Nové výzkumy: vypočtete si kvalitu hesla

V posledních dvou letech došlo v odborné komunitě ke změně postoje vůči heslům. Statistické průzkumy neustále ukazují, že lidé volí hesla slabá. Proto jsme v minulém čísle doporučili volit hesla náhodně a seznam hesel chránit superheslem. Všechny zásady, které se po dlouhou dobu vytvářely pro tvorbu hesel, nyní aplikujeme pouze na tvorbu superhesel.

Ochrana superhesla

Superheslo je heslo, které si snažíme pamatovat za všech okolností a pokud možno ho nikdy nikam nezapisujeme. Nikdy ho nepoužíváme v cizím prostředí, v internetových kavárnách apod. Výjimkou jsou lidé, kteří by si prostě takové heslo nezapamatovali nebo mají strach, že by ho mohli zapomenout. V tom případě nechť si ho někam zapíší, ale musí počítat s větším rizikem jeho prozrazení. Naproti tomu ho mohou volit zcela náhodné.

Nepredikovatelnost superhesla

Kvalitní superheslo musí být co nejvíce nepredikovatelné. Útočník, který nemá možnost jeho predikce, je nucen zkoušet všechny jeho kombinace. Těch musí být takové množství, že je nemůže vyzkoušet žádnými jemu dostupnými prostředky. Za takové číslo můžeme považovat 2^{128}. Pro méně náročné postačí číslo 2^{100}, někomu stačí 2^{80}. Pokud nemáte obavy z tajných služeb, poslední uvedená složitost by vás měla bohatě chránit před všemi ostatními útočníky. Dlouhodobé distribuované výpočty na Internetu zatím dosáhly počet operací pouze cca 2^{64}, takže i tato složitost je pro „běžné použití“ dostatečná. Počtem stejně pravděpodobných možností hesla 2^S, neboli jeho entropií S, se zabýváme podrobněji proto, že z ní vyplyne všechno ostatní. Jestliže uvažujeme, že útočník bude mít výpočetní kapacitu na vyzkoušení 2^S hesel, musíme zvolit heslo s entropií o něco větší než S.

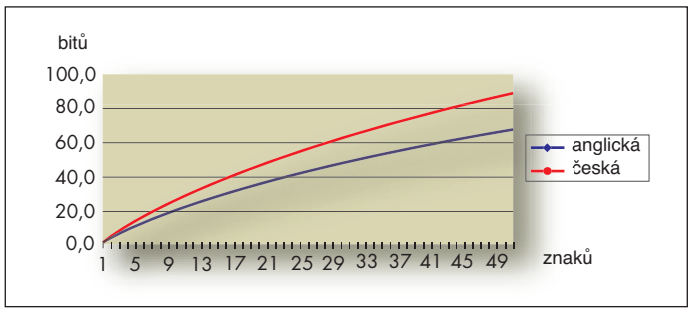
Délka superhesla

Délka superhesla závisí zejména na tom, z jaké znakové sady ho tvoříme. Pokud se znaková sada skládá pouze z číslic a heslo tvoříme zcela náhodně, budeme k dosažení počtu jeho možností 2^S potřebovat x číslic, kde $10^x=2^S$, tj. $x=S*\log_{10}2$. Pro S = 64, 80, 100 a 128 máme (po zaokrouhlení) x 20, 25, 31 a 39. Jistě budete souhlasit, že náhodné číslo o 39 číslicích byste si opravdu nechtěli pamatovat, takže musíme rozšířit znakovou sadu.

Dejme tomu, že použijeme znakovou sadu, která se skládá z malých a velkých písmen, číslic a zvláštních znaků (,.;-_/!?"@#%&*§$\<>:), celkem z 26 + 26 + 10 + 20 = 82 znaků.

Pokud heslo tvoříme náhodně, budeme k dosažení složitosti alespoň 2^S potřebovat x znaků, kde $82^x \geq 2^S$, tj. $x \geq S*\log_{10}2 / \log_{10}82$. Pro S = 64, 80, 100 a 128 máme $x \geq$ 11, 13, 16 a 21.

