasti ochrany autorských práv[20], skrytá informace v dokumentu snadno prokáže jeho legitimního vlastníka.

Důvod k ochraně svých údajů je větší než kdy dříve. Technologie i její uživatelé se závratným tempem zdokonalují a nám tedy nezbývá, než se chránit. Růst v používání internetu, chytrých telefonů a jiných moderních zařízení má velký dopad na každodenní život. Tyto technologie jsou velkým přínosem, umožňují nám zpracovávat informace nenákladným a široce dostupným způsobem, avšak tím roste i strach o únik těchto dat. [2]

Naštěstí existuje mnoho steganografických metod které dokáží ukrýt citlivé informace tak, aby se běžný pozorovatel bez speciálních znalostí, neměl šanci k těmto informacím dostat.

Kapitola 2

Rešerše k tématu bakalářské práce

2.1 Základní pojmy

Ještě než stručně shrnu historii steganografie a popíšu odlišné způsoby jejího provedení, bylo by vhodné na začátek objasnit, co tento termín znamená. Steganografie je vědní obor, zabývající se utajenou komunikací. Původem tento pojem pochází z řečtiny a doslovně znamená „skryté psaní“ [9]. Cílem steganografie je nepozorovaně zašifrovat zprávu a bezpečně ji tak dopravit požadovanému adresátovi.

2.1.1 Rozdíl mezi steganografií a kryptografií

Steganografie spadá pod vědní obor nazývaný se kryptografie. Hlavní rozdíl těchto dvou disciplín je v přístupu k ukrytí informace. Steganografie usiluje o ukrytí informace nepozorovaně tak, aby zpráva, obrázek, ani zvukový soubor nebudily žádné podezření. Zatímco kryptografie šifruje zprávu zcela otevřeně a výsledná podoba vypadá jako nesmyslná kupa znaků [9].

	Steganografie	Kryptografie
Cíl	Ukrytí zpráv bez šifrování obsahu	Zašifrovat obsah zprávy
Utajení	Skrytá zpráva není nijak patrná. Text v pozorovateli nebudí žádné podezření	Je okamžitě patrné, že se jedná o zašifrovanou zprávu
Zabezpečení	Záleží na použité metodě	Záleží na diskrétnosti klíče
Útoky	Náročné odhalení/Náročná extrakce	Snadné odhalení/Náročná extrakce

Tabulka 2.1: Srovnání Steganografie a Kryptografie [23]

2.1.2 Vznik steganografie

První použití steganografie je připisováno Řekům. Zdokumentováno bylo antickým historikem Herodotem, podle nějž steganografie zachránila Řecko před Xercem [9].

Demaratus, řecký žijící v Persii se rozhodl varovat Spartany před Xerxesovou invazí do Řecka. Demaratus na dřevěné skládací destičky nanesl vrstvu vosku, pod kterou se skrývala skrytá zpráva [9].

2.2 Historie steganografie a její vývoj

V této kapitole jsem jako primární zdroj používal knihu *Investigator's guide to steganography*[9]. Tento zdroj v této kapitole nadále nebudu uvádět. Uvedené budou pouze citace, které výtažky z této knihy doplňují.

Steganografie má dlouhou historii, její vznik je datován již od starověku 2.1.2, kde byl tento způsob ukrytí zprávy poprvé použit formou dřevěného tabletu a vrstvy vosku. Postupem času se tato věda vyvíjela a zdokonalovala až do dnešní podoby.

V této kapitole vám představím hlavní milníky vývoje steganografie a její nejdůležitější průkopníky, kteří přispěli ke zdokonalení tohoto oboru.

2.2.1 Gaspar Schott

Tento německý vědec vytvořil dílo *Schola Steganographica*, kde poprvé popisuje steganografickou metodu, založenou na přiřazování písmen jednotlivým notám. Píseň by nebyla příjemná k poslechu, avšak nezkušenému člověku by se noty jevily zcela obyčejně.



Obrázek 2.1: Steganografická metoda Gaspara Schotta [9]

2.2.2 Johannes Trithemius

Spisovatel a kryptograf Johannes Trithemius, je jedním ze zakladatelů moderní steganografie. Okolo roku 1500 napsal třídílné dílo *Steganographica*. Tato kniha popisuje rozsáhlý systém k ukrývání tajných zpráv do nevinně vypadajícího textu.

Popsal zde několik nových steganografických metod. Například metodu, která z cover textu¹ bere každé 2. slovo a z něj každé 2. písmeno. Z těchto znaků je následně složena tajná zpráva.

padiel aporsy mesarpon omeuas peludyn malpreaxo

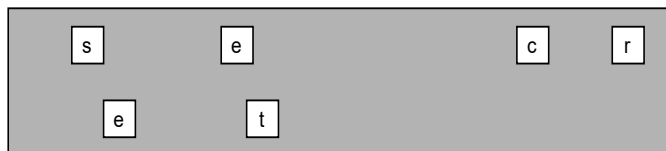
V následující zprávě je ukryta skrytá zpráva *prymus apex*. Další inovativní metodou v knize *Steganographia* byla tzv. šifra *Ave Maria*. Tato metoda obsahuje několik tabulek, které přiřazují každému písmenu nějaké slovo. K napsání tajné zprávy autor textu nahrazuje písmena vybranými slovy, jenž jsou obsaženy v tabulkách.

¹Cover text - text do kterého je vložena tajná zpráva.

2.2.3 Girolamo Cardano

Tento vynikající matematik přišel se způsobem ukrývání zpráv, s kterým jsme dosud neměli možnost se setkat, tzv. grille system. Každý příjemce má papír s několika děrami. Pokud grille² překryjeme přes obvykle vypadající zprávu, díry nám odhalí tajnou zprávu.

Ukázka steganografické metody pro ukrytí tajné zprávy pomocí grille.
Tento způsob ukrytí zprávy vymyslel Girolamo Cardano



Obrázek 2.2: Ukázka vytvoření tajné zprávy pomocí grille [9]

2.2.4 Bishop John Wilkins

Bishop John Wilkins přišel s nápadem pro tajné psaní, pomocí materiálů svítících ve tmě.

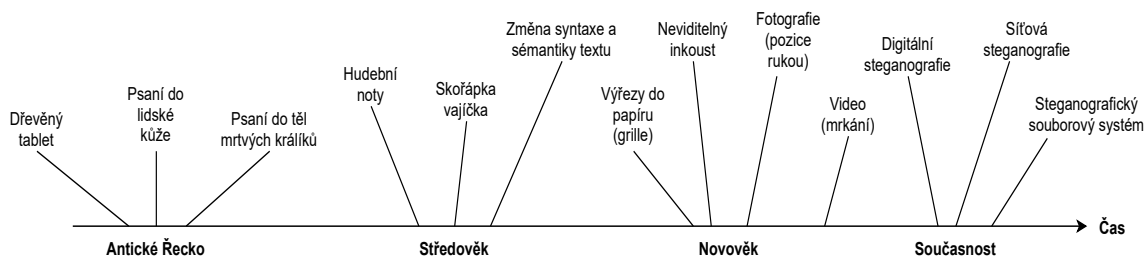
2.2.5 The USS Pueblo

Roku 1968 byla při misi v Severní Koreji zajata americká loď s posádkou. Po několika měsících byla zveřejněna fotka posádky, aby vypadalo že spolupracují s Koreou. Tato fotka však obsahovala tajnou steganografickou zprávu. Pozice rukou členů posádky představovaly tajnou zprávu *snow job*. Tato fráze znamená utajení skutečného motivu ve snaze někoho přesvědčit.

2.2.6 Válka ve Vietnamu

Dalším příkladem využití steganografie je případ amerického velitele Jeremiah Dentona, který byl zajat během války ve Vietnamu. Byl donucen veřejně vystoupit v televizi, aby byl součástí propagandy.

Neschopen cokoliv veřejně zkritizovat, zvolil utajenou konverzaci pomocí mrkání. Během přenosu Jeremiah vymrkal slovo “torture”³



Obrázek 2.3: Časová osa vývoje steganografie [23]

²pomůcka nutná pro rozšifrování zprávy, zpravidla kus papíru, nebo kartónu s děrami.

³z ang. - mučení

2.2.7 Různé techniky steganografie napříč historií

Klasický model pro skrytou komunikaci poprvé navrhl Simmons - Prisoner's problem [16].

Prisoner's problem

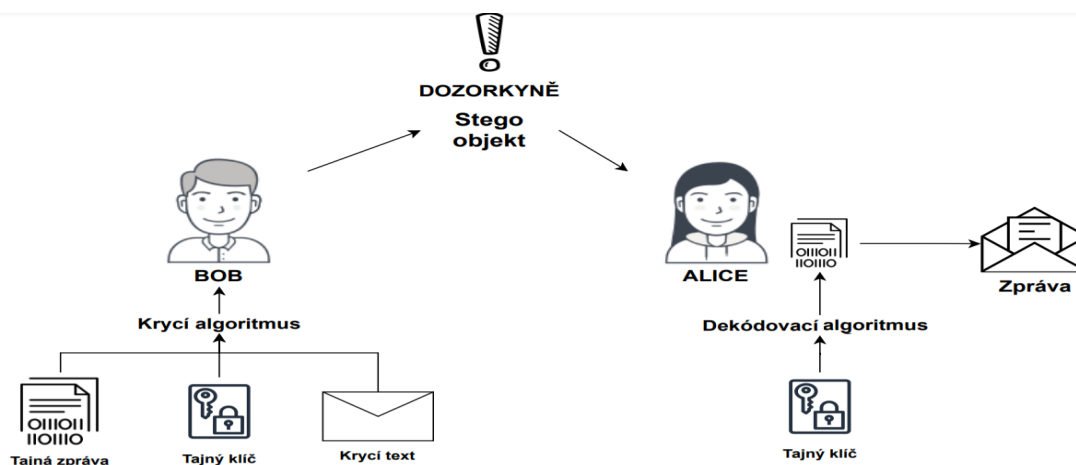
Tento referát byl vytvořen roku 1983 G.J. Simmonsem pro lepší pochopení problematiky tajné komunikace [24].

Dva vězni Bob a Alice jsou uvězněni odděleně, a jediná možnost jejich komunikace je přes dozorkyni. Pokud bude mít dozorkyně podezření na jakoukoliv podezřelou činnost, umístí je na samotku [16].

Jediný způsob tajné komunikace, je nepozorovatelná změna textu, nebo obrázku. Bob proto nakreslí obrázek s modrou krávou a zelenou pastvinou, jenž obsahuje skrytou zprávu. Dozorkyně bez podezření předá obrázek, aniž by si byla vědoma jakékoliv zašifrované zprávy.

Tento způsob tajné komunikace bohužel není stoprocentně spolehlivý a má své slabiny. Pokud by dozorkyně upravila obrázek, mohla by pozměnit, nebo zničit zašifrovanou zprávu. Dalším typem útoku na zašifrovanou komunikaci je tzv. **malicious attack**, v tomto případě by dozorkyně vytvořila vlastní zprávu a předstírala, že je od Boba. Také by dozorkyně mohla zprávu od Boba zapomenout a Alici by jí řekla vlastními slovy, opět by byla tajná zpráva poškozena - tzv. **pasivní útok**.

