Crypto-World

Informační sešit GCUCMP

ISSN 1801-2140

Ročník 12, číslo 4/2010 **15. duben 2010**

4/2010

Připravil: Mgr. Pavel Vondruška

Sešit je přednostně distribuován registrovaným čtenářům.
Starší sešity jsou dostupné na adrese
http://crypto-world.info
(1358 registrovaných odběratelů)

Obsah:

A. Analýza Blue Midnight Wish – srovnání složitosti (bezpečnosti) BMW a dalších kandidátů SHA-3

RNDr. Vlastimil Klíma, kryptolog konzultant, Praha
(http://cryptography.hyperlink.cz, v.klima@volny.cz)
Prof. Danilo Gligoroski, Norwegian University of Science and Technology, Norway (danilog@item.ntnu.no ,
http://www.item.ntnu.no/people/personalpages/fac/danilog/start)

Článek navazuje na příspěvky v číslech 1 - 3 Crypto-Worldu 2010, s nímž má společnou skoro celou úvodní stranu a několik obrázků. Volně také navazuje na články o BMW v 12/2009, 3/2009 a 7-8/2009. V čísle 1 jsme se zabývali hledáním vzoru (úloha první), v čísle 2 hledáním kolize (úloha druhá), v čísle 3 různými bloky BMW.

Chceme stimulovat analýzy a útoky na BMW a prezentovat otevřené problémy. Ty by se mohly stát předmětem studentských prací. Proč? Velkou výhodou oproti jiným tématům je, že tyto rozbory jsou nyní velmi žádané, ať s negativním nebo pozitivním výsledkem. Když bude problém vyřešen nebo naopak bude ukázáno, že je složitý, je to v obou případech žádaný a velmi dobře publikovatelný výsledek.

Označení
Článek bude využívat označení zavedené v Crypto-Worldu 12/2009. Připomeňme jen šířku slova w = 32 nebo 64 bitů, délku bloku zprávy a průběžné haše n = 16*w (mají 16 slov) a výpočet haše:

1. Předzpracování
(a) Doplň zprávu M jednoznačným definovaným způsobem o délku zprávy v bitech a doplněk
(b) Rozděl zprávu na celistvý násobek (N) m-bitových bloků $M^{(1)}$, ..., $M^{(N)}$.
(c) Nastav počáteční hodnotu průběžné haše $H^{(0)}$ na konstantu ($CONST^0$).
2. Výpočet haše
For i = 1 to N: $H^{(i)} = f(M^{(i)} , H^{(i-1)})$.
3. Finalizace
$H^{final} = f(H^{(N)} , CONST^{final})$, kde $CONST^{final}$ je konstanta.
4. Závěr
H(M) = dolních n bitů z hodnoty H^{final}.

O míře složitosti
V následujícím se pokusíme najít přijatelnou a prakticky využitelnou míru složitosti kandidátů na SHA-3. Víme, že pro každého kandidáta je úloha nalezení kolize nebo vzoru převeditelná na řešení systému Booleovských rovnic. Proto se pokusíme porovnat jejich složitost. Navrhneme míru, která sice není perfektní, ale uvidíme, že složitost těchto kandidátů dobře odráží. V každém případě získáme horní odhad složitosti.

Definice složitosti
Budeme porovnávat složitost zápisu Booleovských rovnic v algebraické normální formě. Tato forma zápisu využívá operace XOR a AND. Pochopitelně, že více nám vadí operace AND, neboť při absenci operací AND by se jednalo o snadno řešitelný lineární systém. Avšak i

operace XOR budeme počítat a budeme též počítat počet všech meziproměnných. Jedinými proměnnými jsou sice jen bity zprávy, avšak než se vypočítá hašovací hodnota, vznikne řada meziproměnných, což je nejlépe vidět z programové realizace dané funkce. Každou dílčí funkci hašovací funkce můžeme napsat jako polynom a každý polynom můžeme zapsat postupně pomocí meziproměnných tak, že vznikají pouze elementární rovnice typu
$a = b \oplus c$,
$a = b*c$,
kde * označuje AND a kde proměnné a, b, c jsou Booleovské proměnné.

Takto můžeme například zapsat aritmetické sčítání w-bitových slov $a = b + c$ modulo 2^w, a to následovně. Označíme bity $i = 0 ,..., w-1$ indexem i, tj. $a = (a_{w-1}, ..., a_0)$, kde a_{w-1} je nejvyšší bit a a_0 nejnižší. Bity přenosu označíme $carry_i$ pro $i = 1, ..., w-1$, což jsou právě vznikající meziproměnné (kromě $a_{w-1}, ..., a_0$).

Máme
pro bit $i = 0$:
$a_0 = b_0 \oplus c_0$, $carry_1 = b_0 * c_0$,
pro bit $i = 1, ..., w - 2$:
$a_i = b_i \oplus c_i \oplus carry_i$, $carry_{i+1} = b_i*c_i \oplus b_i*carry_i \oplus c_i*carry_i$,
pro bit $i = w - 1$:
$a_{w-1} = b_{w-1} \oplus c_{w-1} \oplus carry_{w-1}$.

Funkci $carry_{i+1} = b_i*c_i \oplus b_i*carry_i \oplus c_i*carry_i$ rozložíme na elementární operace, ale rovnou uděláme drobnou optimalizaci, když b_i vytkneme z prvních dvou členů a součet $c_i \oplus carry_i$ použijeme také k výpočtu a_i.

Máme:
$s_i = c_i \oplus carry_i$,
$t_i = b_i * s_i$,
$u_i = c_i * carry_i$,
$carry_{i+1} = t_i \oplus u_i$.

Tedy sčítání slov $a = b + c$ můžeme zapsat normalizovaně elementárními operacemi takto:
pro bit $i = 0$:
$a_0 = b_0 \oplus c_0$, $carry_1 = b_0 * c_0$,
pro bit $i = 1, ..., w - 2$:
$s_i = c_i \oplus carry_i$, $t_i = b_i * s_i$, $u_i = c_i * carry_i$, $a_i = b_i \oplus s_i$, $carry_{i+1} = t_i \oplus u_i$,
pro bit $i = w - 1$:
$v = b_{w-1} \oplus c_{w-1}$, $a_{w-1} = v \oplus carry_{w-1}$.

