TO IN ONO O TAJNOPISIH

V prejšnji številki Preseka smo na ugankarski strani objavili šifrirano pisemce, katerega namen je bila bolj vaja iz koordinatnih sistemov kot pa njegovo dešifriranje. Razvozlati ga ni bilo prav nič težko, saj je bila priložena tudi mreža - ključ za dešifriranje. Brez ključa pa bi bila ta drobcena naloga kar lep kriptološki problem. Zakaj, bomo razložili nekoliko kasneje.

Kriptologijo, vedo o tajnih pisavah, sestavljata dve veji. Kriptografija uči, kako lahko sporočila bolj ali manj dobro šifriramo, kriptanaliza pa se ukvarja s prav nasprotnim: Kako razvozlati prestreženo sporočilo, če ključa ne poznamo; drugače povedano, kako "zlomiti" kodo sporočila.

Obe vedi imata od nekdaj pomembno vlogo na diplomatskem in vojaškem področju. Iz zgodovine je znan primer, ko je med špansko-francosko vojno leta 1589 francoski matematik Vieta po naročilu svojega kralja razvozlal ključ tajne pisave, ki so jo Španci uporabljali v vojnih načrtih. Pisava je bila za takratne razmere tako zamotana, da so se Španci počutili povsem varne. Njena analiza bi z računalnikom najbrž ne bila prehud problem, tedaj pa so bili Španci zaradi njenega odkritja tako pretreseni, da so se celo pritožili pri papežu, češ da si Francija v vojni pomaga s čarovnijo.

V računalniški dobi se je vloga kriptografije še povečala zaradi potrebe po varnem shranjevanju poslovnih in osebnih podatkov. Tajno kodiranje uporabljajo tudi pri igrah na srečo, da se izognejo goljufijam s ponaredki.

V tem prispevku si bomo ogledali nekaj preprostih načinov šifriranja, ki so jih uporabljali v preteklosti, v prihodnji številki Preseka pa ilustrirali moderno metodo, imenovano šifriranje z javnim ključem. Pri tem se bomo omejili le na razlago osnovnih idej, ob strani bomo pustili njihovo računalniško izvedbo.

Cezarjeva metoda

Najpreprostejši tajnopisi temelje na permutaciji črk v abecedi jezika, v katerem je sporočilo napisano. To pomeni, da zamenjamo, v skladu z nekim pravilom, posamezno črko abecede z natanko določeno drugo črko abecede. Eno takih metod je uporabljal tudi Julij Cezar. Njegova sporočila so šifrirali tako, da so posamezno črko nadomestili s črko, ki stoji tri mesta za njo v abecedi, zadnje tri črke abecede pa so nadomestili s prvimi tremi. V slovenščini predstavlja ključ za Cezarjev način šifriranja naslednja tabela:

A	B	C	Č	D	E	F	G	H	I	J	K	L	M
Č	D	E	F	G	H	I	J	K	L	M	N	O	P

N	O	P	R	S	Š	T	U	V	Z	Ž
R	S	Š	T	U	V	Z	Ž	A	B	C

Sporočilo

TAKOLE JE ŠIFRIRAL CEZAR

se šifrirano glasi:

ZČNSOHMHVLITLTČOEHBČT.

Presledki med besedami so namerno izpuščeni, da bi morebitna nepoklicana oseba sporočilo težje razvozlala. Pri tem smo seveda predpostavili, da bo pravemu naslovniku besedilo TAKOLEJEŠIFRIRALCEZAR dovolj domače, da bo znal postaviti presledke na prava mesta.

Tri seveda ni nobeno magično šifrirno število. Črke abecede bi lahko premaknili za poljubno število mest. Pri tem bi očitno vse bistveno različne načine dobili s premiki za manj kot 25 mest.

Postopek lahko opišemo tudi z matematičnim izrazom, če črke nadomestimo s števili, ki pomenijo njihova mesta v abecedi, pri čemer začnemo številčiti z 0:

A	B	C	Č	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12	13

N	O	P	R	S	Š	T	U	V	Z	Ž
14	15	16	17	18	19	20	21	22	23	24

Če označimo z x število, ki pripada črki sporočila, in z y število, ki pripada njeni šifri, lahko opišemo Cezarjevo metodo s formulo

$$y \equiv x + 3 (\mathrm{mod}\ 25),$$

šifriranje s premikom za d mest pa z

$$y \equiv x + d (\mathrm{mod}\ 25). \qquad (1)$$

Šifri prirejeno število y je torej ostanek, ki ga dobimo, če $x + d$ delimo s 25.