Tento model popisuje obecný princip steganografie, vždy bude dvě a více tajně komunikujících stran a naslouchající, snažící se utajenou komunikaci zachytit. Proto je potřeba při využívání steganografie pro tajnou komunikaci brát zřetel na potencionální útoky.



Obrázek 2.4: Znázornění obecného principu šifrování zpráv pomocí steganografie.

Mikrotečky Tento nacistický vynález se běžně používal během 2. světové války. Spočívá v pořízení fotografie textu, která je následně zmenšena přibližně na velikost poštovní známky. Následně je reverzním mikroskopem zmenšena na velikost jednoho milimetru (zmenšení až 1:200 [24]. Pomocí jehly se poté obrázek nanese na negativ.

Vernamova šifra

Spolehlivým ukrytím tajné zprávy je tzv. Vernamova šifra (v angličtině One-Time Pads). Každý znak zprávy je posunut o náhodný počet míst abecedy.

Tento systém kódování je považován za nerozluštitelný, za předpokladu, že každý klíč je zcela náhodný a je použit maximálně jedenkrát.

Nulová Šifra

Princip nulové šifry spočívá v ukrytí tajné zprávy do většího množství cover textu na základě jednoduchého vzoru. Nevýhodou této metody je snadná rozluštitelnost, navíc text nemusí vždy znít dobře.

Fishing freshwater bends and saltwater coasts rewards anyone feeling stressed. Resourceful anglers usually find masterful leapers fun and admit swordfish rank and overwhelming any day [9].

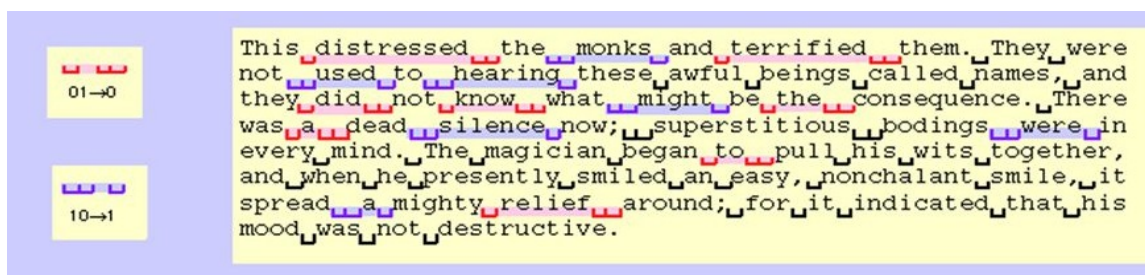
V textu je pomocí nulové šifry schovaná tajná zpráva “Send lawyers guns and money”. V tomto případě je zpráva ukrytá na pozici třetího písmena každého slova, avšak vzor může být libovolný. Které písmena z textu vybrat si komunikující strany musí domluvit předem.

Tato metoda opět našla své uplatnění během 2. světové války[24].

Mezery a Odsazení

Tato forma kódování chytře a nenápadně upravuje původní text ke skrytí taných dat. K zakódování zprávy stačí upravit odsazení řádků, nebo přidat mezeru mezi slova. Tyto slova nacházející se mimo jejich běžnou pozici indikují tajnou zprávu.

Nadbytečné mezery a odřádkování indikují v binární soustavě⁴ 0 a 1.



Obrázek 2.5: Vložení tajné zprávy pomocí odsazení slov.

Rozprostřené Spektrum

Metoda rozprostřeného spektra vezme malý zvukový signál a smíchá ho s širším přenosovým signálem⁵. Tato metoda tajné komunikace je velmi robustní, jelikož používá velké množství krycího signálu, který obklopuje tajnou zprávu.

Techniky Zkreslení

Ukládají informace podle signálu a měří odchylku od původního textu v jednom kroku dekódování [16]

⁴Číselná soustava o základu 2. Využívá pouze číslice 0 a 1.

⁵Šířka pásma zprávy výrazně překračuje koherenční šířku pásma kanálu

Neviditelný Inkoust

Princip této metody spočívá v použití “neviditelné” tekutiny, které jsou viditelné pouze po zahřátí, pod UV světlem, nebo po použití speciálních chemikálií. V průběhu historie se využívalo velké množství tekutin pro tyto účely, například mléko, ocet, citrónová šťáva i lidská moč.

Tento způsob tajné komunikace byl často využíván pro válečné i politické účely napříč celou historií. Tuto metodu používal například i George Washington v 2. polovině 18. století, pro komunikaci se vzdálenými špehy [24].

Používání žargonu

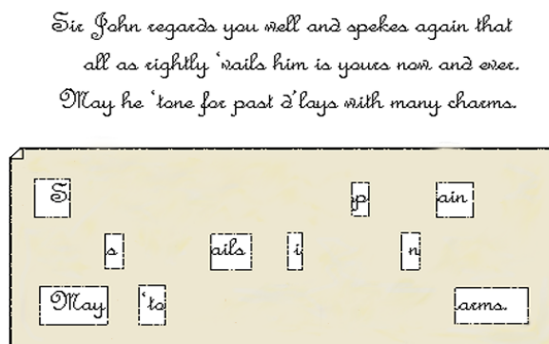
Tato metoda se užívala během 2. světové války pro snadné maskování zpráv. Tato metoda spočívá v nahrazování celých slov.

Během 2. světové války se k FBI⁶ dostal dopis “Polámané anglické panenky budou v nemocnici pro panenky pár měsíců, dokud nebudou opraveny. V nemocnici pro panenky pracují ve dne i v noci”. FBI se podařilo zjistit že panenky reprezentují lodě, každý druh panny jiný typ lodě. Nemocnice pro panenky značila dok [24].

Cardanova mřížka

Tento způsob poslání tajné zprávy se poprvé objevil již ve starověku, kdy se používala deska z pevného materiálu, princip odhalení tajné zprávy je stejný jako u Cardanovi mřížky.

Tato metoda byla převzata vynálezcem Girolamo Cardano, po kterém je pojmenována. Každý s příjemců má kus kartónu s děrami uvnitř, které po překrytí nevinně vypadající zprávy odhalí tajnou zprávu.



Obrázek 2.6: Ukázka odhalení tajné zprávy po přiložení Cardanovy mřížky [22]

Tato metoda se hojně využívala při politické korespondenci během 16. a 17. století [24].

Steganografický souborový systém

Systém si vezme velké místo na disku, zašifruje a schová data tak, že bez hesla nelze dokázat že existují.

Pro správnou funkčnost musí být hesla v hierarchii. Pokud máme 3 soubory uživatel s přístupem k třetímu souboru musí mít také přístup k souborům 1 a 2 [21].

⁶Americký Federální úřad pro vyšetřování. Z ang. Federal Bureau of Investigation

Tento systém nám umožňuje schovávat dokumenty v jiných zdánlivě náhodných souborech. Navíc nám zajišťuje dvou-úrovňovou ochranu. Zašifrované soubory jsou viditelné a stego soubory jsou “neviditelné” a pozorovatel nemůže odhalit soubor, který “neexistuje” [21].

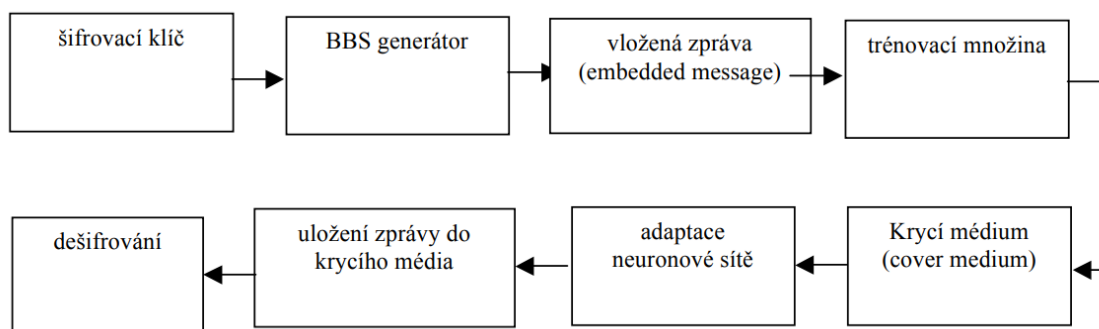
2.3 Současný stav steganografie

V současnosti by se mohlo zdát, že steganografie je přežitek. V dnešní digitální době, plné výkonných počítačů a zařízení by se kryptografie mohla jevit jako lepší a bezpečnější řešení pro přenos zpráv. Pokud ale pro přenos tajné zprávy zvolíme kryptografii, všichni okamžitě vědí, že je jedná o tajnou komunikaci a ta se rázem stane cílem jejich útoku.

To je hlavní a největší výhoda steganografie, umožnit odesílateli nepozorovaně doručit tajnou zprávu. Proto je i dnes steganografie hojně využívána. Nechvalně známým příkladem je používání obrazové steganografie teroristickou organizací Al-Kaidá, která posílá tajné informace svým zaoceánským členům pomocí pornografických obrázků [8].

2.3.1 Steganografie s využitím neuronových sítí

Odhalení tajné zprávy ukrytou steganografickou metodou není snadná záležitost, proto je zpráva dekodována špatně ve více než 20% případů [10]. Z tohoto důvodu je vhodné využít některého z modelů umělé inteligence. V kryptografické oblasti je stále populárnější využití neuronových sítí. Neuronové sítě se dají využít jak k šifrování souborů steganografickými metodami, tak k jejich dešifrování.



Obrázek 2.7: Model steganografického systému využívajícího neuronových sítí [7]

Neuronové jednotky jsou různými modely spojovány do sítí. Každá síť je navržena pro jiné úlohy. Tyto sítě využívají několika různých algoritmů, například Backpropagation, Pruningův algoritmus, gradientní metody a další. [6]

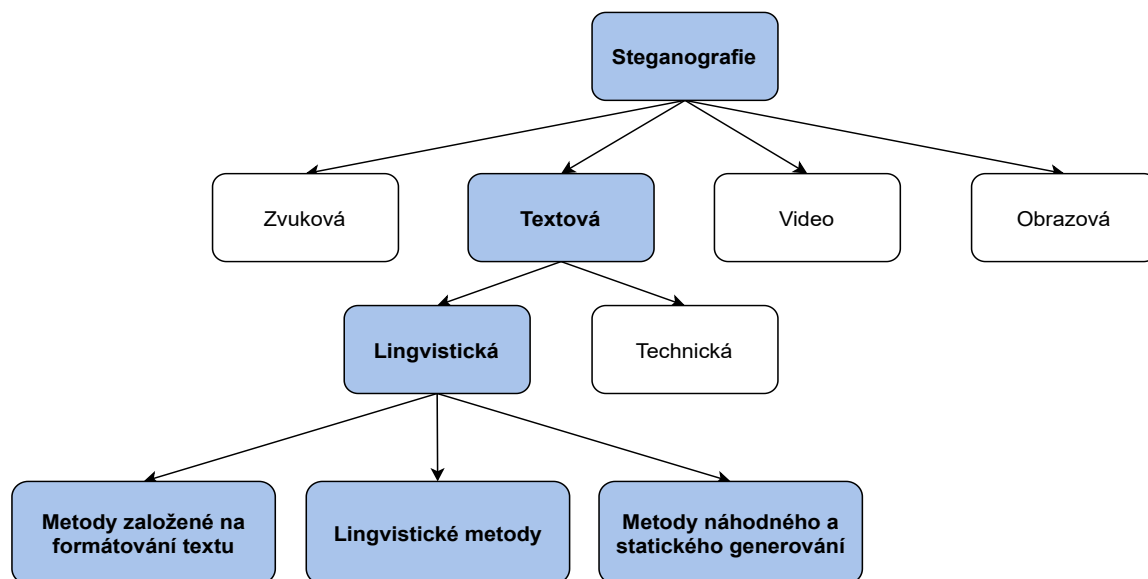
Nejčastěji jsou neuronové sítě využity pro obrazovou a zvukovou steganografii, kde jsou schopné odhalit ukrytou správu téměř ve 100% případů [6][10]. Metodám využívajících neuronových sítí se výrazně zlepšila účinnost z hlediska zachování utajení a kvality zakódované zprávy [4]. V nedávné době se navíc zjistilo, že neuronové sítě lze použít ke kódování více zvuků do jednoho krycího souboru.

