Vigenèrjeva šifra

Alenka Bahovec, *Škofijska klasična gimnazija, Ljubljana*
Rok Kaufman, *Gimnazija Vič, Ljubljana*
Vid Kocijan, *Gimnazija Vič, Ljubljana*
Matej Roškarič, *Fakulteta za naravoslovje in matematiko, Maribor*

POVZETEK

Vigenèrjeva šifra je postopek šifriranja, ki je ostal nerazbit skoraj štiristo let. Predstavili bomo njene lastnosti, zgodovino uporabe in računalniški program, ki jo razbije.

Osnovni pojmi kriptografije

Kriptografija *(iz gr. kryptós - skrit in gráphein - pisati)* je veda o komunikaciji v prisotnosti aktivnega napadalca. Skozi zgodovino se je razvilo veliko različnih postopkov šifriranja, ki običajno sporočilo (čistopis) spremenijo v šifrirano besedilo (tajnopis) z uporabo skritega parametra (ključa). Za obratni postopek (dešifriranje) moramo vedeti, katera šifra in ključ sta bila uporabljena. Postopke, ki bi omogočili razbitje šifre brez poznavanja ključa, preučuje kriptoanaliza.

Zgodovina uporabe Vigenèrjeve šifre

Vigenèrjeva šifra je poliabecedna zamenjalna šifra, kar pomeni, da se znaki šifrirajo po principu zamenjalne šifre z več tajnimi abecedami. Je preprosta za uporabo, a jo je težko streti, zato si je prislužila naziv *le chiffre indéchiffrable (iz fr. nezlomljiva šifra)*.

Prvi jo je opisal Giovan Battista Bellaso leta 1553 v knjigi La cifra del. Sig. Giovan Battista Bellaso. Zmotno se imenuje po Blaisu de Vigenèrju, ki je leta 1586 predstavil podobno šifro na dvoru francoskega kralja Henrika III.

Slika 1: Blaise de Vigenère

Čeprav so se pojavili posamezni primeri uspešnega razbitja šifre že v 16. stoletju, je več stoletij veljala za nezlomljivo, saj zanesljiv postopek za razbijanje ni bil znan. Tudi Charles Babagge je leta 1854 zlomil različico šifre, vendar ni objavil svojega dela. Med ameriško državljansko vojno je šifro uporabljala Konfederacija, ker pa je uporabljala samo tri različne ključe, je njihova sporočila Unija redno razbijala. Leta 1863 je Friedrich W. Kasiski odkril postopek izračuna dolžine ključa, dokončno pa je leta 1920 šifro zlomil William F. Friedman z indeksom sovpadanja.

Postopek šifriranja in dešifriranja Vigenèrjeve šifre

Pri šifrirnem in dešifrirnem postopku uporabimo tabelo večkrat zapisane abecede, ki se imenuje tabula recta, Vigenèrjev kvadrat ali Vigenèrjeva tabela. Izberemo si ključ in ga nad čistopisom brez presledkov in ločil zapišemo tolikokrat, da njegova končna dolžina ustreza dolžini čistopisa. Trenutna črka v ključu predstavlja, katero vrstico tabele uporabljamo, trenutna črka čistopisa pa kateri stolpec.

	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž
A	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž
B	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A
C	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B
Č	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C
D	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č
E	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D
F	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E
G	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F
H	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G
I	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H
J	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I
K	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J
L	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K
M	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L
N	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M
O	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N
P	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O
R	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P
S	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R
Š	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S
T	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š
U	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T
V	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U
Z	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V
Ž	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z

Slika 2: Vigenèrjeva tabela

Če si izberemo ključ MARS in čistopis ŠIFRA, je prva črka zašifriranega besedila na presečišču vrstice, ki ustreza M, in stolpca, ki ustreza Š, torej G. Postopek ponovimo in dobimo celoten tajnopis GIZJM.

	A	B	C	Č	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž
													⋮												
L	L	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K
M	M	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L
N	N	O	P	R	S	Š	T	U	V	Z	Ž	A	B	C	Č	D	E	F	G	H	I	J	K	L	M
													⋮												

Slika 3: Grafični prikaz šifriranja prve črke iz zgornjega primera

Pri dešifriranju ključ zapišemo nad tajnopis po enakem postopku kot pri šifriranju. V vrstici, ki jo predstavlja trenutna črka ključa, poiščemo trenutno črko tajnopisa in tako dobimo stolpec, ki ustreza črki čistopisa, ki jo iščemo.

Postopki razbijanja šifre

Samo z uporabo grobe sile, torej tako, da preizkusimo vse možne ključe, šifre ne moremo razbiti, če ima le-ta dovolj dolg ključ. Za dolžino ključa m namreč obstaja 25^m različnih ključev. Pri dolžini ključa 18 je število možnih ključev $1,5 \cdot 10^{25}$, kar je preveliko tudi za računalnik.