Ključ, s katerim lahko ta tip tajnopisa razvozlamo, je seveda naravno število $d \leq 24$. Od števila, ki v zgornji tabeli pripada šifri, odštejemo d in, v primeru, da je rezultat negativen, prištejemo še 25. Nato poiščemo v tabeli črko, ki pripada dobljenemu številu. Postopek dešifriranja torej opisuje formula

$$x \equiv y - d (\mathrm{mod}\ 25),$$

kjer je x najmanjše nenegativno število, ki ji pri danih y in d ustreza.

Opisani način šifriranja (tudi za splošni d mu bomo rekli kar Cezarjev način) je za uporabnike zelo preprost, a ga je tudi zelo lahko zlomiti. S pogosto menjavo ključa d lahko dosežemo sicer nekoliko večjo varnost, vendar si s tem nakopljemo skrb za varen prenos ključa do naslovnika. Možni so seveda vnaprejšnji dogovori. Ena takih možnosti je, da na posamezni dan velja ključ, ki je enak vsoti števk tistega dela tekočega datuma, ki označuje dneve. Če bi ob takem domenku, denimo, 14. februarja (na valentinovo) dobili sporočilo SNZRNSŠEAJ, bi vedeli, da se moramo za vsako črko besedila vrniti v abecedi za $d = 1 + 4 = 5$ črk nazaj. Prva črka sporočila tako ustreza številu $18 - 5 = 13$ in je M, vse sporočilo pa se glasi MISLIM NATE.

Seveda lahko nepoklicani ugotovi, da uporabljamo Cezarjev način šifriranja. Potem zlahka odkrije tudi vsakokratno vrednost ključa d. Za d preprosto vstavi zapored vseh 24 možnosti in zelo verjetno bo dešifrirano sporočilo smiselno le pri eni vrednosti premika d.

Učinkovitejši je zlom kode na način, ki temelji na pogostosti posameznih črk v besedah jezika, za katerega domnevamo, da je v njem tajno sporočilo napisano. Naslednja tabela, dobljena seveda statistično, prikazuje v promilah izraženo pogostost posameznih črk v slovenščini.

E	A	I	O	N	R	S	L	J	T	V	D	K
108	102	89	88	69	53	52	47	45	45	40	36	35

M	P	U	Z	B	G	Č	H	Š	C	Ž	F
33	31	22	21	18	15	15	11	10	7	7	1

Oglejmo si tak način kriptanalize kar na primeru. Denimo, da smo prestregli sporočilo VKSŽPŠPACLSŠRČJCGRUVZPAG, za katero domnevamo, da je napisano v slovenščini in šifrirano na Cezarjev način z neznanim premikom d. V sporočilu največkrat nastopa znak P, zato najprej poskusimo, če je to morda šifra za črko E, ki je najpogostejša črka v slovenskem jeziku. Izračunamo $d = 16 - 5 = 11$, vendar je s takim d dešifrirano sporočilo

KAGMEHENPBGHFRŽPUFJKLENU nesmiselno, kar odkrijemo že po nekaj prvih črkah. Podobno propade poskus s črko A, za I, tretjo najbolj pogosto črko v slovenščini, pa dobimo $d = 7$, ki nas vodi (potem, ko smo smiselno vstavili presledke) do sporočila ODKRILI STE KLJUČ TAJNOPISA.

Modificirana Cezarjeva metoda

Šifriranje po Cezarjevo torej ni hudo varen način pisanja tajnih sporočil. Obstajajo različne modifikacije Cezarjeve metode, ki izboljšajo varnost tajnopisov. Tako lahko namesto (1) uporabimo kakšno drugo šifrirno funkcijo, na primer

$$y \equiv ax + b (\mathrm{mod}\ 25). \tag{2}$$

Vse možnosti, ki jih daje formula (2), dobimo, ko a in b pretečeta vsa nenegativna števila manjša od 25. Pri tem smemo a izbirati le med števili, ki so tuja s 25, sicer sporočila ne bo moč dešifrirati (zakaj?).

V daljših sporočilih je seveda nujna tudi uporaba ločil in presledkov. Običajno jih dodamo na koncu abecede in jim, podobno kot črkam, priredimo nadaljnja zaporedna števila. Če ima tako razširjena abeceda n znakov, preideta formuli (1) in (2) v

$$y \equiv x + d (\mathrm{mod}\ n) \quad \text{in} \quad y \equiv ax + b (\mathrm{mod}\ n),$$

število a pa mora biti tuje z n.