2.4 Digitální textová steganografie

Textová steganografie využívá textové soubory jako krytí pro přenos tajných dat. Textová steganografie je považována za nejtěžší druh steganografie. Postrádá totiž velké množství redundantních dat na rozdíl ve srovnání s digitálními médii jako je obraz, zvuk, nebo video [16].

Blíže se budu zabývat textovou steganografií lingvistickou, tj. taková, která k ukrytí zprávy nepotřebuje žádné speciální nástroje, nebo zařízení [9].

Digitální textová steganografie se dále dělí na 3 části: metody založené na formátování textu, lingvistické metody a metody náhodného a statického generování[18].



Obrázek 2.8: Rozdělení steganografie podle prostředí, ve kterém jsou ukrytá data a podle způsobu jejich zašifrování. Zvýrazněná je oblast, kterou se budu podrobně zabývat.

2.4.1 Metody založené na formátování textu

Metody založené na formátování textu libovolně upravují text, za účelem vložení tajné zprávy. Všechny metody založené na formátování vkládají do cover textu binární zprávu ⁷. [9]

Existuje spousta druhů formátování textu pro účely ukrytí zprávy nepozorovaně. Mezi hlavní patří například přidání mezer, nebo odsazení mezi řádky, změna fontu, či velikosti písma. Opět se snažíme, aby byly změny co nejméně nápadné běžnému čtenáři.

Tyto metody fungují nejlépe pro text typu ASCII ⁸.

Open-Space Method

Tato metoda využívá přidávání bílých znaků ⁹ do cover textu.

⁷Hodnota složená pouze z nul a jedniček. Binární hodnotou lze vyjádřit cokoliv.

⁸Americký standart definující znaky anglické abecedy a znaky používané v oblasti informatiky.

⁹Netištěné znaky, tj. mezery

Open-space metoda má několik způsobů provedení. Prvním je vložení bílých znaků mezi věty, tento způsob není efektivní. Aby tajná zpráva nebyla odhalena, můžeme za větu přidat velmi malý počet bílých znaků navíc. pro ukrytí informace proto potřebujeme velké množství krycího textu. [9]

Dalším způsobem jsou mezislovní mezery. Metoda mění počet mezer mezi slovy pro ukrytí binární zprávy. Jedna mezera představuje “0”, dvě mezery představují “1”. [9]

Nejefektivnější variantou Open-Space metody je přidávání mezer na konec řádků. Tato metoda je hůře zpozorovatelná a zároveň nám umožňuje ukrytí většího množství tajných bitů.

Zarovnání textu

Zarovnání cover textu doprava je také používáno k zakódování. Princip opět spočívá v počítání mezer mezi slovy 2.4.1. Tato metoda je velice efektivní a těžká na dekódování. Má ale jednu velkou nevýhodu. Jak poznat nevinnou mezeru od té zakódované, která má značit binární 0? [19]

Řešením je tzv. **kódování Manchester**, které tuto nevýhodu eliminuje. Bity “0” a “1” jdoucí po sobě značí “1”. Bity “1” a “0” reprezentují “1”. [19]

Baconova šifra

Tato metoda vytvořena Sirem Francisem Baconem, využívá 5-ti bitů (BSCII¹⁰) pro kódování 24 písmen abecedy (bez “j” a “u”) od 00000 do 10111. Číslice “0” nahrazuje písmenem “A”, číslice “1” nahrazuje písmenem “B” (takže od AAAAA po BABBB). [5]

Ukázka použití Baconovy šifry.

Princip ukrytí zprávy je následující. Máme 2 fonty písma, 1 značí písmeno “A”, druhý font písmeno “B”. V reálné situaci by si tyto fonty měly být co nejpodobnější, pro lepší demonstraci metody, jsem použil výrazně odlišné.

Z textu lze získat schéma: **BAABA AAAAA ABAAA ABBAA AABAA**

Pomocí následující tabulky převedeme schéma na tajnou zprávu: **tajne**

a	AAAAA	e	AABAA	i,j	ABAAA	n	ABBAA	r	BAAAA	w	BABAA
b	AAAAB	f	AABAB	k	ABAAB	o	ABBAB	s	BAAAB	x	BABAB
c	AAABA	g	AABBA	l	ABABA	p	ABBBB	t	BAABA	y	BABBA
d	AAABB	h	AABBB	m	ABABB	q	ABBBB	u,v	BAABB	z	BABBB

Tabulka 2.2: Tabulka pro Baconovo šifrování

Kódování vlastností

Tato metoda se využívá zejména v MS Word souborech. Metoda ukrývá tajnou zprávu pomocí úprav textu, které jsou pro lidské oko neviditelné. Patří zde například změna měřítka znaků, ohraničení odstavců, ohraničení symbolů a různé způsoby vkládání dat. [3]

¹⁰Baconův standart

2.4.2 Lingvistické metody

Princip metod lingvistické steganografie spočívá ve změně struktury textu. Může se jednat o narušení gramaticky správné syntaxe¹¹, nebo sémantickou záměnu slov. [9]

Sémantické metody

Princip ukrytí informace je podobný jako u všech ostatních metod, cílem je do textu nepozorovaně ukrýt sekvenci bitů, které představují skrývanou informaci. Sémantické metody k tomu používají záměny slov.

Jednou z nejznámějších je sémantická metoda, kterou vynalezl Mohammad Hassan Shirali-Shahreza. Tato metoda využívá záměnu výrazů pomocí synonym¹². Při použití této metody vyžadujeme předem danou tabulku, která bude specifikovat které slovo z dvojice představuje bit “0” a “1”. Sémantická metoda je velmi spolehlivá a text s ukrytou zprávou v žádném případě nevypadá podezřele. [18]

Slovo	Synonymum
Hard	Difficult
Movie	Film
Unhappy	Sad
Cookie	Biscuit

Tabulka 2.3: Příklady synonym

Další podobná metoda zaměňuje výrazy z britské a americké angličtiny [18]. Tuto metodu na rozdíl od předchozí lze využít pouze v anglicky psaných textech. Rodilý mluvčí a člověk zdatný v anglickém může velmi rychle nabýt dojmu, že text není v pořádku.

Britské výrazy	Americké výrazy
Flavour	Flavor
Neighbour	Neighbor
Humour	Humor
Dialogue	Dialog
Colour	Color
Cheque	Check

Tabulka 2.4: Velmi podobné britské a americké slova vhodné k záměně slov

Existuje nespočetně mnoho sémantických metod, založených na podobném principu. Slova můžeme zaměnit za zkratky, smajlíky, protiklady a další.

SMS forma

Tyto metody využívají kombinaci zkrácených slov používaných v SMS. Pro implementaci této metody je potřeba vytvořit slovník, který obsahuje slova a k nim odpovídající SMS zkratky. Zkrácený tvar značí bit “1”, zatímco plný tvar slova značí bit “0”. [1]

¹¹Soubor pravidel definujících správné kombinace slov ve větné stavbě.

¹²Slovo se stejným, nebo podobným významem.

Syntaktické metody

Ke skrytí tajné zprávy do cover textu lze například využít manipulace s interpunkcí. [1] Velkou nevýhodou této metody je, že metoda musí identifikovat vhodná místa pro vložení těchto znaků.

červená, zelená, a modrá
červená, zelená a modrá

Nadbytečné použití interpunkce je velmi nápadné a proto tato metoda není vhodná k ukrytí zprávy.

Jiné syntaktické metody se zaměřují na změnu dikce ¹³ a pořadí slov ve větné stavbě. Tento způsob je mnohem bezpečnější, ale obtížněji proveditelný.

2.4.3 Metody náhodného a statistického generování

Aby se steganografové vyhnuli pracování se známým textem, často si generují vlastní cover text pro ukrytí šifrované zprávy.

Výhoda těchto metod je, že třetí strana nezná cover text, avšak velkou nevýhodou je, že tento vygenerovaný text nemusí vždy působit zcela legitimně.

Jeden z nejlepších příkladů pro náhodného generování steganografického textu je aplikace **spammic**, tento nástroj využívá “spammovou” gramatiku a Waynerův algoritmus pro bezkontextovou gramatiku. Zpráva je tvořena ze spamových frází. Jedná se o přesvědčivou formu generování textu, jelikož spousta nesmyslného spamu je volně dostupná všude na internetu. A skutečný spam je většinou napsaný tak hloupě, že je téměř nemožné rozeznat náhodně vygenerovaný spam od toho skutečného. V nástroji **spammic** se vygenerované fráze liší na základě zprávy, kterou chceme zašifrovat. [17]

Mimické funkce

Mimické metody využívají bezkontextovou gramatiku (CNF) ¹⁴. Generovaný text se snaží napodobovat text reálný, použitím statistiky výskytu slov a písmen, délky slov a četnosti jednotlivých znaků pro vygenerování velmi realistických vět. Při každém kroku vybere metoda jednu možnost ze získaných dat.

Kódování:

Start → podmět, přísudek
podmět → Petr | Pavel
přísudek → poslal | odeslal
předmět → email | sms

Tato tabulka nám umožňuje vytvořit 8 různých vět, přičemž každá věta v sobě ukrývá jinou posloupnost ukrytých bitů. (Petr poslal sms: **011**, Pavel odeslal email: **110** atd.). Je patrné, že počet vygenerovaných možností je dán vztahem 2^N , kde N je počet pravidel v dané tabulce. [17]

¹³Způsob vyjadřování a volba výrazových prostředků

¹⁴Množina pravidel. Pravidla tvoří různé řetězce (věty)

2.5 Porovnání klíčových vlastností steganografických metod

Porovnání jednotlivých metod je vytvořeno na základě existujících publikací, obsahujících šifrování textu pomocí těchto metod. Přesnému vyhodnocení a porovnání těchto steganografických metod se budu podrobně věnovat v části **Testování a analýza vybraných metod 5**.

