Nové užitečné statistické testy

Je to s podivem, že byly objeveny nové statistické testy, které jsou jiné a užitečnější, než stávající. Jsou skvělé proto, že jsou univerzálně použitelné na testování náhodnosti binárních (a po modifikaci i nebinárních) dat. Statistika je věda natolik propracovaná, že přijít s něčím novým je opravdu neobvyklé. Proto ta novota vznikla na pomezí statistiky a kryptologie, kdy se tyto statistické testy ukázaly jako velmi citlivé k měření kvality kryptografických nástrojů. Pomocí nich byly po třinácti letech zkoumání nalezeny první „slabiny“ blokové šifry AES, přesněji první odchylky blokové šifry AES od „dokonale náhodné“ šifry. Další výsledky přinesly tyto testy při zkoumání hašovacích funkcí v rámci probíhající světové soutěže na nový standard SHA-3. Mohou dokonce zasáhnout do jejího průběhu a naznačují, kde mohou mít kandidátské funkce slabiny a silné stránky. Testy jsou natolik složité a výkonnostně náročné, že se musí ještě potvrdit nezávislými realizacemi, stejně jakoby se jednalo o nějaké fyzikální experimenty nebo astronomická zkoumání. To, že se testy musí ověřovat nezávislými týmy, aby se potvrdila jejich platnost, je také zajímavá paralela s naším okolním vědeckým světem, avšak pokud víme, první svého druhu v kryptologii. V matematice se toto už stalo mnohokrát, naposledy při ověřování důkazu Velké Fermatovy věty. Ověřování trvalo několik let a bylo děláno nadvakrát – první důkaz byl špatně a až druhý byl ověřen. Možná se čtenářům testy zalíbí a najdou tak možnost jejich využití ve své praxi. Ve své podstatě jsou jednoduché, což je činí ještě atraktivnějšími. Složité na jejich realizaci pro AES nebo hašovací funkce je to, že k rozpoznání odchylek pro odlišení těchto funkcí od náhodných funkcí je potřeba enormní objem dat. Předpokládáme, že se brzy dostanou do učebnic a do standardních balíků statistických testů a že ukáží některé vlastnosti finálních pěti kandidátů na SHA-3.

Sada testů náhodnosti

Předesíláme, že nyní se budeme věnovat sadě čtyř testů, které navrhli turečtí matematici [1]. Netvrdíme, že právě tyto testy testují všechno a jsou jediné pravé pro testování náhodnosti, avšak umí dostat lepší výsledky při použití k testování blokových šifer, než byly obdrženy dosud, a lepší, než balík statistických testů NIST. Jistě je proto lze použít i tam, kde bylo testování náhodnosti binárních posloupností již prováděno nebo se bude provádět. Zajímavé by bylo porovnání i s jinými existujícími statistickými balíky nebo jejich obohacení o nové testy. Konkrétně si ukážeme výsledky sady čtyř testů při použití na pět blokových šifer, z nichž se vybíral vítězný AES (tj. i včetně Rijndaelu, budoucího AES), viz *tabulka 1*. V pravém sloupci tabulky vidíme původní výsledek. Každý z algoritmů postupně nabírá složitost v tzv. rundách, které se opakují až do poslední rundy, která dává výsledek. Při první rundě není výsledek „šifrování“ valný, takže ho všechny testy zachytí jako nenáhodný, při druhé rundě už neprotestuje test Linear Span, při třetí rundě mlčí další dva testy u MARSu a Rijndaelu atd. Například pro splnění testu SAC potřeboval MARS šest rund, zatímco původní testy NIST prohlašovaly už MARS se čtyřmi rundami za náhodný. V tabulce vidíme, že nová sada testů zpřesnila výsledky NIST u všech pěti testovaných blokových šifer.

Tabulka 1 Původní (NIST) a přesnější výsledky nových testů pěti kandidátů AES

Test: / Šifra:	SAC	LST	CoiT	CovT	Výsledky NIST
MARS	6	2	3	3	4
RC6	5	2	5	5	4
Rijndael	4	2	3	3	3
Serpent	4	2	4	4	4
Twofish	4	2	4	4	2

Test lavinovitosti (SAC)

Kdykoliv se změní jeden bit vstupu blokové šifry, měl by se s pravděpodobností 0,5 změnit každý bit výstupu. Toto nazýváme kritériem lavinovitosti (Strict Avalanche Criterion, SAC).