Máme zde
$(1 + (w - 2)* 3 + 2)$ operací $\oplus$
a
$(1 + (w - 2)*2 + 0)$ operací *
s
$2 + (w - 2)*(1 + 1 + 1 + 1 + 1) + 1*(1 + 1)$ novými meziproměnnými (a_i, $carry_i$, s_i, t_i, u_i, v), tj. máme $3w - 3$ operací $\oplus$, $2w – 3$ operací * a $5w - 6$ nových meziproměnných.

Další optimalizace

Existuje ještě lepší optimalizace [P]. Výraz $carry_{i+1} = b_i*c_i \oplus b_i*carry_i \oplus c_i*carry_i$ můžeme zapsat pomocí jedné operace AND, a to takto:
$carry_{i+1} = (b_i \oplus carry_i) * (c_i \oplus carry_i) \oplus carry_i$.

Nyní máme o jednu operaci $\oplus$ navíc, ale ušetřili jsme jednu meziproměnnou (u) a jednu AND operaci, což je ještě lepší optimalizace:
$s_i = c_i \oplus carry_i$,
$t_i = b_i \oplus carry_i$,
$carry_{i+1} = s_i * t_i \oplus carry_i$,

Sčítání slov a = b + c můžeme zapsat normalizovaně elementárními operacemi takto:
pro bit i = 0:
$a_0 = b_0 \oplus c_0$, $carry_1 = b_0 * c_0$,
pro bit i = 1, ..., w - 2:
$s_i = c_i \oplus carry_i$, $t_i = b_i \oplus carry_i$, $carry_{i+1} = s_i * t_i \oplus carry_i$, $a_i = b_i \oplus s_i$,
pro bit i = w - 1:
$v = b_{w-1} \oplus c_{w-1}$, $a_{w-1} = v \oplus carry_{w-1}$.

Máme zde
(1 + (w - 2)* 4 + 2) operací $\oplus$
a
(1 + (w - 2)*1 + 0) operací *
s
2 + (w - 2)*(1 + 1 + 1 + 1) + 1*(1 + 1) novými meziproměnnými (a_i, $carry_i$, s_i, t_i, v), tj.
máme 4w - 5 operací $\oplus$, w – 1 operací * a 4w - 4 nových meziproměnných.

Pokud bychom při zápisu řádku
$a_i = b_i \oplus c_i \oplus carry_i$, $carry_{i+1} = b_i * c_i \oplus b_i * carry_i \oplus c_i * carry_i$
do elementárních operací postupovali bez optimalizace, zápis by vyšel místo pěti na sedm elementárních operací. Při zápisu všech funkcí v každém schématu záleží velmi na tom, zda podobnou možnost optimalizace objevíme. Pokud ne, a optimalizace bude přesto existovat, dostaneme horní odhad skutečné reálné složitosti zápisu. Pokud ano, přiblížíme se více reálné složitosti. Avšak i horní odhad složitosti je dobrým výsledkem, neboť pokud vyjde malý, je jisté, že příslušná funkce je jednoduchá.

Normalizace úlohy

Nechceme poškodit žádného z kandidátů na SHA-3, avšak žádná míra není dokonalá, proto námi navržená míra pravděpodobně některého kandidáta upřednostní a jiného poškodí. Cílem je ale spíše uvést tuto míru obecně, nikoli konkrétní variantu.

Také při rozhodování, jakou úlohu měřit, jestli úlohu nalezení kolize nebo vzoru, narazíme na různá dilemata. Tyto úlohy mohou jako řešení mít zprávy o milionech bloků, musíme tedy jak úlohu samu, tak délku zprávy nějak normalizovat. Proto nebudeme formulovat ani úlohu získání kolize ani úlohu získání vzoru, ale srovnáme složitost soustavy Booleovských rovnic, které popisují vznik hašovacího kódu co nejmenším počtem rovnic, proměnných a operací. Tomu přizpůsobíme i délku zprávy. Volíme ji takovou, aby měla maximální délku, ale takovou, aby vyžadovala co nejméně operací, což je tolik, tolik je potřeba na zpracování

zprávy prázdné. Konkrétně u SHA-1 by to byla zpráva délky 512 – 65 bitů, vyžadující jedno volání kompresní funkce. U SHA-256 by to také byla zpráva délky 512 – 65 bitů, u BMW256 by to shodou okolností také byla zpráva délky 512 - 65 bitů. U BMW512 by to byla zpráva délky 1024 – 65 bitů, u SHA-512 také 1024 – 65 bitů. Všechny tyto zprávy vyžadují stejné volání kompresní funkce jako zpráva o délce 0 bitů (prázdný řetězec). Pochopitelně budeme porovnávat složitost kandidátů pro stejný hašový výstup, tedy například pro varianty s 256 bitovým hašovým kódem.

Hašovací funkce typu XAR, XAS

Tito kandidáti na SHA-3, kam patří například BMW a Skein, používají pouze operace XOR, ADD a ROT (SHIFT). U těchto funkcí musíme vypočítat pouze složitost sčítání, operace rotací nebo posunů nevnáší do soustavy rovnic žádné nové meziproměnné ani operace (shift se neprovede, pouze se v dané rovnici, která se shiftovanými proměnnými pracuje, použije proměnná, odpovídající danému posunu).

Ostatní funkce

Všechny ostatní funkce, které se v kandidátských algoritmech objevují, rozložíme na jednodušší operace, až dospějeme k základním bitovým funkcím typu $y_i = f(a_i, b_i,...)$, které vyjádříme v algebraické normální formě. Měli bychom se rozhodně zabývat tím, jak tyto funkce optimalizovat. Protože to může být značně náročná úloha, pro první přiblížení můžeme optimalizovat jen zjevné výpočty.

Přesnost naší míry

Domníváme se, že toto je velmi dobrá míra složitosti (nikoli bezpečnosti), protože když je schéma uděláno dobře, není v něm žádná zkratka, nevíme, jak řešit soustavu rovnic, neznáme slabiny schématu, a proto řešení námi zapsané soustavy rovnic je jediná cesta, jak prolomit schéma. Pokud předpokládáme, že všichni kandidáti na SHA-3 jsou uděláni dobře, je řešení systému normalizovaných rovnic u každého odrazem jeho skutečné bezpečnosti. A protože tento systém je normalizovaný pro všechna schémata stejně, lze jeho složitostí porovnávat bezpečnost schémat.

Závěr

Záměrně zde nevypočítáváme složitost kandidátů podle navržené metody a ponecháváme to na zájemcích, studentech. Porovnání je velmi potřebné a tato míra je pro některé kandidáty docela jednoduše spočitatelná, pro jiné postačí k velmi potřebnému výsledku velmi málo.