2.5.1 Postřehnutelnost

Postřehnutelnost je v oblasti steganografie nejdůležitější vlastnost. Celý princip steganografie staví právě na nepostřehnutelnosti. Tzn. že všechny metody usilují o co nejmenší postřehnutelnost.

Postřehnutelnost	
Open-space method	Velmi nízká
Baconova šifra	Vysoká
Metoda synonym	Nízká
Nulová šifra	Vysoká
Mimické funkce	Nízká
Kódování vlastností	Velmi nízká

Tabulka 2.5: Z vybraných metod je nejméně nápadná Open-space metoda

2.5.2 Robustnost

Robustnost, neboli náchylnost na změnu při manipulaci s textem je další faktor při hodnocení steganografických metod. Ke změně struktury textu, která by poškodila ukrytou zprávu může dojít i neúmyslně, proto je robustnost u steganografických metod velmi důležitá vlastnost.

Robustnost	
Open-space method	Velmi nízká
Baconova šifra	Nízká
Metoda synonym	Velmi vysoká
Nulová šifra	Nízká
Mimické funkce	Vysoká
Kódování vlastností	Nízká

Tabulka 2.6: Vyšší robustnost implikuje lepší využití steganografické metody

Robustnost je testována změnou výstupního textu, který již obsahuje skrytou zprávu. Text můžeme modifikovat přidáním/odebíráním slov, změnou fontu, nebo velikosti písmen. Při dekódování nás zajímá, jaký vliv měly změny na podobu ukryté zprávy.

Pokud je steganografická metoda vysoce robustní, modifikace výstupního textu by neměly mít na ukrytou zprávu žádný vliv.

2.5.3 Kapacita

Kapacita metod nám říká, kolik lze maximálně uložit bitů do daného souboru, za použití vybrané metody.

$$\text{Kapacita} = \text{počet ukrytých bajtů} / \text{velikost cover textu [B]}$$

Kapacita	
Open-space method	Velmi nízká
Baconova šifra	Velmi vysoká
Metoda synonym	Velmi nízká
Nulová šifra	Nízká
Mimické funkce	Vysoká
Kódování vlastností	Vysoká

Tabulka 2.7: Tabulka nám říká, která z metod dokáže ukrýt největší množství bitů pro stejně velký cover text

Nejnižší kapacitu z vybraných metod má metoda synonym, ta pro ukrytí tajné zprávy potřebuje velké množství cover textu a rozsáhlé slovníky pro nahrazování slov.

Nízkou kapacitu má také Open-space metoda, ta nemůže ukrývat větší množství bitů, neboť větší množství mezer mezi slovy by bylo opravdu velmi nápadné. Avšak existuje modifikace této metody, která přidává bílé znaky na konec řádku. Tento způsob je, co se týče kapacity metody, mnohem lepší [2.4.1](#).

Kapitola 3

Návrh a rozbor řešení

3.1 Formát souboru a typ skrávané informace

Pro ukrytí zprávy do textového souboru lze využít mnoho textových formátů, například `.txt`, `.docx`, `.doc`, `.adoc` a další. Já se rozhodl pracovat pouze s prvními 2 výše zmíněnými formáty. Výstupní soubor bude vždy formátu `.docx`, který mi umožní použití všech možných steganografických metod (`.txt` například neumožňuje změnu fontu jednotlivých slov).

Každá metoda má navíc vlastní znakovou sadu, kterou je ukrytá informace kódována. Nejčastěji využívaná sada je ASCII. V současnosti se častěji používá rozšířené ASCII, které kóduje znaky do 8 bitů (z původních 7). 8-bitové ASCII dokáže kromě písmen a číslic, zobrazit i velké množství speciálních znaků. Tuto sadu jsem se rozhodl použít u všech svých metod, kromě Baconovy šifry, jenž má své vlastní kódování.

3.2 Realizační prostředky

3.2.1 Implementační jazyk

Jako implementační jazyk jsem zvolil `Python version 3.8.1`. Jedná se o objektově orientovaný jazyk s jednoduchou syntaxí, která zaručuje snadno čitelný kód. Pro tuto BP nabízí mnoho užitečných knihoven, jako je *Matplotlib* pro vykreslování grafů⁵.

Knihovny tohoto jazyka jsou uvnitř psané stejně jako v jazyce C, proto tento jazyk zaručuje i vysokou rychlost programu.

Dalším faktorem pro výběr tohoto jazyka je snadná manipulace s textem, která představuje hlavní náplň této práce.

3.3 Aplikace

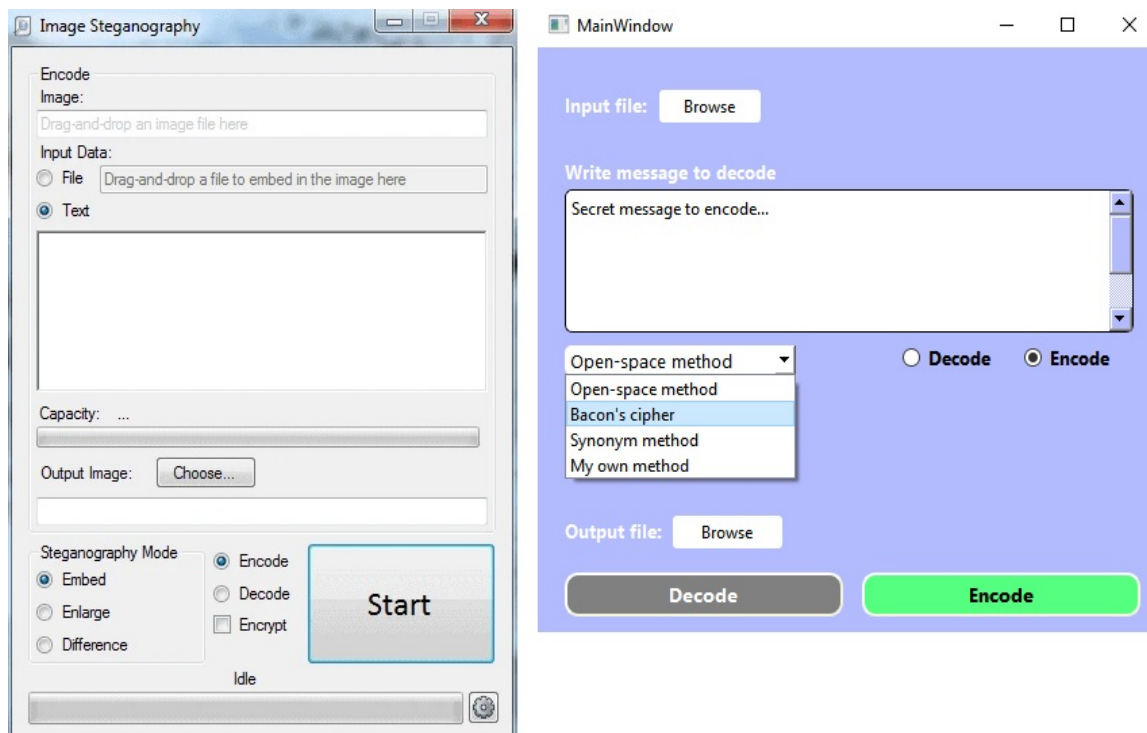
3.3.1 Průzkum existujících aplikací

Jako součást řešení jsem se rozhodl vytvořit aplikaci, pro lepší demonstraci jednotlivých steganografických metod. Při tvorbě návrhů své steganografické aplikace jsem vycházej z již existujících aplikací, jako jsou například *Steghide*, *Hide'N'Send*, nebo *Image Steganography*.

3.3.2 Návrh GUI

Pro tvorbu uživatelského rozhraní jsem si vybral platformu QT, s knihovnou `pyside6`. Tato knihovna je kompatibilní s jazykem Python, ve kterém je psána hlavní část bakalářské práce.

Aplikace umožní uživateli vybrat textový soubor, zvolit si jednu z implementovaných metod a zašifrovat do tohoto souboru zprávu, nebo ji naopak ze souboru dešifrovat.



Obrázek 3.1: Nalevo je aplikace pro obrazovou steganografii, kterou jsem se inspiroval [11]. napravo je prvotní návrh mojí aplikace.

3.4 Návrh řešení

3.4.1 Vybrané metody textové steganografie

V rámci praktické části této bakalářské práce jsem se rozhodl implementovat následující metody z oblasti steganografie. Dvě z nich jsou založené na formátování textu **Open-space method** (varianta s mezerami mezi slovy) 2.4.1 a **Baconova šifra** 2.4.1. Z oblasti lingvistických metod jsem si zvolil **Metodu synonym** 2.4.2. Na závěr se pokusím implementovat vlastní metody vycházející z těchto metod.

Všechny tyto metody pozměňují cover text pro ukrytí bitů, obsahujících tajnou zprávu. Tyto metody jsem záměrně vybral podle odlišného stylu kódování a dekodování tajné zprávy.

Dalším kritériem pro výběr těchto metod byly jejich přednosti. Vybral jsem tedy metody takové, z nichž každá vyniká v jiném aspektu steganografie. Některá metoda dokáže ukrýt mnohem větší množství bitů než jiné, jiná je odolná vůči všem možným změnám textového

souboru. Použití některých metod si všimneme okamžitě, jiné nemáme šanci postřehnout. Porovnání těchto metod se věnuje jiná část práce 5.4.

3.4.2 Vlastní metoda

Součástí této práce je i návrh vlastní steganografické metody. Na poli steganografie je velmi obtížné přijít s nápadem na novou metodu. Existuje enormní množství nejružnějších metod a všechny dobré nápady byly již vyčerpány. Proto svou vlastní metodu vytvořím z již existujících metod. Skloubením jejich principů vytvořím novou metodu, která bude kombinací jejich nejsilnějších stránek.

Baconova šifra + Open-space metoda

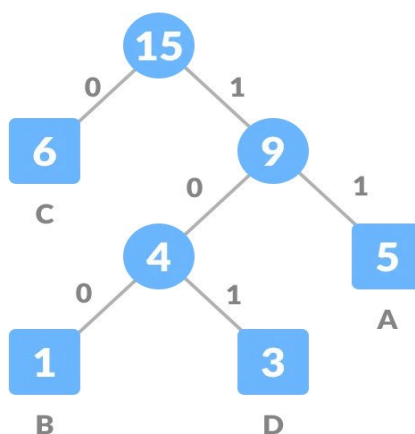
Prvním příkladem je šifrování Baconovou šifrou, u kterého jsou pro nás relevantní pouze řádky s mírně pozměněnou hodnotou odřádkování (Open-space metoda). Vzniklá metoda by tedy kombinovala přednosti obou metod, špatnou postřehnutelnost a dobrou kapacitu pro ukrytí zprávy.

Baconova šifra + Metoda synonym

Mým dalším návrhem je kombinace Metody synonym a Baconovy šifry. Metoda synonym by měla mít ze všech metod nejnížší kapacitu, což kompenzuje vysokou robustností. Proto navrhuji použít způsob šifrování dat, které využívá Baconova šifra k zvýšení kapacity metody synom.