Je poměrně překvapivé, že i při tak velké znakové sadě potřebujeme na kvalitní heslo 11 až 21 znaků, což je velmi mnoho, uvážíme-li, že bychom tyto znaky měli volit zcela náhodně jejich taháním z klobouku. Takové superheslo jako například 1j,9Q;%P6gT&U7ba*9hV si ovšem běžný uživatel nemůže zapamatovat, proto je nutné zvolit jiný postup.



Počet bitů informace v klíčové větě

Klíčová věta – passphrase

Často se doporučuje jako heslo použít první písmena nějaké věty, kterou si dobře pamatujeme. Třeba z věty „Kočka leze dírou, pes oknem, nebude-li pršet, nezmoknem.“ dostáváme třináctiznakové heslo „Kld,po,n-lpn.“. Mnozí by takové heslo považovali za superkvalitní, ale není tomu tak! Pokud bychom uvažovali, že je náhodně vybírané z množiny malých písmen (pouze první je povinně velké) a interpunkčních znamének (čárka, pomlčka) uprostřed a případně čtyř znamének na konci {!, ?, ., nic}, dostáváme celkem $26*28^{11}*4 < 2^{60}$ možností. Je velmi nemilé překvapení, že tak dlouhé heslo poskytuje tak malou složitost! Jedna z příčin je, že jsme využili jen zlomek znaků klíčové věty. Klíčová věta má ve skutečnosti 55 znaků a my jsme jich použili jen 13. Pravda, nemusíme psát celou větu, ale přesto si ji musíme celou pamatovat. Když už si ji celou pamatujeme, je vyloženě škoda nevyužít takové informace a „zkrátit“ ji na 13 znaků.

Složitost klíčové věty

Pro výpočet kvality klíčové věty použijeme výsledků teorie informace. Uvažujme všechny řetězce, které se skládají z N znaků, které jsou tvořeny písmeny a až z. Je jich 26^N. Kolik z nich je jazykově smysluplných? Je jasné, že v té množině najdeme jak české smysluplné věty (bez háčků, čárek a mezer), tak věty anglické, španělské, maďarské, finské apod. Každý jazyk má trochu jinou obsažnost a nadbytečnost, takže počty smysluplných vět se liší. Anglických vět je zde cca $2^{1,5N}$, českých cca 2^{2N}. Zapíšeme-li 26^N jako $2^{4,7N}$, vidíme, že jedno písmeno může přinést informaci 4,7 bitu, ale v přirozených jazycích přináší méně – pouze 1,5 bitu v angličtině, 2 bity v češtině apod. Přirozené jazyky jsou ve srovnání s náhodným shlukem písmen pochopitelně příliš „ndbtčn“.

Nové výzkumy anglických hesel přinášejí přesnější hodnoty entropie, obsažené v klíčové větě v závislosti na její délce. Pro češtinu, která je obsažnější vůči angličtině zhruba v poměru 2 : 1,5, jsme provedli aproximaci. Počet bitů entropie české (anglické) klíčové věty počítáme takto: 4,5 (4,0) bitu za první znak, 2,67 (2,0) bitu za 2.–8. znak, 2,0 (1,5) bity za 9.–20. znak a 1,33 (1,0) bitu za 21. a každý další znak. Výsledky ukazuje obrázek.

V naší klíčové větě je 44 písmen. Nepočítáme mezery a interpunkci, protože ty si v hesle „Kockalezediroupesoknemnebudeliprsetnezmoknem“ můžeme doplnit my i útočník. Podle uvedené metodiky výpočtu nám klíčová věta dává entropii 79 bitů. To je mnohem více než 60 bitů, které jsme obdrželi jen z prvních písmen jejích slov. Odtud vyplývá první jednoduchý závěr: místo výběru prvních písmen z klíčové věty ji použijeme jako heslo celou!

Závěr

Pokud používáme klíčovou větu jako superheslo, můžeme si nyní vypočítat jeho kvalitu. V příštím čísle ukážeme nový přístup, jak heslo zásadně zkvalitnit jeho rozbitím neboli rozrušením klíčové věty.

Vlastimil Klíma, Tomáš Rosa,
v.klima@volny.cz, trosa@ebanka.cz