[P] Michael Fischer, Rene Peralta: Counting Predicates of Conjunctive Complexity One YALEU/DCS/TR-1222, December 2001 (Revised February 2002), Yale University Department of Computer Science

B. Rozhovor: Šifry mistrů kryptologů

Rozhovor připravil: Lukáš Soukup , Hosté: Vlastimil Klíma a Pavel Vondruška

Tento rozhovor byl pořízen jako upoutávka k diskusnímu pořadu Dobrodružství kryptologie, který v rámci svých pravidelných setkání uspořádalo Science Café v úterý 9. února 2010 v tradičním místě pražských setkání v kavárně Potrvá (Srbská 2, Praha 6) a byl s laskavým svolením organizátorů akce převzat ze stránky projektu Science Café.

Nejprve, co to vlastně Science Café je:

Science Café jsou setkávání vědců a veřejnosti v prostředí kaváren a restaurací měst České republiky. Večery zasvěcené společnému hledání otázek a odpovědí na předem domluvené téma dávají návštěvníkům možnost se v neformálním prostředí zeptat na to, co je zajímá. Více informací a hlavně program plánovaných akcí naleznete na domácí stránce tohoto zajímavého projektu http://sciencecafe.cz/;

Rozhovor

Tajná písma, šifry a jejich luštění jsou inspirací pro řadu populárních knih a filmů, které baví čtenáře a diváky po celém světě. Vytváření šifer, respektive jejich luštění může být ale pro někoho jeho každodenní prací. O kryptologii jako nauce o metodách utajování informací a o tom, jaké to je být kryptologem jsme hovořili s předními českými kryptology-profesionály Vlastimilem Klímou a Pavlem Vondruškou.

Jak byste definovali kryptologii?

Pavel Vondruška (PV): Kryptologie je aplikovaná věda, která využívá poznatky z řady oborů – zejména z matematiky, statistiky, ale i z fyziky nebo chemie.

Vlastimil Klíma (VK): Dodal bych, že kryptologie je také považována za umění a někdy za šarlatánství, v kryptologii, a to i moderní, totiž spoléháme na tvrzení, která se nedají dokázat.

Dá se říci, kdy kryptologie vznikla?

VK: V zásadě platí, že s každým druhem komunikace se objevil nějaký druh šifrování. Nejstarší známé šifry pocházejí z roku 1900 př. n. l., nicméně potřeba šifrování vznikla už v okamžiku, kdy se rozvinulo písmo. Brzy začali lidé vyvíjet znaky, kterými by nahradili znaky dosud známé. Často se to používalo v náboženských textech. A s každým druhem komunikace se vždy objevil nový druh šifrování, příchod rádia, televize nebo počítačů vždy znamenal pro kryptologii nový impuls.

PV: Je to málo známé, ale utajování – šifrování nebylo opravdu nikdy nutně vázáno jen na písmo, ale již od starověku se používalo i v přímé řeči. Kódování v řeči není jen o tom známém „pět na stole v českých“, ale byly zde celé systémy, které za tímto účelem byly vyvíjeny. Víme o tom, že obchodníci v Číně nebo Fénicii si vymýšleli

speciální řeč, aby spolu mohli komunikovat v průběhu obchodního jednání. Zmínky nalezneme také v Kamásutře, kde se doporučovalo, aby žena uměla tajnou řeč, kterou může používat v komunikaci se svým milencem na veřejnosti.

Jak se člověk stane kryptologem?

VK: Kryptologem se můžete stát, když budete mít teoretické základy z matematiky, důležitá je ale zejména praxe a zkušenost. Já se zabývám kryptologií 27 let. Po vysoké škole jsem nastoupil do ozbrojených složek, později jsem pracoval v soukromém sektoru a teď se kryptologii věnuji na volné noze. Slovy klasika, vyučil jsem se kryptologem a jsem kryptologem.

PV: Šifrování mne lákalo od dětství. Profesně pak byla u mne cesta podobná, pracoval jsem rovněž jako kryptolog ve státní sféře a později jsem se v Ústavu pro ochranu osobních údajů zabýval elektronickým podpisem. V současnosti pracuji v soukromém sektoru a kryptologii přednáším na MFF UK.

V souvislosti s kryptologií se používá zejména dvou pojmů, a to kódování a šifrování. Jaký je mezi nimi v kryptologii rozdíl?

VK: Rozdíl mezi kódováním a šifrováním je v existenci klíče. Kódování nepoužívá žádnou tajnou informaci, zatímco šifrování ano – k šifrování potřebujete takzvaný klíč. V dnešní době může být dokonce jeden z klíčů veřejný a jeden tajný, stále tak lze sdělení zašifrovat. Příkladem kódu je například morseovka, což je přepis jedné abecedy do druhé. Příkladem šifrování je třeba Caesarova šifra, kdy se písmeno posune o tajný počet pozic v abecedě. Caesar použil jako to tajné číslo tři, Augustus jedničku.

Jak se vyvíjely klíče k šifrování?

PV: U nejstarších šifer byla takovým klíčem převodová tabulka. To byl ovšem systém, který nepřítel mohl na základě frekvenční analýzy luštit, a tak se začaly hledat dokonalejší systémy. Šifry se začaly vylepšovat, lidé do

nich vkládali například klamače – písmena, která nemají smysl a stěžují analýzu šifer. Toto vše, co se do sdělení vkládá, či co je nahrazováno, je základ klíče. Jedním z dokonalejších systémů je tzv. polyalfabetická šifra, kdy je v každém kroku pro náhradu písmene otevřeného textu použita jiná převodová tabulka, výběr tabulky se řídí klíčem. Tato šifra měla zpočátku punc nerozluštitelnosti, ale nakonec i ta byla rozluštěna… Dlouho se diskutovala otázka, zda algoritmus tajit nebo zda jej zveřejnit. U dobrého systému by měla být bezpečnost zajištěna pouze utajením klíče. Dnes se algoritmy-šifry zveřejňují a procházejí veřejnou analýzou, stačí proto utajovat pouze klíče.

Jaké jsou tedy základní metody šifrování?