Huffmanovo kódování

Zvýšit kapacitu metod by šlo i dalšími způsoby. Velmi oblíbené je například Huffmanovo kódování, které se běžně používá pro bezztrátovou kompresi dat. Princip šifrování s tímto typem kódování by probíhal následovně. V tajné zprávě kterou si přejeme zašifrovat si zjistíme nejčastější výskyty znaků, nebo posloupností znaků. Následně vytvoříme Huffmanův strom, který častěji používaným znaků, přiřadí kratší bitový vzor.



Obrázek 3.2: Huffmanův strom pro vstupní řetězec bcaaddccacacac [13]

Znak	A	B	C	D
Četnost	5	1	6	3
Huffmanův kód	11	100	0	101

Tabulka 3.1: Znaky byly zapomocí Huffmanova kódování převedeny na kratší sekvence bitů

Výsledný kód 1000111110110110100110110110 vytvořený na základě předchozí tabulky se nazývá prefixový. To znamená, že žádné kódové slovo není prefixem slova jiného.

Časová složitost zakódování znaku na základě jeho četnosti v tajné zprávě je $O(n \log n)$.

3.4.3 Automatizované testy

Existující publikace pro zhodnocení vlastností steganografických metod provádí testování všech klíčových vlastností nad vybraným textovým datasetem ¹. Žádné z těchto benchmarkových dat/datasetů jsem při hledání nenašel. Důvodem může být, že existuje velké množství steganografických metod, z nichž každá má jiný způsob ukrytí informace, bylo by tedy potřeba mnoho různých testovacích sad. Dalším důvodem je, že je velmi snadné vytvořit si vlastní cover text.

Jelikož jsem nenašel žádné konkrétní testovací soubory, vytvořil jsem si vlastní testovací data, na míru svým metodám. Například nebudu testovat metodu synonym na jiných než anglických testech, takovéto testy by negativně zasáhly do výsledků této metody.

Testy důkladně prověří základní steganografické vlastnosti implementovaných metod. V plánu mám testovat **rychlost metod**, **úspěšnost ukrytí**, **kapacitu** a **robustnost**. Na některé důležité vlastnosti, jako například **postřehnutelnost** nelze napsat testy. Vyhodnocení této vlastnosti bude na základě mého osobního dojmu a názoru dotázaných osob.

Rychlost metod

Testy na rychlost nám řeknou, která metoda dokáže nejrychleji zašifrovat všechny poskytnuté soubory. K vyhodnocení této vlastnosti využiju knihovny `timeit`

```

1  #soucasny adresar + presun do slozky s cover soubory
2  path = os.getcwd() + '/cover_files/synonyms'
3  list_of_files = os.listdir(path)
4  #zacatek mereni
5  start = timeit.default_timer()
6  for file in list_of_files:
7      #sifrovani se vzolenymi vstupnimi argumenty
8      steganography.main(['-i', file, '-e', '-s', secret_message, '-r'])
9  #konec mereni
10 end = timeit.default_timer()
11 time = (end - start)

```

Výpis 3.1: Výpočet rychlosti pro metodu synonym

¹kolekce dat

Úspěšnost ukrytí

Úspěšnost steganografické nám udává v procentech, do kolika souborů z celkového počtu se podařilo vložit celou tajnou zprávu.

$$\text{Úspěšnost} = \frac{\text{počet zašifrovaných souborů (obsahující celou zprávu)}}{\text{celkový počet cover souborů}} * 100 \text{ [\%]}$$

Kapacita metod

Kapacita nám udává množství dat, které je možno skrýt do vstupního textu. [3]

```
1 #kompletni text cover souboru
2 full_text = steganography.print_text(file)
3 if method is "synonyms":
4     words_available = synonyms.count_dictionary_words(full_text)/8
5 elif method is "bacon":
6     words_available = len(full_text.split())/5
7 elif method is "spaces":
8     words_available = len(full_text.split())/8
```

Výpis 3.2: Zjištění maximálního počtu slov, který lze ukryt

Pro výpočet kapacity je zásadní vědět, kolik znaků je maximálně metoda schopna ukryt do daného souboru. Kapacitu následně spočítám jako poměr maximálního počtu ukrytých bajtů, ku velikosti souboru (viz 2.5.3).

3.4.4 Robustnost

Robustnost udává míru schopnosti metody zachovat původní ukrytou zprávu. V souboru zašifrovaném vysoce robustní metodou, zůstane skrytá zpráva zachována i po manipulaci s tímto souborem.

Tuto vlastnost budu testovat následovně. Vytvořím si několik formátovacích funkcí, z nichž každá bude jiným způsobem manipulovat s zašifrovaným textem. Měnit budou například velikost písma a jeho font, hodnotu odřádkování, budou přidávat odstavce a různě měnit strukturu souboru.

Zajímat mě bude, která funkce je odolná vůči jakým změnám. Tzn., po kterých formátovacích úpravách zůstane vložená tajná zpráva nepoškozená.

Kapitola 4

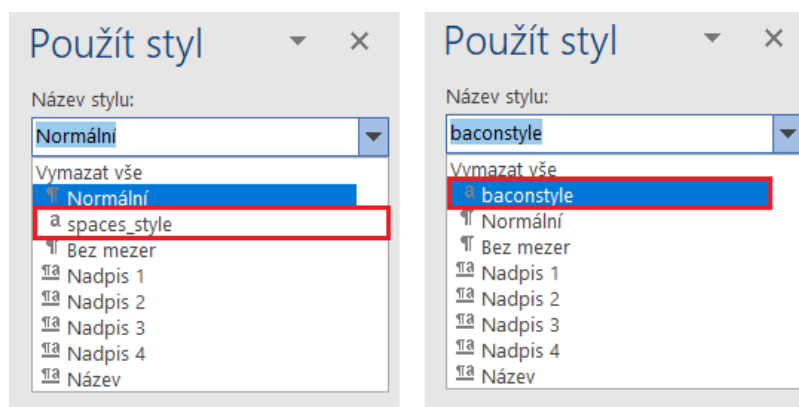
Implementační část práce

4.1 Způsob implementace

Při implementaci zvolených metod jsem využil několika python knihoven, které mi značně usnadnily práci. Mezi ty nejdůležitější knihovny patří `docx`, `XML.etree.ElementTree` a `numpy`.

4.1.1 Vytvoření vlastního stylu

Knihovna `docx` mi nejen umožní vytvořit dokument, ale navíc do něj i nahrát vlastní styl, což je nezbytné při použití metod, které formátují text.



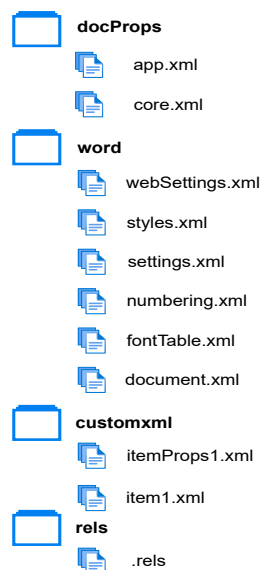
Obrázek 4.1: Vlastní styly, které jsou použity pro vložení tajné zprávy

Styly vytvářím ve funkcích `add_bacon_style` a `add_spaces_style`, přiřadím k nim vlastnosti, jako je velikost textu, font a následně je styl přidám do dokumentu.

4.1.2 Zachování stylu původního dokumentu

Při snaze vložit utajenou zprávu do dokumentu jsem zjistil, že knihovna `docx` nedokáže přenést styl (formátování textu, font, velikost písma, odřádkování atd.) do nově vytvořeného dokumentu.

Tento problém mi pomohla vyřešit již zmíněná XML knihovna. Původní soubor `.docx` jsem rozbalil a dále pracoval pouze se souborem `word/document.XML`, který obsahuje hlavní textové prvky, včetně jejich stylu.



Obrázek 4.2: Struktura dokumentu .docx

Ve funkci `split_document` procházím původní XML soubor element po elementu a kopíruji jej do nového XML souboru. Text kopíruji včetně všech XML tagů, které nám udávají styl textu. Jednotlivé styly se nachází mezi tagy `<w:rPr>` a `</w:rPr>`

```
<w:r w:rsidRPr="00B52930">
  <w:rPr>
    <w:rFonts w:ascii="Agency FB" w:hAnsi="Agency FB" w:cs="Aharoni"/>
    <w:b/>
    <w:bCs/>
    <w:i/>
    <w:iCs/>
    <w:sz w:val="32"/>
    <w:szCs w:val="32"/>
  </w:rPr>
  <w:t>Lorem</w:t>
</w:r>
```

Obrázek 4.3: XML tagy

4.1.3 Implementace - Baconova šifra

Šifrování

Jak jsem již dříve vysvětlil [2.4.1](#), princip Baconovy šifry spočívá v ukrývání bitů do textu, z nichž jsou tvořeny vzory, které značí příslušnou znakovou interpretaci.

Bity jsou touto šifrou nejčastěji ukryty využitím 2 stylů písma. Já se rozhodl k tomu přiřadit navíc změněnou velikost písma. Díky čemuž snížím šance na poškození tajné zprávy. To by mohlo nastat v případě, kdy by původní cover text již obsahoval font, který slouží pro ukrytí bitu “1”.

Při tvorbě vlastního stylu jsem vycházel z faktu, že většina textů je psána fontem **Times New Roman**. Snažil jsem se tedy najít font, který je tomuto vzhledově nejbližší. Experimentováním s různými fonty písma jsem dokázal najít velmi podobný font, který je téměř k nerozeznání od zmíněného **Times New Roman**.

Schoolbook font Times New Roman font

Obrázek 4.4: Styly použité pro Baconovu šifru

Jako font písma, který bude sloužit k ukrytí jedničkových bitů jsem tedy zvolil **Century Schoolbook**, velikosti písma 10. Stejně jako u **Times New Roman** se jedná o takzvaný serifový font ¹.

```
font_charstyle = font_styles.add_style('baconstyle', WD_STYLE_TYPE.CHARACTER)
font_object = font_charstyle.font
font_object.size = Pt(10)
font_object.name = 'Century Schoolbook'
```

Poté co jsem vytvořil vlastní styl jej potřebuji aplikovat na vhodná místa. Uživatel si na vstupu zvolí tajnou zprávu kterou si přeje ukrýt. Zprávu si převedu do binární podoby a dále pracuji pouze s XML.

