Cryptology

Náhodnost (background)

Definition 1 (Probabilistic TM)**.** *Probabilistic TM is a NTDM with a function* $P : (Q \times \Sigma \times Q \times \Sigma \times M) \to [0, 1]$ *s.t.*

$$\forall (q, a) \in Q \times \Sigma : \Sigma_{(q', \sigma', m') \in Q \times \Sigma \times M} P(q, \sigma, q', \sigma', m) = 1,$$

Q being a set of possible states and $M = \{\leftarrow \rightarrow, \_\}$.

Definition 2 (Zanedbatelná funkce)**.** *$\mu : \mathbb{N} \mapsto \mathbb{R}_+$ je zanedbatelná, jestliže pro každý polynom p existuje $n_p \in \mathbb{N}$ takové, že*

$$\forall n \geq n_p : \mu(n) \leq \frac{1}{p(n)}.$$

V této části nás budou zajímat především soubory náhodných veličin $\{X_i\}_{i \in \mathbb{N}}$, X_i nabývá hodnot z $\{0, 1\}^{l(i)}$ pro $l(i) : \mathbb{N} \mapsto \mathbb{N}$. Obvykle $l(i) \geq i$, nicméně často budeme uvažovat $l(i) = i$. Pro tyto soubory se budeme snažit zjistit podobnost se soubory $\{U_n\}_{n \in \mathbb{N}}$ příslušných rovnoměrně rozdělených náhodných veličin.

Definition 3 (Blízkost souborů)**.** *Buďtež $\{X_i\}_{i \in \mathbb{N}}$ a $\{Y_i\}_{i \in \mathbb{N}}$ dva soubory náhodných veličin. Pak rozlišujeme tyto stupně blízkosti souborů náhodných veličin:*

- *Soubory jsou identické, pokud pro všechna $n \in \mathbb{N}$ platí $X_n = Y_n$, tj. X_n a Y_n mají stejné rozdělení.*
- *Soubory jsou statisticky blízké, jestliže jejich statistická diference $\Delta_{X,Y}(n)$ je zanedbatelná funkce.*

$$\Delta_{X,Y}(n) = \frac{1}{2} \Sigma_{v \in \{0,1\}^n} |Pr[X_n = v] - Pr[Y_n = v]|$$

- *Soubory jsou výpočetně nerozlišitelné, jestliže pro každý polynomiální pravděpodobnostní algoritmus D je $\delta_{D,X,Y}(n)$ zanedbatelná funkce.*

$$\delta_{D,X,Y}(n) = |Pr[D(X_n) = 1] - Pr[D(Y_n) = 1]|$$

- *Soubory jsou silně výpočetně nerozlišitelné, jestliže pro každý soubor obvodů $\mathcal{C} = \{C_n\}_{n \in \mathbb{N}}$ polynomiální velikosti je $\sigma_{\mathcal{C},X,Y}(n)$ zanedbatelná funkce.*

$$\delta_{D,X,Y}(n) = |Pr[C_n(X_n) = 1] - Pr[C_n(Y_n) = 1]|$$

Definition 4. *Řekneme, že soubor náhodných veličin $\{X_i\}_{i \in \mathbb{N}}$ je (silně) pseudonáhodný, jestliže je (silně) výpočetně nerozlišitelný od souboru $\{U_n\}_{n \in \mathbb{N}}$ náhodných veličin s uniformním rozdělením.*

Definition 5 (jednosměrná funkce)**.** *Funkce $F : \{0, 1\}^* \mapsto \{0, 1\}^*$ je jednosměrná, jestliže je spočitatelná v polynomiálním čase a je těžko invertovatelná, tj. pro každý pravděpodobnostní polynomiální algoritmus B je funkce*

$$Pr_{x \sim U_n}[B(f(x)) \in f^{-1} \circ f(x)], \; x \sim U_n$$

zanedbatelná v n.

Zvláštním případem jsou *jednosměrné permutace*, prosté jednosměrné funkce zachovávající délku vstupu.

Definition 6 (Těžký bit)**.** *Zobrazení $b : \{0,1\}^* \mapsto \{0,1\}$ je těžký bit funkce f, jestliže platí:*

- *b je spočitatelné v polynomiálním čase*
- *není odhadnutelné s nezanedbatelnou výhodou, tj. pro každý pravděpodobnostní algoritmus B je funkce*

$$|Pr_{x \sim U_n}[B(f(x)) = b(x)] - \frac{1}{2}|$$

  zanedbatelná v n.

Je-li f prostá, pak má těžký bit, právě když je jednosměrná. V opačném případě bychom mohli hodnotu bitu $b(x)$ spočíst invertováním $f(x)$. Pokud by naopak bylo možné s nezanedbatelnou pravděpodobností spočíst libovolný bit x_i, pak bychom mohli vzor spočíst s nezanedbatelnou pravděpodobností i celé x.

Pro neprosté funkce je situace složitější, jelikož požadujeme nalezení právě použitého x, nikoliv jiného se stejným obrazem. Pak tedy například $f(x)$ má dva stejně pravděpodobné vzory x a x' a $b(x) \neq b(x')$, pak je odhad nemožný.

