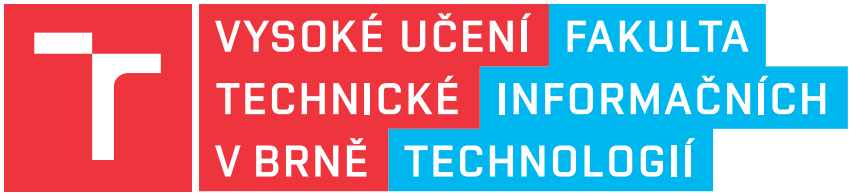


Implementace a prolomení Afinitní šifry

KRY - Kryptografie

8. března 2023

Bc. Jakub Komárek (xkomar33)

1 Úvod

Tento projekt se věnuje implementaci a prolomení substituční šifry, konkrétně Afinitní šifry, která se používá k šifrování textových zpráv. Její základní princip spočívá v lineární transformaci jednotlivých písmen v textu. Tento program umí šifrovat, dešifrovat a prolamovat tuto šifru. Afinitní šifra se skládá ze dvou klíčů "a"a "b".

2 Implementace

Zdrojový kód je členěn na soubory podle jejich funkce následujícím způsobem:

- main.cpp/hpp - Hlavní soubor, který parsuje vstupní argumenty, kontroluje korektnost konfigurace a následně volá jednotlivé moduly programu.
- decrypt.cpp/hpp - Dešifrovací modul
- encrypt.cpp/hpp - Šifrovací modul
- break.cpp/hpp - Modul na prolamování šifry
- myLib.cpp/hpp - Obecné, pomocné funkce programu
- config.hpp - Struktura konfigurace
- externalCode.cpp/hpp - Převzaté funkce a konstrukty.

2.1 Šifrování/dešifrování

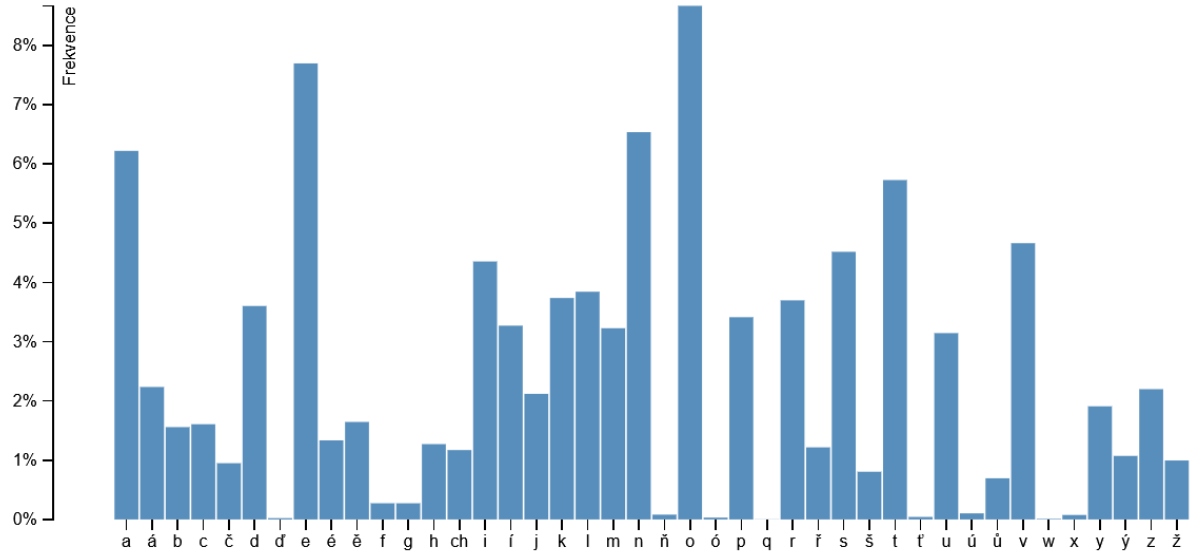
Šifrování/dešifrování probíhá velmi podobně. Šifrování probíhá znak po znaku. V prvním kroku se znak (a-Z) převede do normalizované formy (interval 0-25). Poté je znak transformován pomocí odpovídající funkce. Takto substituovaný znak je následně zpětně denormalizován a přiveden na výstup. Znaky, které nelze šifrovat (mezery, čísla, atd.), se nešifrují a jsou ponechány v nezměněné podobě.