XML podobu původního cover textu zkopíruji do nového XML souboru a zároveň se dívám na binární podobu zprávy. Ke slovům, které se vyskytují na pozicích jedničkových bitů, vložím do XML vlastní tag `<w:rStyle>`, který značí mnou vytvořený styl. Ještě než jej do xml dokumentu vložím, musím odstranit tagy, které by s mým stylem byly v konfliktu. Jedná se o tagy `<w:rFonts>`, `<w:sz>` a `<w:szC>`.

```
▼<w:r>
  ▼<w:rPr>
    <w:rStyle w:val="baconstyle"/>
  </w:rPr>
  <w:t xml:space="preserve">Ukázka </w:t>
</w:r>
```

Obrázek 4.5: Styly použité pro Baconovu šifru

¹Některé tahy tvoří příčné ukončení. Mají takzvanou patku

Dešifrování

Dešifrování Baconovy šifry probíhá ve funkci `Bacon_decode`. Procházím slova jednotlivých běhů ² a kontroluji jaký mají nastavený styl.

```
1     if run.style.name == "baconstyle":
2         binary += '1';
3     elif run:
4         if run.text != ' ':
5             binary += '0';
```

Výpis 4.1: Dešifrování ukryté zprávy Baconovou šifrou

Poté co jsem prošel celý zašifrovaný text jsem získal binární podobu ukryté zprávy. Tu musím rozdělit do skupinek po 5 bitech (Baconova šifra je 5-ti bitové binární kódování) a každou skupinku převést na odpovídající znak. To probíhá ve funkci `bacon_pattern_to_string`

```
bacons_table = ["00000", "00001", "00010", "00011", "00100", "00101",
                "00110", "00111", "01000", "01001", "01010", "01011", "01100",
                "01101", "01110", "01111", "10000", "10001", "10010", "10011",
                "10100", "10101", "10110", "10111"]

alphabet      = ["A", "B", "C", "D", "E", "F", "G", "H", "(I,J)", "K", "L",
                "M", "N", "O", "P", "Q", "R", "S", "T", "(U/V)", "W", "X",
                "Y", "Z"]
```

```
1     bacons_decoded_message = ""
2     for k in range(len(bacons_patterns)):
3         for l in range(len(bacons_table)):
4             if(bacons_patterns[k] == bacons_table[l]):
5                 bacons_decoded_message += alphabet[l]
```

Výpis 4.2: Převod binární sekvence na čitelnou podobu

`Bacons_patterns` jsou již vytvořené binární skupiny. Jednotlivé skupiny postupně testuji, zdali odpovídá některému ze vzoru z `bacons_table`, následně uložím do zprávy znakovou podobu tohoto vzoru.

²Run je v knihovně `docx` sekvence znaků se stejným formátem, může to být odstavec, zvýrazněné slovo a další

4.1.4 Implementace - Open-space metoda

Open-space metoda 2.4.1 podobně podobně jako Baconova šifra spočívá ve formátování textu. Jako variantu této metody jsem zvolil mezislovní mezery. Tuto metodu jsem ještě upravil podle sebe, aby dosahovala lepších výsledků. Námísto přidání mezery, jen zvětším velikost stávající mezery

Šifrování

Stejně jako u Baconovy šifry i zde pracuji s XML podobou dokumentu. Ve funkci `create_whitespace_el` vytvářím XML element `<w:sz>`, kterému nastavuji hodnotu na 19, tato hodnota odpovídá velikosti mezery 9.5. Velikost mezery jsem záměrně zvolil takovou, aby byla minimální šance na zničení tajné zprávy a zpráva byla zároveň nepostřehnutelná.

Velikost mezery je vždy totožná s velikostí textu, nejčastěji používaná velikost textu je 11. Rozdíl mezi velikostí mezery 9.5 a 11 je minimální. Stejně tak šance, že by uživatel použil velikost textu 9.5, což by znemožnilo použití této metody s použitými parametry, je velmi nepravděpodobná.

Opět procházím XML soubor a jednotlivé elementy, včetně stylu, kopíruji do nového XML souboru. Při šifrování touto metodou nás zajímají pouze mezery. Mezeru v XML struktuře poznáme tak, že je prázdný text element. To znamená, že tagy `<w:t>` a `</w:t>` mezi sebou neobsahují žádný text. Pokud se takový element nachází na pozici odpovídající bitu "1" (Opět na základě binární podoby zprávy, kterou si zvolí uživatel), vložím do něj nový element značící velikost mezery 9.5.

```
▼<w:r>
  ▼<w:rPr>
    <w:rFonts w:ascii="Times New Roman" w:cs="Times New Roman" w:hAnsi="Times New Roman"/>
  </w:rPr>
  <w:t xml:space="preserve"> </w:t>
</w:r>
▼<w:r>

▼<w:r>
  ▼<w:rPr>
    <w:rFonts w:ascii="Times New Roman" w:cs="Times New Roman" w:hAnsi="Times New Roman"/>
    <w:rStyle w:val="spaces_style"/>
    <w:sz w:val="19"/>
  </w:rPr>
  <w:t xml:space="preserve"> </w:t>
</w:r>
```

Obrázek 4.6: První blok kódu reprezentuje mezeru beze stylu, druhý blok reprezentuje mezeru, které jsem přiřadil nový element

Dešifrování

K dešifrování jsem využil funkcí knihovny `docx`. Knihovna je schopná zjistit styl jednotlivých běhů, ale není schopná zjistit jejich velikost. Z toho důvodu jsem do elementů mezer

přidal prázdný styl `spaces_style`, který neobsahuje žádné vlastnosti a slouží tedy pouze jako flag, značící bity “1”.

Procházím jednotlivé slova dokumentu a zjišťuji jejich styl. Mezeru se stylem `spaces_style` (tedy mezeru velikosti 9.5) převedu na bit “1”, ostatní mezery na bit “0”.

```
1 for paragraph in doc.paragraphs:
2     for run in paragraph.runs:
3
4         if run.style.name == "spaces_style":
5             binary = binary + '1';
6         elif run.text == " ":
7             binary = binary + '0';
```

Výpis 4.3: Dešifrování ukryté zprávy Open space metodou

Získám tedy ukrytou zprávu v binární podobě, tu jen snadno převedu na textovou podobu pomocí funkce `binary_to_str`. Kódování pro tuto metodu jsem zvolil rozšířené ASCII, to znamená, že každý znak je nutno kódovat 8 bity.

4.1.5 Implementace - Metoda synonym

Tato metoda 2.4.2 se od předchozích liší ve způsobu ukrytí zprávy, neřadí se mezi metody založené na formátování textu, nýbrž mezi metody lingvistické.

Šifrování

Pro účely ukrytí tajné zprávy v binární podobě, jsem nejdříve potřeboval vytvořit slovníky anglických slov a jejich synonym ³.

```
dictionary_of_zeros = OrderedDict(((“smart”,{0}),(“dry”,{0})...))
dictionary_of_synonyms = OrderedDict(((“wise”,{1}),(“arid”,{1})...))
```

Oba slovníky obsahují okolo 150 slov. Jeden slovník obsahuje často používaná anglická slova, které budu v cover textu hledat a druhý slovník obsahuje synonyma k těmto slovům, kterými je budu nahrazovat.

Začínám procházením cover textu, jakmile narazím na slovo, které se vyskytuje v prvním slovníku, zaměním jej za odpovídající synonymum. Pokud původní cover text obsahuje nějaké slovo ze slovníku synonym, přiřadím mu pomocí XML `skip tag`.

```
1 for word in text:
2     text_el = xml.etree.ElementTree.Element(text_tag)
3     if(word.lower() in dictionary_of_zeros):
4         index = list(dictionary_of_zeros.keys()).index(word.lower())
5
6         #jednickove bity nahradim jejich synonymem
7         if bit == '1':
8             #nalezení prislusneho synonyma ve druhem slovníku, podle indexu
9             syn_word = list(dictionary_of_synonyms.keys())[index]
10        elif(word.lower() in dictionary_of_synonyms):
11            # apply <w:rStyle w:val="skip"/>
12            tag = skip_tag()
13            prop_el.append(tag)
14            text_el.text = syn_word
```

Výpis 4.4: Šifrování metodou synonym

³Slova stejného, nebo podobného významu

Dešifrování

Při dešifrování procházím jednotlivé běhy zašifrovaného textu a dělím je na slova. Pokud se slovo nachází ve slovníku synonym a zároveň jeho XML reprezentace neobsahuje `skip tag`⁴ reprezentuje bit “1”. Pokud se slovo nachází ve druhém slovníku, značí bit “0”.

```
1     for paragraph in doc.paragraphs:
2         text = text + paragraph.text + " "
3
4     for run in paragraph.runs:
5         t = run.text
6         split = t.strip().split(" ")
7
8         for word in split:
9             word = word.lower()
10            if ((word in dictionary_of_synonyms) and
11                (run.style.name != "skip")):
12                binary += "1"
13            elif word in dictionary_of_zeros:
14                binary += "0"
15
16    secret_message = steganography.binary_to_str(binary)
17    return secret_message
```

Výpis 4.5: Dešifrování metodou synonym

⁴značí, že slovo se nacházelo ve slovníku synonym již před šifrováním

Kapitola 5

Testování a analýza vybraných metod

5.1 Testovací data

Pro testování implementovaných metod jsem vytvořil rozsáhlý dataset textových souborů. Některé soubory jsem vytvořil, jiné převzal [15][14][12]. Soubory jsou různorodé, při testování používám více textových formátů, různé fonty písma a formátování.

5.2 Úspěšnost implementovaných metod

Testování jsem prováděl na vlastních benchmarkových testech. Za měřil jsem se na tyto vlastnosti metod: rychlost metody, úspěšnost ukrytí zprávy, maximální velikost zprávy vzhledem k velikosti textu, změna velikosti souboru, robustnost a postřehnutelnost.

5.2.1 Rychlost a efektivita metod

V testu rychlosti překvapivě metoda synonym, která je ze všech implementovaných metod nejrychlejší a tedy i nejefektivnější.

Efektivitu počítám jako čas potřebný k zašifrování 1KB dat.

Průměrná efektivita [sekundy]	Baconova šifra	Metoda synonym	Open-space metoda
Krátká zpráva	0.011632	0.010929	0.016486
Dlouhá zpráva	0.012854	0.010577	0.018910

Tabulka 5.1: Čas potřebný k ukrytí jednoho kilobajtu dat u jednotlivých metod

5.2.2 Úspěšnost ukrytí zprávy

Pokud cover text nemá dostatek slov na ukrytí celé zprávy je ukrytí považováno za neúspěšné.

Při testování nejmenší úspěšnost zaznamenala metoda synonym. Pro ukrytí tajné zprávy touto metodou je zapotřebí obrovské množství anglického textu (nejlépe nějaké knihy, články, texty písní atd.).

Nejlépe v tomto testu skončila Baconova metoda, která k zašifrování informace potřebuje pouhých 5 bitů.

Úspěšnost ukrytí [%]	Baconova šifra	Open-space metoda	Metoda synonym
Krátká zpráva (do 10 znaků)	≈ 100	94,444	68.4211
Dlouhá zpráva (do 50)	70.5882	67.6471	42.8571
Velmi dlouhá zpráva (nad 100)	61.9048	59.5238	14.2857

Tabulka 5.2: Úspěšnost ukrytí tajné zprávy u jednotlivých metod

Z tabulky lze vyčíst, že je metoda synonym téměř nepoužitelná pro ukrytí delších zpráv a lze ji spolehlivě použít jen pro opravdu krátké zprávy. Zbylé metody mají nadpoloviční úspěšnost i u delších zpráv.

5.2.3 SIR (Size Increasing Ratio)

SIR nám udává, o kolik je nově vzniklý zašifrovaný soubor větší, než soubor původní.

$$\text{SIR} = (\text{stego.docx size} - \text{cover.docx size}) / \text{cover.docx size} * 100 \text{ [3]}$$

Average Size Increasing Ratio [%]	
Baconova šifra	19,638
Open-space method	57,07
Metoda synonym	22,722

Tabulka 5.3: Procentuální navýšení velikosti nově vzniklého souboru

Největší nárůst velikosti souboru zaznamenala Open-space metoda, která navyšuje velikost souboru o více než 50% . Pokud bychom tedy chtěli šifrovat soubory o velkém objemu dat, bylo by vhodné zvolit jednu ze zbylých 2 metod.