VK: Základní metody jsou tři: substituce, transpozice a přičítání. I když se to zdá málo, přesto to dostačuje – vždyť počítače také pracují jen se dvěma prvky, nulou a jedničkou. Přelomem v šifrování byla chvíle, kdy se od metod oddělila informace. Dobře to ilustruje příklad z první světové války: vojáci se naučili jednu metodu a pak jen dostávali odlišné klíče. Distribuce klíčů pak měla ve válkách rozhodující roli. Například Rusko prohrálo řadu válek špatnou nebo opožděnou distribucí klíčů. V bitvách padly tisíce vojáků jen proto, že dostali staré klíče. Protivník je již znal, rozluštil radiové zprávy a znal přesuny vojsk. Podobně tomu bylo v Japonsku, kde se klíče těžko distribuovaly na jednotlivé ostrovy, to se stalo osudné například generálu Yamamotovi. Dalším příkladem může být rozluštění Enigmy, které zkrátilo druhou světovou válku asi o rok.

PV: Mimochodem Československá lidová armáda používala ke svému spojení Enigmy až do poloviny padesátých let.

Existují nějaké nerozluštitelné šifry?

PV: Ano, absolutně bezpečné šifry existují a jsou algoritmicky velmi jednoduché. Používaly se například ke spojení mezi USA a Sovětským svazem během studené války.

VK: Jednou z klasických nerozluštěných šifer a záhad je kniha známá také jako Voynichův rukopis, která pochází ze 17. století a která dodnes nebyla rozluštěna.

PV: Tato kniha skutečně odolává sofistikovaným pokusům o rozluštění. Může jít ale o knihu nesmyslných znaků. O rukopisu se poprvé mluví v souvislosti s jeho prodejem císaři Rudolfu II. Dodnes nevíme, kdo ji napsal, jediné co lze říci je, že připomíná alchymistické knihy z té doby.

Kdo byli nebo jsou pro vás hrdinové kryptologie?

VK: Zaujala mě epizoda z francouzské historie, kdy Georges Painvin zachránil v létě 1918 Paříž díky rozluštění klíče k německé šifře. Pak na mě udělal dojem Charles Babbage, který začal s programovatelným počítačem, protože chtěl luštit šifry. Dále pak Turing a lidé z týmu v Bletchley Parku. Hrdinou je pro mě i historik David Kahn, který v roce 1967 vydal první knihu o historii kryptologie, která byla vysoce odborná a zároveň pravdivá z hlediska historie. Z těch novodobých bych zmínil Merckleho, který je spoluobjevitelem myšlenky veřejného klíče. Hrdinou je pro mě i americký úřad pro standardizaci, který přišel jako první s tím, že udělá soutěž na veřejnou šifru a v roce 1977 byla zveřejněna první bloková šifra pro komerční použití, která se stala nejpoužívanější šifrou v historii.

PV: Všechny osobnosti, které jmenoval Vlastimil, jsou hrdinové i pro mě, existuje ale také mnoho bezejmenných hrdinů kryptologie, o kterých se asi nikdy nedozvíme, protože pracovali v době, kdy tato činnost byla utajená. Před rokem 1989 se o kryptologii téměř mlčelo, mimo jiné proto, protože kryptologie byla vnímána jako zbraň.

Jaké jsou vaše osobní plány v souvislosti s kryptologií?

VK: Chystám se na Voyniche, koneckonců s Pavlem jsme již začali.

PV: Já se chystám na zmapování dějin české kryptologie od třicátých let, kdy v zásadě vznikla, až do současnosti.

Za Science Café se ptal Lukáš Soukup

Fotograie z tohoto setkání lze najít na facebooku:

http://www.facebook.com/album.php?aid=154895&id=78321222125

C. Index koincidence

Lucie Fojtová, lucie@fojtova.cz, Ústav telekomunikací FEKT VUT v Brně

Statistické metody publikované americkým kryptologem Williamem Freidmanem r. 1920 ve sbírce *The Index of Coincidence and Its Applications to Cryptography* zcela pozměnily tehdejší možnosti kryptoanalýzy šifer. Svého času byla tato kniha dokonce označována jako nejvýznamnější počin v oblasti kryptografie a přinášela do tohoto odvětví vyšší matematiku, statistiku a pravděpodobnost. Tento článek popisuje některé ze statistických testů a jejich využití při kryptoanalýze klasických šifer.

William F. Friedman (1891 – 1969)

Luštitel klasických šifer svůj úspěch obvykle může znásobit použitím některého ze statických testů, které pomáhají učit typ luštěné šifry. Prvotní analýza textu může často ušetřit spoustu marných pokusů.

Index koincidence

Index koincidence (dále jen IC) je míra relativní četnosti písmen v (šifrovém) textu. Tento statistický test může výrazně usnadnit kryptoanalýzu klasických šifer – zejména pak polyalfabetické. Pro určení délky klíče této šifry se dá použít Kasiského metoda nebo jako alternativa právě odhad pomocí IC.

IC je tedy definován jako pravděpodobnost, že dva znaky náhodně vybrané z šifrového textu budou stejné. Tato pravděpodobnost se vyjádří následujícím vztahem:

$$IC = \sum_{i}^{c} \frac{n_i \cdot (n_i - 1)}{N \cdot (N - 1)}, \qquad (1)$$

kde N znamená celkový počet znaků testovaného textu, ze kterého je IC počítán, c označuje počet znaků abecedy (např. česká abeceda bez speciálních znaků jich má 26) a n_i je počet znaků s indexem i, kde $i \in \{0,1..25\}$ - tedy např. pro mezinárodní abecedu je n_1 počet všech

znaků **a** v celém textu, n_2 počet všech znaků **b** atd. Čitatel zlomku vyjadřuje počet dvojic (tzv. bigramů) písmen obsahujících dva stejné znaky, jmenovatel vyjadřuje počet všech bigramů v analyzovaném textu.

Z uvedeného vzorce vyplývá, že čím větší jsou rozdíly v četnosti výskytu jednotlivých písmen u daného jazyka, tím větší je jeho index koincidence a naopak - v případě analýzy náhodně vygenerovaného sledu znaků je obvykle IC znatelně menší (blíží se hodnotě 0,0385).
Každý jazyk je charakteristický svým indexem koincidence, který hodnotí kolísání četností znaků textu. Pro náhodný text je IC vypočítán jako hodnota 1/26, což vyplývá z následujícího: uvažujeme-li mezinárodní abecedu (26 znaků), pravděpodobnost náhodného výběru jednoho z nich je $P(znak) = 1/26 = 0{,}0385$. Tato hodnota zároveň vyjadřuje velikost pravděpodobnosti (označovaná jako κ_r), že bude náhodně vybrán identický pár:

$$\kappa_r = \frac{1}{26}\cdot\frac{1}{26} + \frac{1}{26}\cdot\frac{1}{26} + ... + \frac{1}{26}\cdot\frac{1}{26} = 26\cdot\frac{1}{26^2} = \frac{1}{26} = 0{,}0385 .$$

