Zamenjalna šifra

Rok Kaufman

21. junij 2011

1 Uvod

V zadnjem času je močno zrasla popularnost matematičnih ugank, kot sta sudoku in kakuro. Tako se v časopisih danes poleg križank pojavljajo tudi te matematične uganke. Seveda pa obstaja še veliko več različnih matematičnih ugank, kot samo tiste, ki jih danes videvamo v časopisih. Tu bomo spoznali eno izmed njih, ki je pravzaprav celo ena izmed najstarejših, saj jo človeštvo pozna že več kot 2000 let. Prav tako kot že prej omenjeni uganki omogoča neverjetno število različnih kombinacij. Ta uganka se imenuje zamenjalna oziroma s tujko substitucijska šifra.

Zamelnjalna šifra temelji na tajni abecedi, s pomočjo katere zapiše besedilo. Tajna abeceda večinoma zaradi praktičnih razlogov sestoji iz črk abecede, ki pa so premešane. Da si tajno abecedo lažje zapomnimo, je najbolje, da jo oblikujemo tako, da je v njej nek vzorec. Temu vzorcu pravimo ključ.

Poglejmo si primer tajne abecede. Po dogovoru se v kriptografiji zaradi preglednosti nešifrirano besedilo piše z malimi, šifrirano pa z velikimi tiskanimi črkami. Nešifrirano besedilo poimenujemo odprto sporočilo, šifrirano beseilo pa tajno sporočilo.

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

OPOMBA. V kriptografiji se po dogovoru sicer vedno uporablja 26-črkovna angleška abeceda, to pa iz dveh razlogov. Prvi razlog je ta, da je besedilo težje odšifrirati, če ne vemo, v katerem jeziku je napisano, kar bi lahko izvedeli iz nabora znakov. Drugi razlog pa je, da so nekatere zapletenejše šifre narejene prav posebej za angleško 26-črkovno abecedo in jih zato ne moremo uporabiti na drugih abecedah.

Če želimo šifrirati besedilo, moramo samo pogledati, katera črka tajne abecede ustreza posamezni črki navadne abecede, in preprosto vsako črko v odprtem sporočilu zamenjamo z ustrezno črko tajne abecede. Tako odprto sporočilo črko za črko pretvorimo v tajno sporočilo.

Šifrirajmo torej ta stavek.

Črki š ustreza črka R, črki i črka A, črki f črka O... Ta postopek ponovimo za vse črke odprtega sporočila.

Dobimo tajno sporočilo:

RAOCACSGUD PDCZG PS IPSMZE.

Da besedilo še bolj zakrijemo, lahko izbrišemo presledke in ločila:

RAOCACSGUDPDCZGPSIPSMZE

Besedilo je tako še bolj nepregledno in s tem še dodatno otežimo razbijanje ključa.

Število možnih ključev je enako številu možnih različno premešanih abeced. To število izračunamo tako, da zmnožimo vsa števila od 1 do števila črk, kar je v našem primeru 25. Temu zmnožku se v matematiki reče fakulteta ali faktoriela, zapiše se pa kot $n!$, torej 25!. Na prvem mestu je lahko namreč 25 različnih črk, na drugem mestu nato 24 različnih, ker je ena izmed črk abecede že na prvem mestu. Na vsakem naslednjem mestu je ena možnost manj. Na predzadnjem mestu sta tako lahko le še 2 različni črki, črka na zadnjem mestu pa je že določena z prejšnjimi izbranimi črkami.

Kaj pa se zgodi, če zamenjalno šifro uporabimo dvakrat zapored? Ali je takšno šifriranje bolj varno kakor enkratno šifriranje?

Predstavljamo si, da tajno abecedo dobimo tako, da črke premešamo, kot bi mešali karte. Z dovolj dolgim mešanjem lahko dobimo vseh 25! možnih tajnih abeced. Če črke tajne abecede še enkrat premešamo, ne moremo dobiti nobene tajne abecede, ki je že s prvim mešanjem ne bi mogli dobiti. Število različnih šifriranj se z večkratnim mešanjem, torej z več tajnimi abecedami, ne poveča, ker bi lahko vsako kombinacijo večih tajnih abeced lahko opisali z le eno tajno abecedo.

OPOMBA: ALI BI DALI PRIMER??