5.2.4 Kapacita metod

Za předpokladu, že jeden znak zabírá 1 bajt v paměti jsem spočítal procento kapacity, jako poměr kapacity 2.5.3, vynásobený 100.

Metody	Cover text [bytes]	Počet slov (v cover textu)	Maximální počet bitů, který lze ukrýt	Počet znaků, které lze metodou ukrýt	Kapacita [%]
Baconova šifra	46615	18975	18975	4033	8,004
	81768	26700	26700	5300	
Open-space metoda	46615	18975	18975	2520	5,211
	81768	26700	26700	3312	
Metoda synonym	46615	18975	662	82	0,158
	81768	26700	1094	136	

Obrázek 5.1: Výsledná kapacita je spočítána pro všechny testovací soubory, ne pouze pro soubory uvedené v tabulce

5.2.5 Robustnost

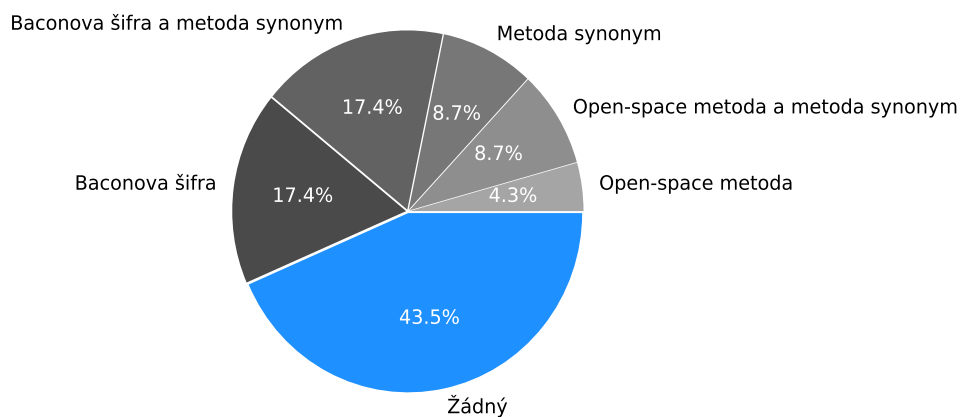
TODO

5.2.6 Postřehnutelnost

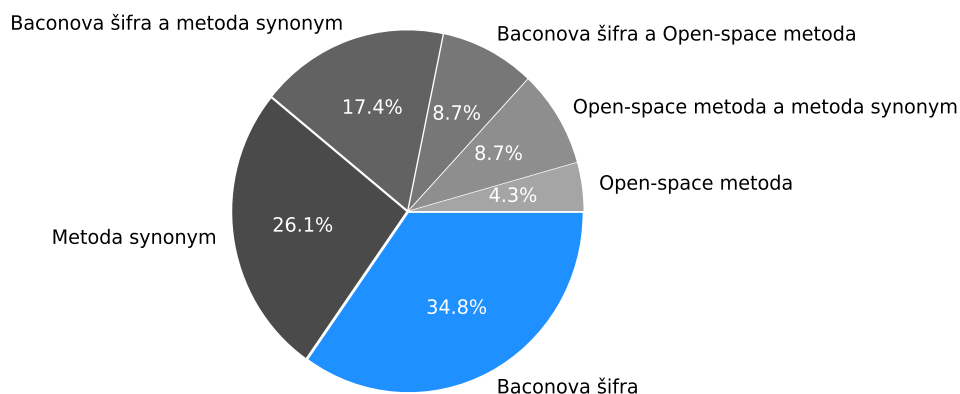
U steganografických metod je nejdůležitější vlastností nepostřehnutelnost. Na tuto vlastnost je velmi obtížné napsat benchmarkové testy. Rozhodl jsem se tedy jednotlivé texty obsahující ukrytou zprávu, zašifrovanou určitou metodou poskytnout většímu množství nezaujatých pozorovatelů.

Zajímalo mě, který text na ně působí neupraveně, a který naopak nejvíce nápadně. Výsledky byly přeneseny do grafu.

Metodami v grafu jsou myšleny texty, zašifrované těmito metodami.



Obrázek 5.2: Výsledky dotazu na otázku: “Přijde vám některý ze souborů podezřelý”



Obrázek 5.3: Výsledky dotazu na otázku: “Dokážete říci, který soubor je zašifrovaný?”

Ukázky zašifrovaných souborů jednotlivými metodami

<p>Lorem ipsum dolor sit amet , consectetur adipiscing elit . Integer vulputate sem a nibh rutrum consequat . Nam quis nulla . Fusce wisi . Etiam ligula pede , sagittis quis , interdum ultricies , scelerisque eu . Duis condimentum augue id magna semper rutrum . Nullam eget nisl . Donec iaculis gravida nulla . Sed convallis magna eu sem. Duis bibendum , lectus ut viverra rhoncus , dolor nunc faucibus libero, eget facilisis enim ipsum id lacus . Sed elit dui, pellentesque a, faucibus vel, interdum nec, diam.</p>	<p>Lorem ipsum dolor sit amet , consectetur adipiscing elit . Integer vulputate sem a nibh rutrum consequat . Nam quis nulla , Fusce wisi . Etiam ligula pede , sagittis quis , interdum ultricies , scelerisque eu . Duis condimentum augue id magna semper rutrum . Nullam eget nisl . Donec iaculis gravida nulla , Sed convallis magna eu sem. Duis bibendum , lectus ut viverra rhoncus , dolor nunc faucibus libero, eget facilisis enim ipsum id lacus . Sed elit dui, pellentesque a, faucibus vel, interdum nec , diam.</p>
---	--

Obrázek 5.4: Text zašifrovaný Baconovou šifrou. Nalevo ideální cover text psaný fontem Times New Roman. Napravo je nepozorované použití metody znemožněno zvýrazněným textem.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vulputate sem a nibh rutrum consequat. Nam quis nulla. Fusce wisi. Etiam ligula pede, sagittis quis, interdum ultricies, scelerisque eu. Duis condimentum augue id magna semper rutrum. Nullam eget nisl. Donec iaculis gravida nulla. Sed convallis magna eu sem.

Obrázek 5.5: Open-space metoda je téměř nepostřehnutelná

THE OBSOLETE MAN AND THE SEA

He was an obsolete man who fished alone in a skiff in the Gulf Stream and he had gone eighty-four days now without taking a fish. In the first forty days a boy had been with him. But after forty days without a fish the boy's parents had told him that the obsolete man was now definitely and finally *salao*, which is the worst form of unlucky, and the boy had gone at their orders in another boat which caught three

Obrázek 5.6: Postřehnutelnost u metody synonym závisí na čtenářově zdatnosti v anglickém jazyce.

5.3 Výsledky vlastních metod

5.3.1 Metoda synonym + Baconova šifra

Další možnou modifikací je navýšení kapacity metod. Open-space metoda i metoda synonym dosahovaly nižších hodnot kapacity, než Baconova šifra. Kapacita by šla navýšit jiným typem kódování. V rámci testů jsem se rozhodl tuto hypotézu ověřit. Vytvořil jsem vlastní metodu, která kombinuje hlavní přednosti metody synonym a Baconovy šifry.

Metodu synonym jsem tedy upravil tak, aby byla informace kódována 5-ti bity, namísto původních 8.

Úspěšnost ukrytí [%]	Metoda synonym (8-bit)	Metoda synonym + Baconovo šifrování (8-bit)
Krátká zpráva (do 10 znaků)	68,4211	78,5714
Dlouhá zpráva (do 50)	42,8571	71,4286
Velmi dlouhá zpráva (nad 100)	14,2857	57,1429
Kapacita	0,158	0,2631

Tabulka 5.4: Úspěšnost ukrytí po použití Baconovy šifry v kombinaci s metodou synonym

Z výsledků lze vidět, že nově vzniklá metoda zaznamenala výrazné zlepšení z hlediska kapacity a tím pádem i úspěšnosti ukrytí tajné zprávy.

5.3.2 Metoda synonym + Huffmanovo kódování

Huffmanovo kódování jsem nejprve testoval na vstupních zprávách, u kterých byl výžkyt jednotlivých znaků zcela náhodný.

Počet batů potřebných k ukrytí tajné zprávy			
Tajná zpráva	Metoda synonym	MS + BŠ	MS + Huffman
zprava1 6[B]	48	30	14
zprava2 20[B]	160	100	86
zprava3 50[B]	400	250	218
zprava4 [200B]	1600	1000	891

Tabulka 5.5: V tabulce 1B dat představuje 1 libovolný znak

Z tabulky je patrné, že šifrování dat Huffmanovým kódováním v porovnání s původní metodou synonym (využívá rozšířené ascii 8-bit) a Baconovým šifrováním (5-bit) potřebuje k ukrytí tajné zprávy nejmenší množství cover textu.

Tohle šifrováním nám tedy zaručí nejlepší možnou kapacitu ze všech zkoumaných způsobů ukrytí zprávy.

Pokud by vstupní zpráva navíc obsahovala pouze omezený počet znaků a jejich výskyt by byl tedy procentuálně vyšší, dosáhla by metoda ještě lepších výsledků. Viz následující tabulka.

Zpráva 30[B]	“Huffman coding incoming uh oh”
Metoda synonym + Baconovo šifrování	K ukrytí je potřeba 150 bajtů cover textu
Metoda synonym + Huffmanovo kódování	K ukrytí je potřeba pouhých 100 bajtů cover textu

Tabulka 5.6: Pokud je zpráva omezena na menší počet znaků, lze očekávat výrazné zlepšení z hlediska kapacity

Znak	"	n	h	o	i	u	f	m	c	g	a	d
Četnost	5	4	3	3	3	2	2	2	2	2	1	1
Vzor	100	101	010	001	000	0111	0110	1101	1100	1111	11101	11100

Tabulka 5.7: Vygenerované vzory [13]. Znaku s častějším výskytem je přiřazen kratší vzor.

Hlavní nevýhodou této metody je, že je potřeba znát předem odpovídající Huffmanův strom. Pokud není předem znám, je potřeba jej vložit do cover textu spolu s tajnou zprávou.

5.4 Zhodnocení výsledků a porovnání metod

V této kapitole se budu věnovat celkovému srovnání všech implementovaných metod.

Jejich steganografické vlastnosti hodnotím na škále 1 (nejlepší) - 3 (nejhorší).

Metoda	Úspěšnost	Kapacita	Robustnost	Postřehnutelnost
Baconova šifra	1	1	3	2-3
Open-space	2	2	2	1
Metoda synonym	3	3	1	2-3

Tabulka 5.8: Srovnání implementovaných metod

Výsledky testování potvrdily téměř všechny vlastnosti metod, které jsem na začátku odhadoval z existujících publikací 2.5. Některé se neshodují zcela, například postřehnutelnost Baconovy šifry dle mých výsledků nevyšla jako nejlepší, jak slibovaly publikace, ale řadí se téměř nakonec žebříčku.