Pa se povrnimo k formuli (2). Pošiljatelj in prejemnik sporočila se morata na neki način dogovoriti glede izbire števil a in b, ki tokrat predstavljata ključ šifre, ali pa bo treba ključ prenesti. Samo šifriranje in dešifriranje poteka najenostavneje s tabelo, katere pripravo si oglejmo kar na primeru za $a = 7$ in $b = 4$. Iz $y \equiv 7x + 4 (\mathrm{mod}\ 25)$ sledi, da je šifra za A($x = 0$) znak D($y = 4$), šifre nadaljnih črk abecede pa dobimo tako, da pri cikličnem sprehodu skozi abecedo izpisujemo vsako sedmo črko, začenši z D. S tako dobljeno tabelo

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

N	O	P	R	S	Š	T	U	V	Z	Ž
C	I	P	Z	E	L	Š	B	H	O	V

hitro preberemo, da sporočilo KITKCRCDPDGIKŠZNJ pomeni BOMBNI NAPAD OB TREH.

To sporočilo bi povzročilo nekaj več sivih las vsiljivcu, ki bi se polotil njegove analize. Čeprav bi vedel, da je šifrirano s formulo (2), bi moral pravilno uganiti vsaj dve črki, da bi lahko izračunal ključ $a = 7$, $b = 4$. Uporaba tabele za pogostost črk bi mu bolj malo koristila, saj je B, ki v sporočilu nastopa največkrat, v tabeli šele na osemnajstem mestu. Če pa bi domneval, da govori sporočilo o bombardiranju, bi morda uganil, da je K($y = 11$) šifra za B($x = 1$) in I($y = 9$) šifra za O($x = 15$). Z vstavljanjem vrednosti za x in y v (2) bi sledilo, da a in b ustrezata sistemu kongruenc

$$a + b \equiv 11 (\mathrm{mod}\ 25)$$
$$15a + b \equiv 9 (\mathrm{mod}\ 25).$$

Sistem lahko hitro rešimo s standardnimi metodami, podobnimi metodam za reševanje linearnih sistemov. Če pa nam to ni po volji, lahko še vedno preverimo, pri katerem izmed 500 možnih parov (a, b) (ne pozabimo, da mora biti a tuj s 25) dobimo smiselno sporočilo.

Druge zamenjalne metode

Naslednji korak k večji zapletenosti šifre je uporaba poljubne permutacije (razširjene) abecede. V takem primeru predstavlja ključ tajnopisa tabela, s katero je vsaki črki abecede prirejena natanko določena druga črka abecede. Največji problem te metode je varen prenos ključa, saj se ga ne da nadomestiti s tako preprostim opisom, kot je bil opis enega ali dveh števil pri prejšnjih metodah. Zgled takega primera tajnopisa je tudi Urškino šifrirano pisemce iz prejšnje številke Preseka, čeprav so bile tam (zaradi vaje iz koordinatnih sistemov), namesto črk, izbrane nekoliko bolj nerodne oznake za šifre.

Prav gotovo je način šifriranja na osnovi permutacije varnejši od metod, opisanih v prejšnjih dveh razdelkih, seveda ob predpostavki, da nepoklicani ni prestregel ključa. Vendar se da tudi tak tajnopis dokaj hitro razvozlati z metodo pogostosti črk, tako da ne moremo za šifriranje stalno uporabljati iste permutacije. O kriptanalizi na osnovi pogostosti črk obstajajo namreč za posamezne jezike cele študije. Razlog je preprost. Že pri naših primerih kriptanalize smo opazili, da se vrstni red pogostosti znakov v opazovanih besedilih ne ujema povsem z vrstnim redom v tabeli, ki kaže pogostost znakov v slovenščini. Čim krajše je besedilo, tem manj verjetno se to zgodi. Zato v kriptanalizi dodatno upoštevajo, katere črke najpogosteje nastopajo na začetku, katere na koncu besed, nadalje pogostost parov zaporednih črk, itd.

V vsakem primeru pa velja, da je sporočilo tem laže razvozlati, čim daljše je. V daljših besedilih pridejo namreč statistične značilnosti jezika bolj do izraza. Zato morajo biti tajna sporočila kratka, ključ je treba pogosto menjati.

Predlagam vam, da poskusite na osnovi pogostosti črk razvozlati Urškino pisemce, ne da bi pri tem kukali v priloženi ključ za dešifriranje. Čeprav besedilo ni najkrajše, boste videli, da je kar trd oreh.