Jinými slovy, na každých 100 dvojic písmen připadají necelé čtyři shody znaků. Tato hodnota pravděpodobnosti však platí pouze v případě, kdy písmena jsou rozložena naprosto pravidelně, tj. žádné z nich nemá vyšší četnost výskytu než ostatní. Pokud však vezmeme v úvahu charakteristické rysy konkrétního jazyka - např. rozložení hlásek (mezinárodní abeceda) v českém jazyce (viz tabulka č.1), pravděpodobnost se změní následujícím způsobem:

$$\kappa_p = \sum_{i=1}^{26} p_i^2 = (0{,}0865)^2 + (0{,}0163)^2 + (0{,}0355)^2 + (0{,}0356)^2 + (0{,}1029)^2 + (0{,}0038)^2 + (0{,}0034)^2 +$$
$$+ (0{,}0226)^2 + (0{,}0752)^2 + (0{,}0194)^2 + (0{,}0368)^2 + (0{,}0402)^2 + (0{,}032)^2 + (0{,}0662)^2 + (0{,}0815)^2 +$$
$$+ (0{,}0339)^2 + (0{,}00)^2 + (0{,}051)^2 + (0{,}0000)^2 + (0{,}0506)^2 + (0{,}0533)^2 + (0{,}0548)^2 + (0{,}0377)^2 +$$
$$+ (0{,}043)^2 + (0{,}0007)^2 + (0{,}0009)^2 + (3{,}44)^2 = 0{,}0569$$

kde p_i je pravděpodobnost výskytu i-tého znaku v abecedě, tj. na každých 100 dvojic znaků českého textu připadá něco mezi pěti a šesti shodami. Tato hodnota je tedy charakteristická pro každý jazyk.
Známe-li tedy četnost rozložení písmen v daném jazyce, můžeme pomocí zobecněné rovnice

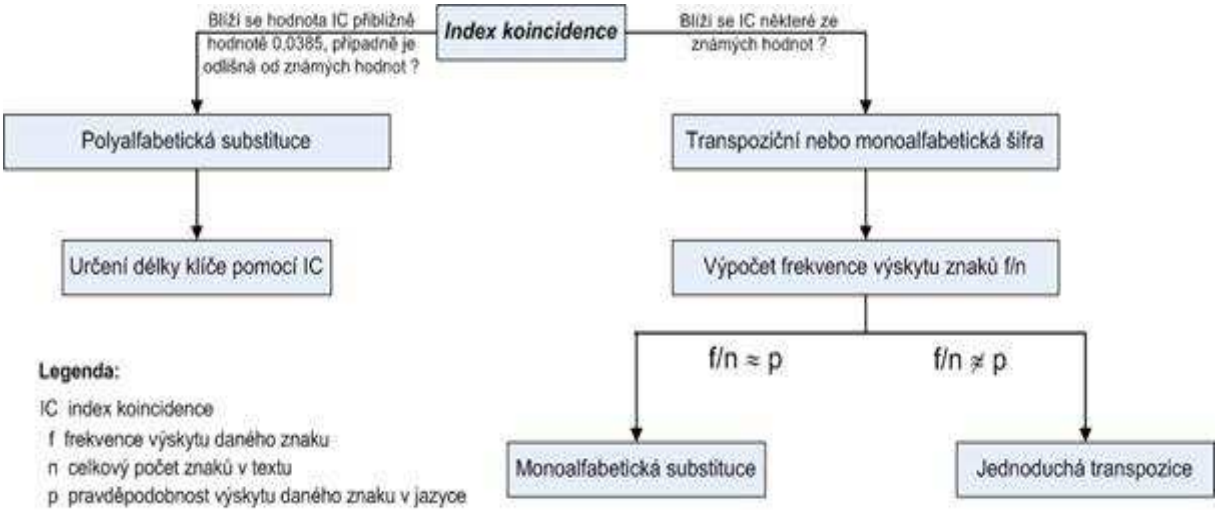
$$\kappa_p = \sum_{i=1}^{N} p_i^2 , \qquad (2)$$

kde N je počet znaků abecedy, určit očekávaný *index koincidence příslušného jazyka*. Hodnota κ pro náhodný text $N = 26$ je označována jako κ_r. Jedná se o tzv. *kappa – test*, který vyjadřuje index koincidence daného jazyka [1].
Jednotlivé hodnoty κ_p pro různé jazyky uvádí tabulka č. 2 [1,3].

Jazyk	κ_p
Francouzština	0,0778
Španělština	0,0775
Němčina	0,0762
Italština	0,0738
Angličtina	0,0667
Čeština	0,0569
Ruština	0,0529
náhodný text	0,0385

Tabulka č.1 Hodnota κ_p

Určení typu šifry
Jak tedy postupovat v případě obdržení zašifrovaného textu? Následující obrázek znázorňuje využití indexu koincidence pro určení typu šifry.



Obr.č.1 Využití indexu koincidence

Monoalfabetická substituce a jednoduchá transpozice zachovávají rozložení četnosti znaků (tyto šifrovací systémy tedy nemění index koincidence). V případě, že se vypočítaný IC šifrového textu (dle (1)) blíží některé ze známých hodnot (viz tabulka č. 1), je pravděpodobné, že se bude jednat právě o jeden z těchto šifrových systémů s otevřeným textem v příslušném jazyce.

Pro rozlišení, zda se jedná o substituci nebo transpozici, lze využít metodu frekvenční analýzy: stačí spočítat rozložení výskytu jednotlivých znaků v šifrovém textu. Pokud je toto rozložení shodné s obecným rozložením výskytu znaků v jazyce určeném pomocí IC, jedná se pravděpodobně o transpozici, naopak pokud jsou pravděpodobnosti pro totožné znaky odlišné, jde o substituci. Tabulka č. 2 uvádí frekvenční rozložení písmen v českém jazyce [2].