OPOMBA. Postopek šifriranja s zamenjalno šifro lahko označimo kot preslikavo, ki elementom množice črk abecede priredi elemente množice črk tajne abecede. Ta preslikava je bijektivna, saj se v vsako črko tajne abecede preslika natanko ena črka navadne abecede. Če preslikava šifriranja ne bi bila bijektivna, bi bilo nemogoče odšifrirati besedilo, saj imajo le bijektivne preslikave svojo inverzno preslikavo, ki v našem primeru omogoča odšifriranje. Dve bijektivni preslikavi pa lahko zamenjamo z eno samo bijektivno presli-

kavo. To pomeni, da lahko dvakratno šifriranje s zamenjalno šifro dobimo tudi z enkratnim šifriranjem z drugim ključem, kar pomeni, da dvakratno šifriranje po enakem postopku ne oteži razbijanja šifre.

2 Zgodovina

Prvi, ki so dokumentirano uporabljali zamenjalno šifro so bili Hebrejci. Uporabljali so postopek, imenovan atbaš, ki ga bom kasneje tudi opisal. Za pošiljanje tajnih sporočil je zamenjalno šifro uporabljal tudi Julij Cezar, ki si je pomagal s poenostavljenim postopkom. Problem, kako brez znanja ključa rešiti uganko zamenjalne šifre in prebrati tajno besedilo, je dolgo veljal za nerešljivega, vendar so arabski učenjaki do 9. stoletja že našli način, kako najti ključ in odšifrirati tajno sporočilo. Z začetkom renesanse so mnoga arabska odkritja na področju matematike prodrla v Evropo. Leta 1586 je Thomas Phellipes s tem, ko je po arabskem postopku zlomil šifro, ki jo je uporabljala Marija Stuart, preprečil zaroto proti kraljici Elizabeti I. Takrat je postalo jasno, da je zamenjalna šifra preveč preprosta, da bi lahko zagotavljala tajno komunikacijo in kriptografi so začeli iskati bolj zapletene postopke za šifriranje besedil.

3 Preprostejši šifrirni ključi

Posebni primeri ključev so dobili tudi svoja poimenovanja. Te ključe si zaradi posebnosti lažje zapomnimo, in zato so se prav ti kot prvi tudi uporabljali. Vendar so iz istega razloga tudi manj varni od ostalih. Zakaj so takšni ključi manj varni, bom razložil kasneje v članku. Oglejmo si nekaj takšnih ključev z zgodovinskim pomenom.

3.1 Atbaš

Šifra atbaš, ki so jo uporabljali stari Hebrejci, za tajno abecedo uporablja abecedo napisano v obratnem vrstnem redu. Prva črka se tako zamenja z zadnjo, druga s predzadnjo in tako dalje do zadnje, ki se zamenja s prvo. Šifra atbaš omogoča zgolj en način šifriranja v vsaki abecedi.

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

Pri šifri atbaš pa obstaja manjši problem. **UOK HIBJŽKI HIFDIHŠM, AŠ LŽPMI HGŠZŠGŠKI CFŽ EOSGOGŽJŽ ZŠFŠTOLŽ.** Problem je zapisan v šifriranem sporočilu.

OPOMBA. Atbaš pa ne ustreza Kerckhoffovemu principu, ki pravi, da je kriptosistem varen, če nasprotnik ne more odšifrirati sporočil tudi v primeru, da pozna postopek. Šifra atbaš nima ključa, saj omogoča zgolj en način šifriranja.

3.2 Premična šifra

Drug preprost način za premešanje abecede je ta, da vso abecedo prestavimo za neko število mest naprej. V tem primeru govorimo o premični šifri. Ker je v abecedi 25 črk, obstaja 24 različnih ključev v premični šifri, saj bi se pri petindvajsetem ključu abeceda preslikala sama vase. Takšno obliko šifriranja je uporabljal Julij Cezar. Vsako črko v sporočilu je zamenjal s črko, ki stoji tri mesta naprej v abecedi.

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
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

NSENČ MH ŠČGOČ! Seveda je bilo Cezarjevo sporočilo v latinščini. Cezarjev ključ je seveda izgledal nekoliko drugače, saj takratna abeceda še ni imela črk J, U in šumnikov, vsebovala pa je Q, X in Y.

Premično šifro se da zlomiti že tako, da poskusimo z vsemi štiriindvajsetimi možnimi tajnimi abecedami, saj to ne vzame veliko časa. Če črke abecede naključno premešamo, dobimo veliko večje število možnih ključev. Teh ključev je 25!, kar je 15511210043330985984000000 - 15 bilijonov bilijonov. V tem primeru vseh ključev vsekakor ne bi mogli pregledati.

3.3 Substitucijska šifra s ključno besedo

Avtor šifriranega besedila lahko uporabi nekoliko preprostejši ključ. Tajno abecedo sestavi tako, da si izbere ključno besedo, in nato vse črke ključne besede zapiše na začetek tajne abecede in seveda izpusti črke, ki se ponovijo. Preostanek abecede zapolni tako, da zapiše preostale črke abecede v pravem vrstem redu začenši z zadnjo črko v ključni besedi. Namesto ključne besede lahko uporabimo tudi stavek. V primeru sem kot ključno besedo uporabil besedo presek.