Na následující tabulce demonstрую výsledky mých upravených metod. V tomto srovnání vyšla původní metoda jako nejhorší, ve všech možných aspektech.

Metoda	Úspěšnost	Kapacita	Robustnost	Postřehnutelnost
Metoda synonym	3	3	stejná	3
Metoda synonym + BŠ	2	2	stejná	2
Metoda synonym + HK	1	1	stejná	1

Tabulka 5.9: Srovnání steganografické metody z pohledu šifrování

Postřehnutelnost by měla zůstat stejná, neboť se nemění způsob ukrytí informace. Teoreticky ale zaměňujeme menší počet slov, a tak se šance, že čtenář nabyde podezření, alespoň trochu snižuje.

V poslední srovnávací tabulce můžete vidět, všechny plusy a mínusy jednotlivých metod. Obecně nelze říci, která metoda je nejlepší, každá si dokáže ve steganografickém světě najít své místo.

Metody	Výhody	Nevýhody
Baconova šifra	Vysoká kapacita a úspěšnost ukrytí.	Není odolná vůči žádným formátovacím změnám.
Open-space	Vysoká neviditelnost. Dobrá kapacita i úspěšnost ukrytí.	Není odolná vůči všem formátovacím změnám.
Metoda synonym	Pokud čtenář nevyniká v anglickém jazyce, je metoda špatně postřehnutelná. Nejvíce odolná vůči jakýmkoliv formátovacím změnám	Vysoká složitost, z důvodu použití rozsáhlých slovníků. Může docházet ke změně významu původního dokumentu.
Metoda synonym Baconova šifra	Zachovává robustnost jakou poskytuje metoda synonym. Lehce navyšuje kapacitu	Stejně jako u metody synonym.
Metoda synonym Huffmanovo kódování	Zachovává vysokou robustnost. Výrazně zvyšuje kapacitu a celkovou úspěšnost ukrytí.	Je zapotřebí předem znát konkrétní Huffmanův strom, nebo jej poslat spoluse skrytou zprávou.

Tabulka 5.10: Srovnání hlavních předností metod a jejich nedostatků

5.5 Možné modifikace metod

Žádná metoda není dokonalá, každá má své silné i slabé stránky. S již implementovanými metodami by šlo samozřejmě manipulovat, různě je mezi sebou kombinovat a snažit se tak docílit ještě lepších výsledků.

V Baconově šifře by bylo ideální měnit font podle fontu okolo šifrovaného slova. Bylo tedy potřeba ke každému fontu najít font alternativní (ten nejpodobnější) a nahrazovat jej tímto. To by zajistilo, že bez ohledu na font původního cover textu by byl nově vzniklý text méně nápadný.

Text šifrovaný metodou synonym jak jsme již zjistili je sice nejvíce robustní, ale pro čtenáře s dobrou úrovní angličtiny i dosti nápadný. V anglickém jazyce existuje vždy více slov stejného významu, avšak jejich použití se mění dle kontextu. Příkladem je slovo **silný** – **strong**, můžeme jej nahradit slovy jako **firm**, **powerful**, **potent** a jinými. **Powerful** lze například použít u osob a lidských vlastností (**powerful voice** = silný hlas), **potent** lze použít, pokud chceme říci, že má na nás něco silný efekt (**potent drink** = silný drink) a **firm** se nejčastěji pojí s materiálem. Pokud bychom tedy v textu zvolili špatnou kombinaci těchto slov (**potent voice**, **firm drink**...), text by působil zcela nesmyslně.

Taktéž existují anglická slova, která mají jiný význam jako podstatné jméno a jiný jako přídavné jméno (**play** = hra/hrát, **workout** = trénink/cvičit)

Opět by tedy bylo ideální nahrazovat slova na základě kontextu, přídavná jména nahradit pouze za přídavná a podstatná pouze za podstatná. Dále by bylo potřebné zjistit si slovo následující, zařadit jej do určité skupiny slov a na základě této skupiny správně nahradit slovo s odpovídajícím kontextem.

Kapitola 6

Závěr

Shrnutí výsledků práce. Bylo dosaženo stanovených cílů? Návrh na rozšíření/pokračování práce.

Nejenže tato úprava zrychlila metodu a snížila SIR, ale především výrazně zvýšila kapacitu této metody a tím pádem i celkovou úspěšnost ukrytí zpráv.

Literatura

- [1] AGARWAL, M. *Text Steganographic approaches: A Comparison* [online]. Department of Computer Science and Engineering, PDPM-IIITDM, Jabalpur, India, 2013 [cit. 2021-21-12]. Dostupné z: <https://arxiv.org/ftp/arxiv/papers/1302/1302.2718.pdf>.
- [2] AHVANOOEY, T., MILAD, LI, QIANMU, HOU et al. AITSteg: An Innovative Text Steganography Technique for Hidden Transmission of Text Message via Social Media. *IEEE Access*. 2018, s. 65981–65995. Dostupné z: https://www.researchgate.net/publication/327034906_AITSteg_An_Innovative_Text-Steganography_Technique_for_Hidden_Transmission_of_Text_Message_via_Social_Media.
- [3] BAAWI, S., AL NASRAWI, D. a ABDULAMEER, L. Improvement of "Text Steganography Based on Unicode of Characters in Multilingual" by Custom Font with Special Properties Improvement of "Text Steganography Based on Unicode of Characters in Multilingual" by Custom Font with Special Properties. *IOP Conference Series Materials Science and Engineering*. Červenec 2020, sv. 870. DOI: 10.1088/1757-899X/870/1/012125.
- [4] DA, A., WAHI, J. S., ANAND, M. a RANA, Y. *Multi-Image Steganography Using Deep Neural Networks*. 2021 [cit. 2022-24-03]. Dostupné z: <https://arxiv.org/pdf/2101.00350.pdf>.
- [5] GATHEN, J. von zur. Chapter E Steganography. In: *CryptoSchool*. Springer Berlin Heidelberg, 2015, s. 418. ISBN 978-3-662-48425-8. Dostupné z: https://doi.org/10.1007/978-3-662-48425-8_14.
- [6] HOLOŠKA, J. *Steganografie Jako Hrozba Úniku Kritických Obchodních Informací A Její Detekce Pomocí Umělých Neuronových Sítí* [online]. Trilobit, 2009 [cit. 2022-30-03]. Dostupné z: <http://trilobit.fai.utb.cz/steganografie-a-jeji-detekce-pomoci-umelych-neuronovych-siti>.
- [7] JANUŠEK, R. *Steganografie s využitím neuronových sítí* [online]. Ostravská univerzita v Ostravě, 2015 [cit. 2022-24-03]. Dostupné z: <https://konference.osu.cz/svk/sbornik2015/pdf/budoucnost/informatika/Jarusek.pdf>.
- [8] JURÁŇ, M. *Steganografie*. 2008 [cit. 2021-12-30]. 71 s. Diplomová práce. Mendelova zemědělská a lesnická univerzita v Brně. Vedoucí práce FOLTÝNEK, T. Dostupné z: <https://adoc.pub/steganografie-mendelova-zemdlaska-a-lesnicka-univerzita-v-brn.html>.

- [9] KIPPER, G. *Investigator's guide to steganography*. 1. vyd. Auerbach publications, 2004. 209 s. ISBN 0-8493-2433-5.
- [10] KISHORE, V., CHEN, X., WANG, Y., LI, B. a WEINBERGER, K. Q. Fixed Neural Network Steganography: Train the images, not the network. In: *International Conference on Learning Representations*. 2022. Dostupné z: <https://openreview.net/forum?id=hcMvApXGSzZ>.
- [11] KOLLA, A. *List of 10 Best Steganography Tools to Hide Data* [online]. Geek Dashboard, 2020 [cit. 2021-21-12]. Dostupné z: <https://www.geekdashboard.com/best-steganography-tools/>.
- [12] LIT, W. V. *Email Spam* [online]. kaggle, 2019 [cit. 2022-30-03]. Dostupné z: <https://www.kaggle.com/datasets/veleon/ham-and-spam-dataset>.
- [13] LTD., P. L. P. *Huffman Coding* [online]. Programiz, no date [cit. 2022-30-03]. Dostupné z: <https://www.programiz.com/dsa/huffman-coding>.
- [14] MOONEY, P. *Song Lyrics* [online]. kaggle, 2018 [cit. 2022-30-03]. Dostupné z: <https://www.kaggle.com/datasets/paultimothymooney/poetry>.
- [15] MOSSER, S. *Steganography-dataset* [online]. Github, 2015 [cit. 2021-21-12]. Dostupné z: <https://github.com/mosser/steganography-dataset/tree/master/txt>.
- [16] RAFAT, K. F. a SHER, M. *Survey report - state of the art in digital steganography focusing ascii text documents* [online]. Department of Computer Science International Islamic University Islamabad in Pakistan, 2010 [cit. 2021-28-09]. Dostupné z: <https://arxiv.org/ftp/arxiv/papers/1003/1003.1470.pdf>.
- [17] SALOMON, D. Data Hiding in Text. In: *Data Privacy and Security*. Springer Science Business Media, 2012, s. 465. ISBN 978-1-4419-1816-1. Dostupné z: https://link-springer-com.ezproxy.lib.vutbr.cz/chapter/10.1007%2F978-0-387-21707-9_11.
- [18] SHARMA, S., GUPTA, A., TRIVEDI, M. C. a YADAV, V. K. *Analysis of Different Text Steganography Techniques: A Survey* [online]. IEEE, 2016 [cit. 2021-6-10]. Dostupné z: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7546588>.
- [19] THAMPI, S. M. *Information Hiding Techniques: A Tutorial Review*. 2008 [cit. 2021-26-10]. Dostupné z: <https://arxiv.org/ftp/arxiv/papers/0802/0802.3746.pdf>.
- [20] TOFTEGAARD, A. *A System for Hiding Steganography in Plain Sight* [online]. Kongens Lyngby, 2016 [cit. 2021-13-10]. Dostupné z: <http://www2.imm.dtu.dk/pubdb/edoc/imm7049.pdf>.
- [21] WAYNER, P. *Disappearing cryphography: Information hiding: steganography and watermarking*. 2. vyd. Morgan Kaufmann Publishers, 2002. 409 s. ISBN 1-55860-769-2.
- [22] WIKIPEDIA, THE FREE ENCYCLOPEDIA. *CardanGrille.png*. 2006. Dostupné z: <https://cs.wikipedia.org/wiki/Soubor:CardanGrille.png>.

- [23] ZIELINSKA, E., MAZURCZYK, W. a SZCZYPIORSKI, K. *Development Trends in Steganography* [online]. researchgate, 2012 [cit. 2021-11-10]. Dostupné z: https://www.researchgate.net/publication/258725207_Development_Trends_in_Steganography.
- [24] ŽILKA, R. *Steganografie a stegoanalýza [online]*. 2008 [cit. 2021-10-21]. 66 s. Diplomová práce. Masarykova univerzita, Fakulta informatiky, Brno. Vedoucí práce ŘÍHA, Z. Dostupné z: <https://is.muni.cz/th/p3l0w/>.