Oglejmo si še način šifriranja, ki mu z metodo na osnovi pogostosti črk vsiljivec ne more biti kos. Osrednja ideja je, da vsako črko sporočila zamenjamo s črko, ki stoji v abecedi d mest dalje, pri čemer premik d spreminjamo, v skladu z nekim pravilom, od črke do črke, natančneje od mesta do mesta, na katerem črka v sporočilu stoji. Pravilo običajno uvaja neka dogovorjena beseda, ki je ključ šifre. Ključno besedo ponavljaje se zapišemo pod sporočilo, črko pod črko, in nato vsako črko sporočila premaknemo za število, ki pripada podpisani črki ključne besede.

Oglejmo si metodo spet kar na primeru. Denimo, da je dogovorjena ključna beseda BOGASTVO in da želimo šifrirati sporočilo TAKOJ PRODAJ LIRE. Takole gre:

T	A	K	O	J	P	R	O	D	A	J	L	I	R	E
B	O	G	A	S	T	V	O	B	O	G	A	S	T	V

U	O	S	O	Č	K	N	E	E	O	R	L	C	L	C

Šifrirano sporočilo se torej glasi UOSOČKNEEORLCLC. Črko P($x = 16$) smo na primer premaknili ciklično za dvajset mest, ker je 20 zaporedno število podpisane črke T, in zanjo dobili šifro K($y = 11$). To ustreza računu $16 + 20 \equiv 11 (\mathrm{mod}\ 25)$. Koristno si je pripraviti tabelo 25×25 črk, ki pripadajo na opisani način parom črk v abecedi. Tako stolpce kot vrstice označimo s črkami od A do Ž. Na križišču stolpca P in vrstice T stoji v tej tabeli po zgornjem računu črka K.

Tabela je seveda uporabna pri poljubni ključni besedi, ki pa jo je priporočljivo pogosto menjati. Lahko se, recimo, domenimo, da na posamezni dan velja ključna beseda, ki jo sestavlja prvih osem črk s tretje strani en dan starega časopisa Delo.

Šifriranje zaporednih parov črk

Obstaja še veliko šifrirnih metod. Vsem je cilj čimbolj otežiti kriptanalizo. Kot zadnjo v tem sestavku si oglejmo metodo, s katero namesto posameznih črk šifriramo zaporedne pare črk v sporočilu. Eden od načinov, kako to napravimo, je, da si izberemo štiri števila a, b, c, d in za šifriranje uporabimo sistem

$$y_1 \equiv ax_1 + bx_2 (\mathrm{mod}\ 25)$$
$$y_2 \equiv cx_2 + dx_2 (\mathrm{mod}\ 25).$$

Za $a = 1$, $b = 3$, $c = 7$ in $d = 12$ preide par črk ET z zaporednima številoma $x_1 = 5$ in $x_2 = 20$ v šifro OA, ker je

$$y_1 = 1 \cdot 5 + 3 \cdot 20 = 65 \equiv 15 (\mathrm{mod}\ 25)$$

in

$$y_2 = 7 \cdot 5 + 12 \cdot 20 = 275 \equiv 0 (\mathrm{mod}\ 25),$$

kar sta števili, ki pripadata črkama O in A. Tako par za parom zaporednih črk šifriramo vse sporočilo.

Za dešifriranje moramo pri danih y_1 in y_2 razrešiti zgornji sistem na x_1 in x_2. To gre v primeru, če je število $ad - bc$ tuje s 25.

Kriptanaliza takega tajnopisa bi seveda potekala z uporabo tabel o pogostosti parov zaporednih črk v jeziku sporočila.

– – –

Računalnikarji med bralci Preseka boste morda za kakšnega od obravnavanih načinov izdelali program za šifriranje in dešifriranje. Takšen je dandanes tudi način dela v kriptografiji in kriptanalizi. Svojčas pa je bilo to tudi za preprostejše sisteme šifriranja hudo zamudno in težaško delo, pri katerem so si pomagali tudi s posebnimi šifrirnimi stroji. Eden najbolj znanih šifrirnih strojev je bila Enigma, ki so jo uporabljali Nemci med drugo svetovno vojno. Skupina angleških kriptanalitikov, ki jo je vodil matematik Alan Turing, je razvila metodo in stroj za dešifriranje prestreženih sporočil, kar je igralo veliko vlogo v zmagi zavezniških sil. Nemci so bili namreč trdno prepričani, da je šifriranje z Enigmo brez ključa nezlomljivo, tako da je bil ta vir informacij zaveznikom na voljo ves čas vojne.

Marija Vencelj