Lemma 1. *Pro x délky n označme $b_x(r)$ orákulum splňující*

$$|Pr_{r \sim U_n}[b_x(r) = \langle x, r \rangle] - \frac{1}{2}| \geq \varepsilon(n).$$

Existuje pravděpodobnostní algoritmus A pracující v čase polynomiálním čase v $(\frac{n}{\varepsilon(n)})$ takový, že pro každé x platí

$$Pr[A^{b_x}(1^n) = x] \geq \frac{\varepsilon(n)^2}{2n}.$$

Theorem 2. *Nechť f je libovolná jednosměrná funkce. Definujme funkci g předpisem*

$$g(x, r) := (f(x),\, r), |x| = |r|.$$

Pak bodový součin vektorů x, r modulo 2 je těžkým bitem funkce g.

Proof. Nechť $\langle x, r \rangle$ není těžký bit funkce g a označme G algoritmus, který to dokazuje a označme γ jeho výhodu:

$$\gamma(n) = |Pr_{x, r \sim U_n}[G(f(x), r) = \langle x, r \rangle] - \frac{1}{2}|.$$

Budiž S_n množina takových x délky n, na kterých je výhoda G alespoň $\gamma(n)/2$. Tato množina má velikost alespoň $\frac{\gamma(n)}{2}2^n$, jinak by výhoda G byla menší než

$$\frac{1}{2^n}\left(\Sigma_{x \in S_n} 1 + \Sigma_{x \notin S_n} \frac{\gamma(n)}{2}\right) < \frac{\gamma(n)}{2} + \frac{\gamma(n)}{2} = \gamma(n)$$

Dle Lemmatu 1 lze pro každé $x \in S_n$ v čase, který je polynomiální v $n/\gamma(n)$ a s pravděpodobností alespoň $\mathrm{poly}(\gamma(n)/n)$ najít x s pomocí $f(x)$ ($b_x(r) = G(f(x), r)$). Protože $x \sim U_n$ leží v S_n s pravděpodobností alespoň $\gamma(n)/2$ a $\gamma(n)$ není zanedbatelná, dostáváme spor s jednosměrností f. $\square$

Buď f jednosměrná funkce a b její těžký bit, pak závazkem může být $Z(m, x) := (f(x), m \oplus b(x))$, autorizací k otevření závazku je x.

Generátory pseudonáhodných čísel

Definition 7. *(Neuniformně silný) pseudonáhodný generátor je deterministický algoritmus G, který vstup délky n prodlužuje na výstup délky $l(n) > n$ tak, že soubor $\{G(U_n)\}_{n\in N}$ je (silně) pseudonáhodný.*

Theorem 3. *Nechť je f jednosměrná permutace a b její těžký bit. Pak je zobrazení $G : x \mapsto f(x) \parallel b(x)$ pseudonáhodný generátor.*

Proof. Chceme rozlišit $f(x) \parallel b(x)$ od náhodné posloupnosti, tzn. chceme uhodnout $b(x)$ z $f(x)$.

Nechť G není pseudonáhodný generátor a D je algoritmus, který to dokazuje, tj. $|Pr[D(f(x)b(x)) = 1] - Pr[D(y') = 1]| = \varepsilon(n)$, ε nezanedbatelná funkce. Toto lze přepsat do formy

$$|Pr[D(y \parallel b \circ f^{-1}(y)) = 1] - Pr[D(y \parallel \sigma) = 1]| = \varepsilon(n).$$

Protože s pravděpodobností $1/2$ platí $\sigma = b \circ f^{-1}(y)$, pak také

$$Pr[D(y \parallel \sigma) = 1] = \frac{1}{2}Pr[D(y \parallel b \circ f^{-1}y) = 1] + \frac{1}{2}Pr[D(y \parallel \overline{b \circ f^{-1}y}) = 1]$$

$$|Pr[D(y \parallel b \circ f^{-1}y) = 1] - Pr[D(y \parallel \overline{b \circ f^{-1}y}) = 1]| = 2\varepsilon(n)$$

Rozlišovač $A(y) := \sigma$, jestliže $D(y \parallel \sigma) = 1$, $A(y) := \overline{\sigma}$ v opačném případě. Úspěšnost A je:

$$Pr[A(y) = b \circ f^{-1}(y)] =$$

$$= Pr[D(y \parallel \sigma) = 1 \wedge \sigma = b \circ f^{-1}(y)] + Pr[D(y \parallel \sigma) = 0 \wedge \overline{\sigma} = b \circ f^{-1}(y)]$$

$$\frac{1}{2}Pr[D(y \parallel b \circ f^{-1}(y)) = 1] + \frac{1}{2}Pr[D(y \parallel \overline{b \circ f^{-1}(y)}) = 0]$$

$$\frac{1}{2}Pr[D(y \parallel b \circ f^{-1}(y)) = 1] + \frac{1}{2}(1 - Pr[D(y \parallel \overline{b \circ f^{-1}(y)}) = 1])$$

$$\frac{1}{2} + \frac{1}{2}(Pr[D(y \parallel b \circ f^{-1}(y)) = 1] - Pr[D(y \parallel \overline{b \circ f^{-1}(y)}) = 1]) = \frac{1}{2} \pm \varepsilon(n)$$