A	0,086533	**h**	0,022573	**o**	0,081516558	**v**	0,042923
b	0,016323	**i**	0,075225	**p**	0,033857769	**w**	0,000705
c	0,035533	**j**	0,019439	**q**	5,65067E-05	**x**	0,000907
d	0,035606	**k**	0,036781	**r**	0,050622637	**y**	0,026406
e	0,102933	**l**	0,040161	**s**	0,053296614	**z**	0,030831
f	0,003863	**m**	0,031982	**t**	0,054825413		
g	0,003361	**n**	0,066169	**u**	0,037699489		

Tabulka č. 2 Frekvence českých písmen

Naopak polyalfabetická substituce vyhlazuje rozdíly mezi četností znaků v dané abecedě, to znamená, že IC pro takový šifrový text bude nižší než příklady uvedené v tabulce, pravděpodobněji se bude blížit hodnotě IC pro náhodný text (tj. 0,0385).
Čím delší šifrový text je k dispozici, tím lépe bude IC fungovat - samozřejmě pouze v případě, že nepůjde o otevřený text záměrně napsaný tak, aby nesplňoval základní statistické

charakteristiky příslušné danému jazyku. Za hranici použitelnosti se obvykle považuje text o délce 100 znaků [3].

V případě polyalfabetické substituce je prvním krokem určení délky periody (klíče). Pro luštění šifry Vigénere se používá tzv. Kasiského metoda, vyvinutá Friedrichem Kasiskim (a nezávisle na něm také Charlesem Babbagem). Máme-li k dispozici dostatečně dlouhý text, začneme v něm hledat opakující se polygramy (čím delší, tím lepší – začínáme obvykle hledat bigramy). Pro jednotlivé opakující se polygramy se určí jejich vzájemná vzdálenost. Skutečná délka klíče pak bude s největší pravděpodobností největším společným dělitelem těchto vzdáleností jednotlivých dvojic shodných polygramů. Přitom vycházíme z předpokladu, že opakovaný výskyt stejných polygramů není náhodný a odpovídá stejnému otevřenému textu – tedy vzdálenost těchto polygrafů je rovna násobku délky klíče.

Při tomto postupu je třeba mít na vědomí, že opakované polygramy (nejčastěji bigramy) mohou v šifrovém textu vzniknout i náhodně, a ne všechny vzdálenosti opakovaných polygramů musí být násobkem délky klíče (delších polygramů se to v drtivé většině případů netýká).
Variantně se však můžeme pokusit o hrubý odhad délky klíče za pomocí očekávané hodnoty IC:

$$E(IC) = \frac{1}{t} \cdot \frac{L-t}{L-1} \cdot \kappa_p + \frac{t-1}{t} \cdot \frac{L}{L-1} \cdot \kappa_r , \qquad (3)$$

kde L je délka šifrového textu, t je délka klíče, κ_p a κ_r hodnoty z tabulky č. 2. Pro jednotlivé periody $(t = 1,2,...,20,...50..)$ dosazujeme do vzorce (3) a výsledek porovnáváme s hodnotou IC šifrového textu vypočítaného dle vzorce (1). Hodnota t, pro kterou se *E (IC)* nejvíce blíží IC šifrového textu, přibližně odpovídá délce klíče.

V tuto chvíli již má luštitel určen typ šifry a zbývá ji „pouze“ vyluštit. O metodách luštění klasických šifer se případný zájemce více dozví např. v [4], nebo ve starších číslech e-zinu Crypto-World.

Literatura:

[1] Menezes, Alfred J.; van Oorschot, Paul; Vanstone, Scott A.: Handbook of Apllied Cryptography

[2] Statistické charakteristiky češtiny - http://nlp.fi.muni.cz/cs/stat_cestiny

[3] Elektronické texty doc. Jiřího Tůmy pro výuku MFF, http://www.karlin.mff.cuni.cz/~tuma/nciphers10.html

[4] Vondruška, P.: Kryptologie, šifrování a tajná písma.

D. Chcete si zaluštit? Díl 4.

Martin Kolařík (marram.mail@gmail.com)

Dubnová dávka luštění.

Dostala se mi zpětná vazba od čtenářů, že jim u luštění chybí možnost dozvědět se správné řešení. Bohužel bych tím kazil smysl hry Geocaching a hlavně tím asi nepotěšil autory těchto pěkných šifer. „Kačeři“ jsou obvykle schopní hledači informací, takže by bylo velké riziko, že by takto získali některá řešení a právě o to autoři určitě nestojí.

Nicméně ani mezi „kačery“ nejsou všeználci, takže čas od času potřebuje každý s něčím poradit, proto vám nabízím možnost mi napsat žádost o pomoc mailem, pokusím se u šifer napovědět. Ovšem hlavním smyslem těchto článků je nabídnout zábavu, ne vysvětlovat způsoby luštění, proto ode mě neočekávejte kompletní řešení.

Těším se na vaše dotazy a nyní již slíbené **dubnové GeoŠifry**, tentokrát jsem vybral spíše zábavné způsoby ukrytí informace, berte ohled na skutečnost, že cílem jsou čísla souřadnic:

DENIK (http://coord.info/GC1PMPW)

Stara Posta (http://coord.info/GC1BRQC)

A : 26 12623456 2142653543

B : 1265431346 2165643134 31436453651 2

Poslední není šifra, ale způsob „ukrytí“ souřadnic je velmi pěkný:

Jezte tatarak! (http://coord.info/GC22X1W)

Přeji úspěšné luštění a šťastný lov.

Martin

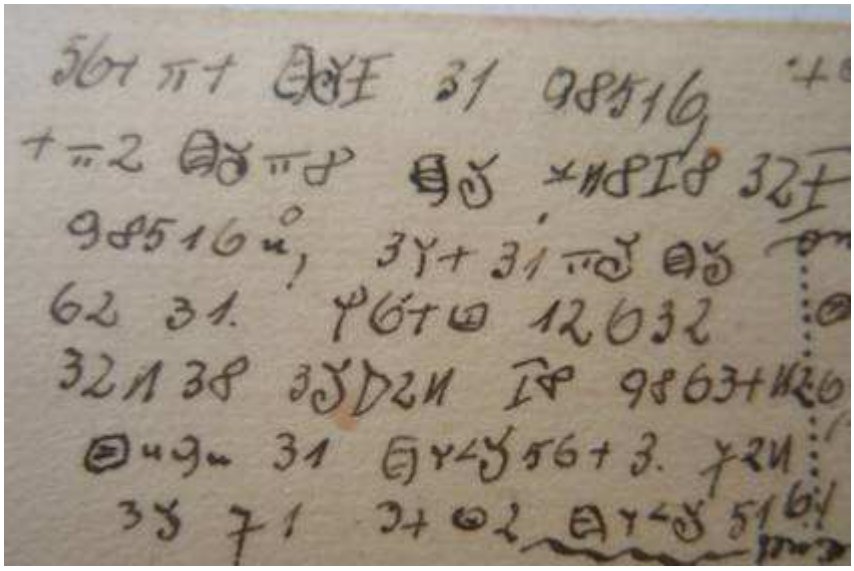
E. Tajemství ukryté v 11-ti pohlednicích (upoutávka)
Pavel Vondruška - Marta Janošová

Zcela náhodou se mi dostal tento měsíc do rukou časopis ***Malovaný kraj*** a v něm jsem našel článek, který mne hned z několika důvodů velmi zaujal. Nebyl to článek pro článek, se kterými se v poslední době často setkávám, ale naopak článek, na němž je vidět, že autorka před jeho sepsáním udělala obrovský kus práce a hlavně že tuto práci dělala s velkým elánem a nadšením. Protože v tomto článku mají své místo i laické šifry používané během první republiky, rozhodl jsem se autorku kontaktovat a požádat ji o článek do našeho e-zinu.