2.2 Prolamování šifry

Při prolamování zašifrovaného textu bez znalosti klíče, jsem využil dvou statistických znalostí textu. První znalostí byla relativní četnost jednotlivých znaků v průměrném českém textu. Druhou znalostí byl seznam 40 nejčastějších bigramů (dvojic písmen) vyskytujících se v českém textu.

Vzhledem k malému počtu substituční funkcí 328 (klíč A může nabýt 13 hodnot a klíč B 26 hodnot - 13*26=328) jsem se rozhodl prozkoumávat všechny kombinace klíčů. Každou kombinaci klíčů jsem algoritmicky ohodnotil metrikou. Za výsledný klíč byl prohlášen takový pár, který měl nejmenší metriku.

Výpočet první části metriky proběhl následovně: V zašifrovaném textu jsem spočetl četnosti všech znaků. Poté jsem každý znak abecedy zašifroval zkoumaným párem klíčů. Provedl jsem rozdíl četnosti tohoto znaku v luštěném textu s průměrnou četností v českém jazyce (viz. obr. 1). Tento rozdíl jsem následně umocnil dvěma a přičetl k metrice. Myšlenka spočívala, že pokud se jednalo o správnou substituci, byl rozdíl četností velmi malý a proto se metrika moc nezvýšila. Pokud rozdíl četností byl velký, metrika se podstatně zvýšila. Umocnění na druhou mocninou ještě více zvýraznilo rozdíly četností.



Obrázek 1: Graf četnosti písmen v Českém jazyce (převzato z [2]).

Tato první metrika se ukázala býti dobrou, hledaný klíč se obvykle nacházel mezi top 10 klíči s nejlepší metrikou. Zhruba u poloviny případů byl nejlépe ohodnocený klíč hledaným klíčem.

Druhou část metriky dvořila analýza bigramů. V zašifrovaném textu jsem spočítal všechny jednotlivé bigramy. Poté jsem z této analýzy vyňal 50 nejčastějších bigramů. Následně jsem šifroval nejčastějších 40 bigramů v českém jazyce (viz. obr. 2) právě zkoumaným párem klíčů, takto zašifrované bigramy jsem se pokoušel nalézt v těchto 50 nejčastějších bigramech. V případě, že bigram nebyl nalezen, byla páru klíčů udělena penalta. V opačné případě metrika nebyla navýšena.

st 74285 en 50645 le 38926 to 36355 ho 31442 al 29682 př 27885 em 26818
35191 do 30665 ed 29622 at 27603 in 26427 ní 60525 na 46737 ko 38688 ou
3 ne 38671 no 32612 os 30530 an 29326 ře 27181 sk 26085 po 56239 je 4243
7168 lo 25981 ov 53818 pr 42099 od 38393 la 32336 se 30454 ce 28280 er 2
ro 51961 te 40393 ra 37531 li 31952 ta 30177 va 27987 ti 26858 ně 25739

Obrázek 2: Nejčastější bigramy v českém jazyce (převzato z [1]).

Obě zmíněné metriky byly použity s rovnou váhou (1:1). Byť je tento přístup poměrně "těžkotonážní", v rámci testování se k mému překvapení ukázal jako extrémně spolehlivý, odolný proti menším abnormalitám v luštěném textu a funkční i s malým vzorkem zkoumaného textu. I když program zkoumá všechny varianty klíčů, při testování jsem nezaznamenal výkonnostní problémy.

Program navíc vypíše na chybový výstup dalších 10 klíčů s největšími metrikami, pro případ, že byla první shoda chybná.

3 Zhodnocení

Podle mého zhodnocení byl projekt úspěšně vypracován v plném rozsahu, nedostatků jsi nejsem vědom.

Odkazy

[1] Centrum zpracování přirozeného jazyka. *Frekvence písmen, bigramů, trigramů, délka slov*. URL: https://nlp.fi.muni.cz/cs/FrekvenceSlovLemmat. (navštíveno: 8.3.2023).

[2] Jan Neckář. *Četnost znaků v českém textu*. URL: https://algoritmy.net/article/40/Cetnost-znaku-CJ. (navštíveno: 8.3.2023).