Testy a zdroje neurčitosti v počítači

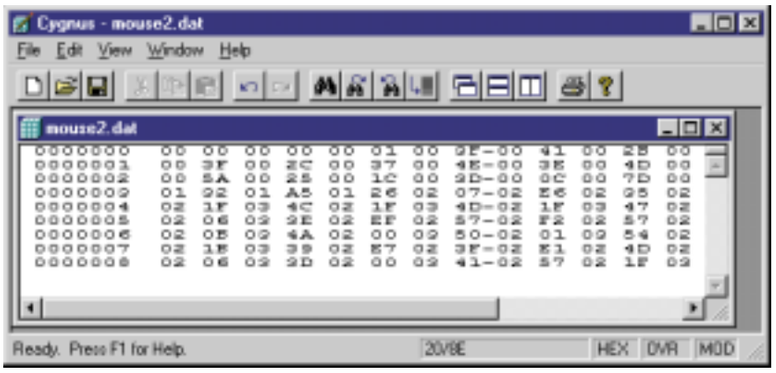
Dá se náhoda měřit?

V oblasti počítačové bezpečnosti se velmi často setkáváme s náhodnými čísly a šifrovacími klíči. Na kvalitě „náhodnosti" jejich generování přitom záleží úplně stejně jako na kvalitě používaných šifer. V tomto článku vás seznámíme s nedávným objevem, který umožňuje měřit kvalitu náhodnosti daného zdroje. Je to poměrně přesná metoda, jejíž význam však sahá daleko za hranice počítačové bezpečnosti.

Možná vás už nějaký program požádal, abyste chvíli náhodně ťukali do klávesnice nebo pohybovali myší. To jsou okamžiky, kdy na náhodnosti záleží natolik, že program odmítá za kvalitu svého zdroje převzít odpovědnost a obrací se přímo na uživatele. Znát míru náhodnosti používaného zdroje je nutné zejména u bezpečnostních aplikací. Kritické je to pak při generování šifrovacích klíčů. Jestliže generátor náhodných bitů nemá dostatečnou kvalitu, může se stát, že vygenerovaných 128 bitů šifrovacího klíče má pouze 40bitovou informační hodnotu (neurčitost, *entropii*). Generátor tak může snadno degradovat silnou šifru na slabou a důsledky mohou být značné. Tyto případy se už staly – a bohužel určitě nikoli naposledy.

Bezpečnost a náhoda

Přestože kvalitní zdroj entropie je při práci na počítači potřeba dost často, s požadavkem vložení náhodného čísla se v praxi setkáváme málokdy. Příslušné programy totiž nechtějí obtěžovat uživatele a generují náhodnost samy – jak umějí nejlépe. Ve většině případů k tomu využívají pouze „náhodnost" odvozenou od systémového času, což je ale z hlediska bezpečnosti silně nedostatečné. Náhodné šifrovací klíče musí například generovat internetový prohlížeč, pokud se se serverem spojuje zabezpečeným spojením prostřednictvím protokolu SSL. Jak možná víte, starší verze prohlížeče Netscape Navigator používala slabý generátor náhodných čísel, a šifrovací klíče tak měly entropii 47 namísto 128 bitů. Tím se degradovala kvalita šifrování a byla z toho ostuda. Od té doby se na kvalitu náhodných generátorů dbá více.



Obr. 1. Záznam pohybu myši na displeji s rozlišením 800 x 600 bodů (799 = 0x031F, 599 = 0x0257) z levého horního rohu obrazovky (00 00 00 00) do pravého dolního rohu (poslední záznam 57 02 1F 03). Každá pozice je zakódována 16 + 16 bity, z nichž „platných" je nejvýše 10 + 10 bitů, ani ty však nejsou ještě zcela náhodné.

Komprimace a náhodnost

Entropie vlastně určuje skutečné množství obsažené informace a měří se v bitech. Jednoduchým a známým měřítkem náhodnosti mohou proto být např. komprimační metody. Pokud nějaký soubor dat zkomprimujeme dejme tomu na 40 % původní délky, můžeme říci, že 60 % obsahu bylo nadbytečných a skutečný informační obsah byl 40 %. V jednom bajtu bylo tedy obsaženo jen 40 %, tj. 8*0,4 = 3,2 bitu skutečné informační hodnoty (entropie), neboli průměrná entropie na jeden bit byla 0,40. A co komprimovaný soubor – bude náhodný? Téměř ano, i když na jeho začátku mohou být prvopočáteční kusy původního textu a v jeho těle některé markantní řetězce. V mnoha případech ale komprimace skutečně velmi přiblíží soubor dat jeho informační hodnotě. Jestliže ale dáme zkomprimovat soubor náhodných dat, komprimační metody zkolabují. A to i v případech, že zdrojová data nejsou zcela náhodná, ale mají entropii například 0,90. Komprimace by měla daný soubor zkrátit na 90 %, ale nestane se tak, protože příslušná metoda prostě nezjistí, o jakou neurčitost vlastně jde. Neumí ji zjistit, změřit ani odstranit. V případech náhodných nebo téměř náhodných souborů tedy běžné komprimační metody jako měřítko neurčitosti použít nelze.

Objev v měření entropie

Průlom v měření entropie znamenal objev Ueliho Maurera z roku 1990, který jej prezentoval na kryptologické konferenci

CRYPTO´90 [1]. Nalezl velmi jednoduchou funkci, jíž dokázal měřit a pomocí statistického testu testovat entropii generátoru. Do té doby byla známa řada důmyslných testů, které zkoumaly partikulární parametry posloupnosti, jako například statistické vlastnosti (autokorelační test, test sérií, frekvenční test apod.) nebo složitostní charakteristiky, ale výsledky se nedaly kvantitativně převést na hodnotu entropie. Jinými slovy – věděli jsme, že posloupnost dejme tomu 200 pozic myši (měřených v časových mikrointervalech při jejím pohybu, viz obrázek 2) není náhodná a jaké má nedostatky (korelace sousedních pozic, nerovnoměrný výskyt jednotlivých bajtů), ale nevěděli jsme, jak dlouho máme myší pohybovat, aby posloupnost jejích pozic už reprezentovala například 128bitovou entropii, tj. ekvivalent 128 náhodných bitů. A Maurerův test právě toto dokázal vypočítat.