Myslím si, že příběh, který na základě několika pohlednic s šiframi sestavila, bude kladně přijat i vámi – tedy čtenáři našeho elektronického časopisu.

Marta Janošová, autorka článku, mi přislíbila spolupráci a domluvili jsme se, že článek, který podle svých časových možností připraví, bude publikován v některém z příštích čísel našeho e-zinu. Také jsme se dohodli, že jako malou upoutávku na tento článek zveřejníme následující text a

otiskneme některé z pohlednic se šiframi, které získala. Dále následuje pár úvodních vysvětlujících slov autorky.



Ráda bych se s vámi podělila o svůj zážitek, který jsem před nějakou dobou prožila v jednom malinkém antikvariátě.

Při prohlížení pohledů mě z nich zaujalo hned několik, které si mezi sebou psali mladí milenci v letech 1929 až 1931, zajímavé na nich bylo to, že **byly psány šiframi**. Ihned jsem zbystřila a začala jsem se dívat, zda jich nenajdu více.

Našla jsem, celkem se mi podařilo objevit jedenáct pohledů, které si tito mladí lidé psali. Bohužel jich v antikvariátě nebylo víc, protože někdo byl pravděpodobně přede mnou rychlejší a koupil si je do své sbírky. I přesto jsem je koupila, protože zvědavost mi nedala a chtěla jsem se pokusit o jejich rozluštění.

Rozluštění nebylo příliš složité, protože se znaky teprve učili a vymýšleli.
Slečna se jmenovala Josefína Halmová a bydlela v Boleradicích (vesnice mezi Hustopečemi a Klobouky u Brna) a její milý byl František Kaisler - vojín, který se účastnil kursu specialistů na vojenském leteckém učilišti v Prostějově.

F. Telegramy pořád žijí

Vladimír David, Informační a asistenční služby, Telefónica O2 Czech Republic

Následující informace byla převzata z intranetu Telefónica O2 Czech Republic, kde byla publikována 6.4.2010

Mnozí z vás si jistě všimli, že došlo v posledních dnech k mediální kampani ze strany České pošty o ukončení telegrafních služeb. Tyto informace jsou trochu zkreslené.

Telefónica O2 jako nástupník státního podniku České telekomunikace je jediným poskytovatelem telegrafní služby na území dnešní ČR a tuto službu provozujeme i nadále. Pouze nelze od 1. 4. 2010 podávat telegram na přepážce pošty.

Pro ty, kteří nevědí, co to TELEGRAM vlastně je. Je to textová zpráva, kterou jsme schopni garantovaně, expresně a pohodlně předat adresátovi a může ji v podstatě využít kdokoliv. Našimi nejvýznamnějšími zákazníky jsou především velké firmy, instituce, nemocnice, soudy, státní správa, finanční úřady apod., které musí bezodkladně informovat své klienty o důležitých skutečnostech. Telegraf je tradiční služba, která se neopírá pouze o listovní zásilku, ale využívá moderní technologie a přidanou hodnotu informačních systémů a lidské energie k efektivnímu předání důležité zprávy.

I vy nám můžete pomoci zachovat tuto službu i nadále, aby nyní nepřicházela o své zákazníky a sloužila jim i do budoucna. Chceme vás požádat o podporu naší jedné z nejstarších služeb. I vy můžete sdělit svým blízkým, známým či obchodním partnerům, že služba neskončila a stále pokračuje.

Můžete také vysvětlit, jak služba funguje. Službu v současnosti poskytuje speciální pracoviště v Praze, 24 hodin denně, 7 dní v týdnu. Měsíčně zpracují kvalifikované operátorky cca 4000 ks telegramů adresovaných jak do České republiky, tak i do zahraničí. Telegram je doručován následující pracovní den.

Telegram může být podán následujícími způsoby:

- telefonem (z pevné linky): na bezplatné lince 133 001 (kde se také zákazník dozví podmínky),
- faxem: na zpoplatněné lince 271 775 003 (požadavek na zaslání telegramu musí být v předepsané formě),
- e-mailem na adrese: telegrafni.sluzby@o2.com (požadavek na zaslání telegramu musí být v předepsané formě).

Jaká je cena za podání telegramu po České republice?

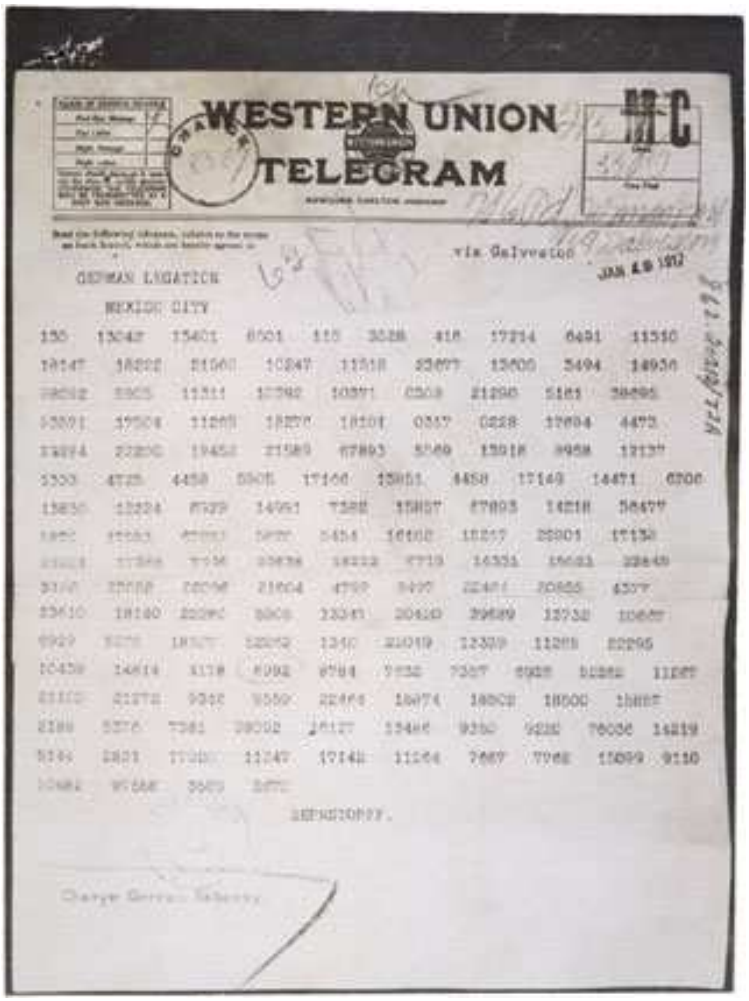
Cena se pohybuje od 130 Kč výše. Cena se skládá u vnitrostátního telegramu ze základního poplatku 80 Kč a poplatku za slovo 5 Kč (minimum slov musí být 10 - počítáno i s adresou). Dále je možné objednat doplňkové služby jako potvrzení o dodání telegramu, ověřený opis telegramu atp. U mezinárodních telegramů se cena slova řídí podle ceníku do dané země.