Test nelinearity (LST)

Správná náhodná Booleovská funkce by měla být nelineární, čili její vzdálenost od množiny všech afinních Booleovských funkcí daného počtu vstupních proměnných (bitů) by měla být velká. Nelinearita je také odrazem základního kritéria, který pro šifru stanovil Shannon jakožto kritérium zvané konfúze (confusion). Odpovídající test se nazývá test rozsahu lineární závislosti (Linear Span Test, LST) a měří, jak mnoho je lineárně závislá množina výstupů dané funkce (šifry), když její vstup tvoří množina vysoce lineárně závislých vstupů.

Test kolize (CoIT)

Nalezení dvou hodnot vstupu, které dávají stejnou hodnotu výstupu, by mělo být u náhodného zobrazení složité. Víme, že u náhodného binárního zobrazení n bitů na n bitů je množina výstupů pouze cca 70 procent z celkového možného počtu a existují hodnoty, které není možno nabýt a naopak některé mají několik vzorů (nastane kolize). Bloková šifra je zvláštní v tom, že při pevném klíči je to náhodná permutace (kolize nikdy nenastane), ale při pevném vstupu a proměnném klíči je to (mělo by být) náhodné zobrazení, u kterého se může měřit míra koliznosti. V praxi se také z blokové šifry (náhodné permutace) vyrobí náhodné zobrazení tak, že se výstup ořízne na menší počet bitů (třeba 10 nebo 20) a kolize se hledají v této množině hodnot. Podobně to lze provést i při pevném klíči.

Test pokryvnosti (CovT)

Tento test měří náhodné zobrazení z opačné strany, tj. dívá se na to, jak mohutný je obraz celé vstupní množiny a jestli náhodou velikost této množiny obrazů není malá nebo velká. Také zde se může uvažovat jen příslušný počet bitů výstupu blokové šifry nebo hašovací funkce.

Další sada testů náhodnosti

Tyto první čtyři testy zabraly na blokové šifry. Další sadu pěti testů turečtí matematici navrhli v [2] a použili je k odhalení možných slabin 14 funkcí ze soutěže SHA-3. V prvním testu jde o zprávy (vstupy hašovacích funkcí), které mají velmi malý počet bitů 1, v druhém testu o zprávy s velkým počtem bitů 1, ve třetím testu se bere náhodná zpráva a postupně se v ní mění každý její bit a ve čtvrtém testu se mění postupně každý bajt zprávy tak, že nabývá všech 255 změněných hodnot. V pátém testu se náhodná zpráva nechá rotovat postupně o jeden, dva až maximum bitů. V každém testu se tak obdrží řada vstupů a jim odpovídajících výstupů, přičemž výstupy se zřetězí do jedné posloupnosti. Ta se testuje klasickou sadou testů NIST a výsledky se vyhodnocují. Technické detaily naleznou zájemci v [1] a [2], ale chtěli jsme ukázat, jak turečtí matematici dobře vyhmátli jaké strukturální závislosti na vstupu volit. Testy (je možné použít nejen sadu NIST, ale jakoukoli oblíbenou sadu testů) pak jen zjišťují, zda se strukturální závislosti na vstupu jakkoli statisticky odrazily na výstupu. Pokud ano, je na zkoumané funkci něco nedokonalého. A právě to bylo u některých funkcí zjištěno.

Vlastimil Klíma
vlastimil.klima@knzsro.cz

LITERATURA

[1] Doganaksoy, A., Ege, B., Kocak, O., Sulak, F.: Cryptographic Randomness Testing of Block Ciphers and Hash Functions. Dostupný z: http://eprint.iacr.org/2010/564.pdf.

[2] Doganaksoy, A., Ege, B., Kocak, O., Sulak, F.: Statistical Analysis of Reduced Round Compression Functions of SHA-3 Second Round Candidates. Dostupný z: http://eprint.iacr.org/ 2010/611.pdf.