Obr. 2. Pohyb myši – pseudonáhodný zdroj s konečnou pamětí.

Použitelnost Maurerova-Coronova testu

V roce 1999 zpřesnil odhady konstant Maurerova testu J. S. Coron [2] a poté navrhl i geniální změnu testovací funkce [3]. Nový test tak oproti dřívějšímu měří entropii přímo a přesněji. Pro vás, kteří byste jej chtěli přímo použít, jej dále popíšeme. Test se týká stacionárních zdrojů s konečnou pamětí. Přesné definice a důkazy tvrzení můžete nalézt v uvedené literatuře. „Stacionární“ znamená, že se v čase nemění charakteristiky zdroje (například na pohyb myši nemá vliv to, zda je měřen v úterý, nebo ve čtvrtek), a konečná paměť (M) znamená, že n-tý výstup zdroje závisí maximálně jen na konečném počtu (M) předchozích výstupů – například poloha myši v daném okamžiku závisí maximálně na tom, kde byla před sekundou, ale už ne na tom, kde byla před deseti sekundami.

Výpočet entropie

Pojďme tedy k výpočtu entropie S podle Maurerova-Coronova (dále jen M-C) testu. Nejprve si zvolíme tři parametry – konstanty L, Q a K. Testovanou posloupnost N bitů si dále rozdělíme na Q + K nepřekrývajících se L-tic bitů b_1, ..., b_{K+Q}, kde b_i je i-tý blok o L bitech a N = (Q + K)*L.

Parametr L by měl být volen v rozmezí {6, ..., 16}, Q by mělo být co největší, minimálně ovšem $10*2^L$ a K alespoň $1000*2^L$. Jestliže např. zvolíme L = 8, zpracováváme posloupnost po bajtech.

Test má dvě fáze – inicializační a výpočetní. V **inicializační fázi** nejprve naplníme tabulku T[0] ... T[2^L-1] indexy prvních Q bloků tak, že pro i = 1, ..., Q postupně definujeme T[b_i] = i. Jinými slovy: prvních Q bloků použijeme na to, abychom naplnili tabulku T. Hodnota T[blok] je místo, kde se naposledy objevil L-bitový blok s hodnotou „blok“. Q by mělo být tak velké, aby se v inicializační fázi korektně naplnila tabulka T, tj. aby se v prvních Q blocích posloupnosti alespoň jednou objevil každý L-bitový blok.

Hodnotu S určíme ve **výpočetní fázi** podle vzorce na obrázku 3. Hodnota, kterou obdržíme, je rovna entropii L-bitového bloku. Chceme-li zjistit entropii zdroje na jeden emitovaný bit, postačí obdrženou hodnotu S vydělit počtem bitů bloku L – zdroj poskytuje neurčitost H = S/L na jeden bit. Pokud se nad vzorcem zamyslíme, zjistíme, že je to vlastně průměrná hodnota jakési funkce g, aplikované na vzdálenost mezi totožnými L-bitovými bloky v dané posloupnosti. Přitom průměr se počítá přes všechny bloky v posloupnosti.

Genialita funkce g je v tom, že uvedenou vzdálenost bloků „přeměňuje“ na entropii a navíc výpočet hodnoty S je velmi jednoduchý. Coron dokázal, že **S** z teoretického hlediska **vyjadřuje hodnotu entropie přesně** – navíc **známe její statistické rozdělení**. Umíme tedy entropii zdroje nejen vypočítat, ale určit i tzv. intervaly spolehlivosti, v nichž se naměřené hodnoty S mohou pohybovat, má-li mít zdroj maximální entropii.

Hodnocení výsledků

Když použijeme M-C statistický test na zkoumanou posloupnost, obdržíme jednu jedinou hodnotu – tzv. *statistiku S*. Tato statistika je ve střední hodnotě rovna přímo entropii L-bitového bloku zkoumaného zdroje, ale zároveň je to náhodná veličina, která má pravděpodobnostní chování. A tak, i když má zdroj dokonalou entropii (například S = 8 na bajt), hodnoty S naměřené na konkrétních posloupnostech se mohou pohybovat v určitých intervalech kolem této dokonalé entropie. Tyto intervaly spolehlivosti (IS) umíme vypočítat, neboť S můžeme aproximovat normálním rozdělením, jehož parametry naposledy zpřesnil právě J. S. Coron [3].

K výpočtu IS si nejprve stanovíme pravděpodobnost **r**, že M-C testem vyřadíme nějakou posloupnost jako špatnou (nemající maximální entropii), přestože byla emitována skutečně náhodným zdrojem;

Vzorec pro výpočet entropie zdroje (S) je geniálně jednoduchý a přesný:

$$S = (1/K) * \sum_{n=Q+1}^{Q+K} g(n - T[b_n]),$$

přičemž

$$g(i) = (1/\log_e 2) * \sum_{k=1}^{i-1} 1/k,$$

kde $T[b_n]$ je index posledního výskytu bloku b_n v posloupnosti bloků b_1, b_2, ... , b_{n-1} a S je entropie zdroje na L-bitový blok, tj. S/L je entropie zdroje na jeden emitovaný bit.

Obr. 3. Vzorec pro výpočet entropie.