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

Število ključev, ki jih dobimo s pomočjo ključne besede ali stavka, je dosti manjše od števila vseh možnih tajnih abeced. Vendar je še vedno preveliko, da bi se dalo šifro razbiti po principu poskušanja. Ključ pa si je dosti lažje zapomniti oziroma posredovati prejemniku sporočila.

STRHE FZISŽ ČSCFRJUŽ GHNF K KČJU NŽČGŽBVJU UJNFRH. V tem primeru je ključ naslov članka.

Koliko pa je možnih ključev v tem primeru? Na začetku tajne abecede je ključna beseda, ki je dolga n črk. Prvih n črk tajne abecede torej tvorijo katerekoli črke abecede, medtem ko so preostale črke že določene, ker gre zgolj za črke v vrsnem redu abecede. Število možnih ključev je torej v primeru ključne besede dolžine n zmnožek števil od 25 navzdol do $25 - n$, oziroma $\frac{25!}{(25-n)!}$. Seveda pa je to število večje, kot je dejansko število možnih ključev, saj so v njem zajeta tudi zaporedja n črk, ki predstavljajo besedo.

4 Razbijanje šifre

Razbijanje šifre v kriptografiji pomeni, da pridemo do odprtega sporočila, ne da bi prvotno poznali ključ. S tujko ga imenujemo tudi kriptoanaliza.

Substitucijsko šifro je zaradi njene preprostosti kljub velikemu številu možnih kljčev lahko zlomiti. Pri razbijanju šifre si pomagamo s statistiko pogostosti pojavljanja črk in znanjem jezika, v katerem je napisano šifrirano besedilo.

Vse črke se ne pojavljajo enako pogosto in prav na tem temelji postopek, s katerim lahko zlomimo zamenjalno šifro. V vsakem jeziku se nekatere črke pojavljajo pogosteje kot druge. Če preštejemo, kolikokrat se posamezna črka pojavi v zelo veliki bazi besedil, lahko ugotovimo, kolikšen delež vseh črk zajema ta črka. Ti procenti so značilni za vsak jezik in podobna razmerja med črkami se pojavljajo v skoraj vseh besedilih v temu jeziku. Daljše ko je besedilo, bolj se procenti približajo splošnim razmerjem med črkami. V slovenščini so najbolj pogoste črke a, e, i, o, n in r; v angleščini e, t, a, o, i in n. Ker se pri šifriranju v zamenjalni šifri vsaka črka abecede preslika vsakič v isto črko tajne abecede, se razmerja med tem, kako pogosto se posamezne črke pojavijo, ohranijo. Le črke so druge. Zato lahko približno sklepamo, katere črke tajne abecede ustrezajo črkam navadne abecede.

Poglejmo si nekaj dejstev, ki veljajo za slovenščino.

Daleč najpogostejše štiri črke v slovenščini so a, e, i in o. To pomeni, da najpogostejši znaki v šifriranem slovenskem besedilu predstavljajo te štiri

samoglasnike. Ker pa se ti štirje samoglasniki pojavljajo približno enako pogosto, ne moremo takoj vedeti, kateri je kateri.

V primeru, da imamo v tajnem besedilu presledke, si lahko pomagamo tudi z eno- in dvočrkovnimi besedami. V slovenščini so enočrkovne besede a, h, k, o, s in z. Torej črke tajne abecede, ki se v besedilu pojavljajo same, predstavljajo prav te črke. Tudi dvočrkovnih besed je bolj malo, vse pa so sestavljene iz enega soglasnika in enega samoglasnika, kjer črka r lahko šteje kot samoglasnik.

Prav tako velja, da se na začetku besede najpogosteje pojavljajo črke (KATERE), na koncu pa (KATERE). Nekatere veččrkovne končnice so bolj pogoste zaradi pravil pregibanja besed.

V vsaki besedi, razen enočrkovnih, nastopi vsaj en samoglasnik. Če nam po tem, ko smo določili vse samoglasnike in r, ostane kakšna beseda brez samoglasnika, nekaj počnemo narobe. Poleg tega sta dva sosednja samoglasnika v besedi v slovenščini zelo redka in tudi to lahko kaže na to, da smo na napačni poti.