Kasiski je leta 1863 prvi javno objavil postopek, ki lahko poišče dolžino ključa, in tako se po njem imenuje test Kasiskega. Če je besedilo dovolj dolgo, se ključ ponovi tolikokrat, da postane verjetnost, da se iste tri črke ključa večkrat ponovijo nad enakim zaporedjem treh črk čistopisa in tako zašifrirajo v enako zaporedje črk, dovolj velika, da test Kasiskega deluje. Test poišče dele tajnopisa, kjer se tri črke ujemajo. Največji skupni delitelj razdalj med temi deli čistopisa, m, je enak dolžini ključa.

Friedman je leta 1920 šifro razbil z indeksom sovpadanja. Ločimo indeks sovpadanja besedila in indeks sovpadanja črke. Če poznamo dolžino besedila n in vemo, da se neka črka v njem pojavi f-krat, je indeks sovpadanja te črke

$$\frac{f(f-1)}{n(n-1)}.$$

Indeks sovpadanja besedila predstavlja vsoto vseh indeksov posameznih črk v besedilu. Za besedila v slovenski abecedi znaša približno $0,063$, za povsem slučajno porazdelitev slovenskih črk pa dobimo indeks $\frac{1}{25} = 0,04$. S temi vrednostmi si lahko pomagamo pri ugotavljanju dolžine ključa.

Sam ključ nato dobimo tako, da za vsako od podzaporedij, sestavljenih iz vsake m-te črke v besedilu, poiščemo čim manjše odstopanje M_g za posamezno črko od povprečnega indeksa sovpadanja besedila v jeziku po formuli

$$M_g = \sum_{i=1}^{25} \frac{p_i f_{i+g}}{n'}.$$

V formuli n' predstavlja količnik med dolžino besedila in dolžino ključa, g indeks črke, za katero preverjamo, če nastopa v ključu, p_i pričakovano verjetnost posamezne črke abecede in f_i število ponovitev te črke v besedilu. Ko izračunamo vrednosti M_g za vsako črko abecede za trenutno podzaporedje, je največja vrednost M_g prava črka ključa. Postopek ponovimo m-krat, saj je m dolžina iskanega ključa.

Opis delovanja programa

Napisali smo tudi program, ki razbija besedila, ki so šifrirana z Vigenèrjevo šifro. Algoritem smo napisali v programskem jeziku Python. Pri iskanju šifrirnega ključa si program pomaga z indeksom sovpadanja in računanjem vrednosti M_g. S programom lahko tudi šifriramo in dešifriramo besedila z znanim ključem.

Program najprej izračuna, kolikšen je indeks sovpadanja za različne dolžine ključev. Za dolžino ključa privzame kar najmanjšo vrednost, ki presega indeks sovpadanja 0,056, kar je dovolji visoka vrednost, da se v veliki večini primerov izogne statistični napaki. Pri večkratnikih te najmanjše možne dolžine je indeks sovpadanja lahko še večji, vendar je zaradi večje dolžine ključa prisotna tudi večja napaka pri računanju vrednosti M_g, zaradi česar se splača preverjati najkrajši možni ključ. Pri računanju vrednosti M_g program preprosto izbere najvišje vrednosti iz tabele. Tako napisan program je sposoben razbiti večino besedil, tudi če so šifrirana z nekoliko daljšimi ključi, če je le besedilo dovolj dolgo, da je statistična napaka majhna. Če pa je besedilo prekratko, lahko program narobe izračuna dolžino ključa zaradi statističnega odstopanja in pri računanju ključa zaide v slepo ulico.

Zaključek

Vigenèrjeva šifra danes ni več uporabna za šifriranje sporočil, saj je algoritem za razbijanje splošno znan. Poleg tega zanjo obstaja problem prenosa ključa, ki je danes rešen s popolnoma drugačnimi postopki šifriranja, npr. sistem RSA z javnimi ključi.

Eno izmed izboljšav našega programa bi dosegli z bazo besed in črk, s pomočjo katere bi program poskusil dodati presledke v končno besedilo in preveril, če je dobljeni čistopis pravilen. Slabost našega programa je, da razbije samo osnovno šifro. Takoj ko je algoritem šifriranja malo spremenjen, program šifre ne more razbiti. Opazili smo, da program šifre ne more razbiti, če dobi premajhno količino podatkov, kar pa v splošnem velja tudi za druge vrste šifer.

Viri

[1] *Vigenère cipher*, `http://en.wikipedia.org/wiki/VigenÃĺre_cipher`, citirano dne 22. 8. 2012.

[2] *Cryptography*, `http://en.wikipedia.org/wiki/Cryptography`, citirano dne 22. 8. 2012.

[3] J. Tonejc, *Matematika šifriranja*, prosojnice iz predavanj za MaRS 2012.