Na závěr si připomeneme, že šifrové telegramy sehrály důležitou úlohu v diplomatické korespondenci. Využijeme k tomu dvě drobné ukázky z knihy Pavla Vondrušky: Kryptologie, šifrování a tajná písma, která vyšla v roce 2006 v edici OKO (nakladatelství Albatros).

1904

1. července vstoupily v platnost v Anglii nové telegrafní předpisy. Mimo jiné měly za následek, že šifrový text byl předáván ve skupinách po pěti znacích. Délka byla stanovena podle průměrné délky přenášeného slova. Pro čtenáře, kteří si již nepamatují posílání telegramů z pošty, připomínám, že se platil poplatek za každé odvysílané slovo. U šifrových telegramů se tedy mohlo stát, že při vysílání dlouhých šifrových slov byl poplatek za telegram výrazně nižší, než kdyby se předával v otevřené podobě. Tato úprava tomu samozřejmě rozumným způsobem zabránila. Zvyk dělit šifrový text do skupin po pěti znacích, i když se nejednalo o telegrafní přenos, se zachoval po celé dvacáté století a dodnes se s ním můžeme setkat při různých příležitostech (popisy šifrových systémů, soutěže v luštění, přepis starých šifrových telegramů).

1917


WESTERN UNION
TELEGRAM
via Galveston
GERMAN LEGATION
MEXICO CITY

Vstup USA do války byl důsledkem vyluštění obsahu šifrového telegramu - dnes známého jako tzv. Zimmermannův telegram. Německý ministr zahraničí Arthur Zimmermann (1864 – 1940) v telegramu zaslaném mexické vládě vyzývá Mexiko k válce proti USA. Slibuje v ní mexické straně podporu a územní zisk. Britové telegram zachytili, rozluštili jej a předali USA (přičemž neprozradili svůj zdroj). Po té, co se prezident Woodrow Wilson (1856–1924) s obsahem telegramu seznámil, svolal Kongres. Kongres 2. 4. 1917 schválil vstup USA do války proti Německu. Tento akt rozhodujícím způsobem změnil poměr sil na evropském bojišti.

G. O čem jsme psali v dubnu 2000 – 2009

Crypto-World 4/2000

Crypto-World 4/2001

Crypto-World 4/2002

Crypto-World 4/2003

Crypto-World 4/2004

Crypto-World 4/2005

Crypto-World 4/2006

Crypto-World 4/2007

Crypto-World 4/2008

Crypto-World 4/2009

H. Závěrečné informace

1. Sešit

Crypto-World je oficiální informační sešit "Kryptologické sekce Jednoty českých matematiků a fyziků" (GCUCMP). Obsahuje články podepsané autory. Případné chyby a nepřesnosti jsou dílem autorů jednotlivých podepsaných článků, GCUCMP za ně nemá odbornou ani jinou zodpovědnost.
Adresa URL, na níž můžete najít tento sešit (zpravidla 3 týdny po jeho rozeslání) a předchozí e-ziny, denně aktualizované novinky z kryptologie a informační bezpečnosti, normy, standardy, stránky některých členů a další související materiály: http://crypto-world.info

2. Registrace / zrušení registrace

Zájemci o e-zin se mohou zaregistrovat pomocí e-mailu na adrese pavel.vondruska@crypto-world.info (předmět: Crypto-World) nebo použít k odeslání žádosti o registraci elektronický formulář na http://crypto-world.info. Při registraci vyžadujeme pouze jméno a příjmení, titul, pracoviště (není podmínkou) a e-mail adresu určenou k zasílání kódů ke stažení sešitu.
Ke zrušení registrace stačí zaslat krátkou zprávu na e-mail pavel.vondruska@crypto-world.info (předmět: ruším odběr Crypto-Worldu!) nebo opět použít formulář na http://crypto-world.info. Ve zprávě prosím uveďte jméno a příjmení a e-mail adresu, na kterou byly kódy zasílány.

3. Redakce

E-zin Crypto-World

Redakční práce:	Pavel Vondruška
Stálí přispěvatelé:	Vlastimil Klíma
	Pavel Vondruška
Jazyková úprava:	Jakub Vrána
Přehled autorů:	http://crypto-world.info/obsah/autori.pdf
NEWS (výběr příspěvků, komentáře a vkládání na web)	Jaroslav Pinkava Vlastimil Klíma Pavel Vondruška Tomáš Rosa
Webmaster	Pavel Vondruška, jr.

4. Spojení (abecedně)

redakce e-zinu	ezin@crypto-world.info ,	http://crypto-world.info
Vlastimil Klíma	v.klima@volny.cz ,	http://cryptography.hyperlink.cz/
Jaroslav Pinkava	jaroslav.pinkava@gmail.com ,	http://crypto-world.info/pinkava/
Tomáš Rosa	tomas.rosa@rb.cz ,	http://crypto.hyperlink.cz/
Pavel Vondruška	pavel.vondruska@crypto-world.info	http://crypto-world.info/vondruska/index.php
Pavel Vondruška,jr.	pavel@crypto-world.info ,	http://webdesign.crypto-world.info
Jakub Vrána	jakub@vrana.cz ,	http://www.vrana.cz/