OPOMBA. Če približno vemo, o čem govori besedilo, si lahko še dodatno pomagamo z naborom besedišča. Med drugo svetovno vojno so si angleški vohuni pri odšifriranju sporočil Enigme pomagali s tem, da so razbili sporočilo z vremensko napovedjo, saj je imelo najbolj predvidljivo strukturo in ga je bilo zaradi tega najlažje razbiti. Poleg tega je bilo oddano vsak dan ob isti uri.

OPOMBA. Podobna analiza je v angleščini dosti preprostejša. Črka e je bistveno pogostejša od ostalih črk. Prav tako nam razbijanje šifre olajšajo določni ter nedoločni členi (the, a, an). Prav tako so kombinacije črk predvidljivejše kot v slovenščini, saj obstajajo zelo pogoste kombinacije, kot so: ch, sh, th...

Na začetku reševanja, ko naredimo analizo pogostosti črk, je dobro preveriti, ali imamo opravka s Cezarjevo šifro, atbašem, oziroma katerokoli drugo zamenjalno šifro z očitnim vzorcem. Takšna vzorca sta lahko recimo, da so razdalje v abecedi med štirimi najpogostejšimi znaki enaki kot so razdalje med črkami a, e, i in o pri Cezarjevi šifri, oziroma da so najpogostejše črke Ž, Š, O in I pri atbašu, saj so te črke ravno nasproti najpogostejšim štirim samoglasnikom. Seveda pa je takšna razporeditev lahko zgolj slučajna. V tem primeru ta bližnjica ne vodi nikamor.

5 Zaključek

Članek o problemu zamenjalne šifre se je v Preseku že pojavil. Presek V/1 je na straneh 40-42 vseboval članek z naslovom Skrivnostno sporočilo (avtor Tomaž Pisanski), v katerem je bil primer tajnega sporočila s šifriranimi presledki, torej je bil presledek obravnavan kot črka. V primeru, ko ima besedilo šifrirane presledke, to takoj opazimo, saj je presledek pogostejši od vseh črk abecede in ga lahko takoj določimo.

Substitucijska šifra ponuja veliko možnosti za matematične uganke, saj predstavlja zanimiv miselni problem z zelo velikim številom različnih primerov za reševanje. Če pa koga zanima še kaj več o šifriranju, mu v branje priporočam v slovenščino prevedeno knjigo Simona Singha z naslovom Knjiga šifer.

6 Izzivi

1. O GHMJSGČZF KMHČC OSK, OJB IH CFZGN. O MZČ OSKCTC ČZ RCOZE O KMSJCB USKCB DJISG, FHUSG CG KCEZG UEHOZD. ŠCE ČZ GZDC MHECD, VS AS GC DFSEN MSDZAS. VZES FN GC ŠCEH FSJ; SFISD GHKCE ČZ HV FHJČS GS KOHČC DHŠCECTC SGAEZLDH KHE, DSJ ČZ ŠCEH IS RZ MCKMCDJSM HKMJH IJZIHOZVSGH. ISPCEC KH AS FZČSUC, VS ŠC AS DČZ GZBHMHFS PSEZPEC; IHLMZGZAS ŠHČS R GČCF KH KZ ŠSEC JSOGH MSDH DSDHJ IHPGZČZ LMZFICBSJČS. DJISG KZ ČZ IS OZVGH NFCDSE CG AEZVSE, VS FN GCKH FHAEC VH RCOZAS.

2. ŽCIJO, ŠO GD ŠTITECLO ZNTRTEDIO TZFOIO FOJLO, JC NTUSRTEDIO, ŠO ZT ŠRŽOUC VUCŠIC ZUTJC ČDMRDRLC ZIVŽGC, BD LOJ GD ROSUDIC LOJGTIJČC HTPTEC ČDMRC DL GDIC TŠPTUTRLC SO UORLT NRCLOČOLJC ZNTRTEDI. KBROFD Z FCH NO ZT LOZNRTFLDBTUD ŠCBRDNFCRJD NTZBVČOID VBROZFD FC ČDMRC DL S LJDHD FVŠD ROSLC ZBRDULTZFD. ŠCBRDNFCRJD ZT JCSDBTULD OIBDHDZFD, S HDFD TGŠOLT NICHC, BD NTZBVČO DS URZFC ZDHGTITU GRCS NTHCLO NRDEOROFD ZHDZCILC GCZCŠC. SPTŠTUDLO FOJLDK NDZOU, BTŠTU DL ČDMCR JC SPTŠTUDLO ZFTICFJO ZFORCPO GTJO HCŠ ČDMRCRJD DL ŠCBRDNFCRJD, ŠVKTULC TGTRTŽCUOILC FCBHC, BD ŠROHOFDELT UNIDUO LO NTFCB SPTŠTUDLC.