Výhoda A, tj. $|Pr[A(y) = b \circ f^{-1}(y)] - \frac{1}{2}|$, není zanedbatelná funkce, což je spor s předpokladem, že b je těžký bit. □

Theorem 4. *Nechť G je pseudonáhodný generátor s prodlužovací funkcí $l(n) = n + 1$ a nechť l' je libovolný polynom. Definujme zobrazení G' předpisem*

$$G'(s) = \sigma_1(s) \parallel \cdots \parallel \sigma_{l'(|x|)}(s),$$

kde

$$x_0 = s,\ G(x_{i-1}) = x_i \parallel \sigma_i(s),\ i = 1, 2, \ldots, l'(|x|).$$

Pak je G' pseudonáhodný generátor.

Proof. Buď $X_1, \ldots, X_{l'(n)}$ nezávislé kopie náhodné proměnné U_1. Budeme chtít ukázat, že $(X_1, \ldots)$ neumíme rozlišit od $(\sigma_1(s), \ldots)$ s nezanedbatelnou výhodou. Zdůrazněme, že $\sigma_i(s)$ vycházejí ze stejné volby s a jsou tedy závislé. Zadefinujme si hybridní distribuci:

$$H^{(i)} = (X_1, \ldots, X_i, \sigma_1(s), \ldots, \sigma_{l'(n)-i}(s))$$

Algoritmus D, který rozlišuje $H^{(0)}$ od $H^{l'(n)}$ s nezanedbatelnou výhodou, rozlišuje taktéž s nezanedbatelnou výhodou nějaké $H^{(i)}$ od $H^{(i+1)}$. Předpokládejme, že jsme takové i zvolili ($Pr = 1/l'(n)$). Pro dané $y \parallel \sigma$ necháme rozhodnout

$$(X_1, \ldots, X_i, \sigma, \sigma_1(y), \ldots, \sigma_{l'(n)-i-1}(y))$$

Je-li $y \parallel \sigma \sim U_{n+1}$, jde o $H^{(i+1)}$. Je-li $y \parallel \sigma = G(s)$ pro $s \sim U_n$, jde o $H^{(i)}$, protože $\sigma = \sigma_1(s)$ a $\sigma_i(y) = \sigma_{i+1}(s)$. □

Důkazy s nulovou znalostí

Definition 8 (ITM). *Interaktivní Turingův stroj je vícepáskový pravděpodobnostní TM se vstupem a výstupem, který kromě vstupní pásky (veřejná), výstupní pásky a pracovní pásky obsahuje ještě*

- *dodatečnou vstupní pásku (soukromá)*
- *vstupní komunikační pásku (read-only)*
- *výstupní komunikační pásku (write-only)*
- *stavový bit (1 políčko s 1/0)*

Stroji je přiřazena identita jedna nebo 0. Program pak obsahuje instrukce pouze pro případ, že je stavový bit roven identitě stroje. V takovém případě stroj pracuje, jinak je nečinný.

Definition 9. *Interaktivní systém je dvojice ITM (A, B), které*

- *mají opačnou identitu,*
- *sdílí veřejnou vstupní pásku,*
- *sdílejí stavový bit,*
- *vstupní komunikační páska A je výstupní komunikační páskou B a naopak.*

Konvence: A dokazovatel, B ověřovatel; tj. výstupem interaktivního výpočtu je pouze výstup B a složitost se bere v potaz pouze u ověřovatele.

Definition 10 (IP). $L \in IP \Leftrightarrow \exists$ *interaktivní systém (A, B) t.ž.*

- *(efficiency) poly-time*
- *(completeness) (A, B) přijme $x \in L$ s pr. alespoň $2/3$*
- *(soundness) $x \notin L$, pak pro lib. ITM A^* je pravděpodobnost, že (A^*, B) přijme x, menší než $1/3$.*

Theorem 5. *1. $BPP \subseteq IP$*

2. $NP \subseteq IP$

Proof. 1: $L \in BPP$ a B ho rozhoduje ve smyslu BPP. Pak $(\emptyset, B)$ rozhoduje L ve smyslu IP – výpočet proběhne v jedné fázi.

2: $L \in NP$, A sdělí slovo y (svědek x) B. B ověří, že y je ve svědecké relaci s x (ověření v poly-time, úplnost daná existencí svědka, spolehlivost z neexistence svědka pro $x \notin L$) □

Definition 11 (Grafový neisomorfismus). *GraphNI*$=\{(G_1, G_2) | G_1 \not\cong G_2\}$

Grafový neizomorfismus je zřejmě v $PSPACE$ (mohu in-place vyzkoušet všechna řešení).

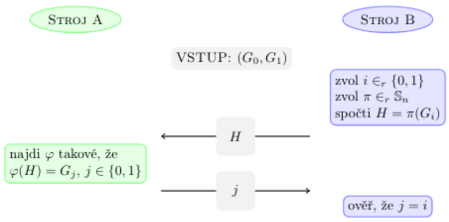
Claim 6. $GraphNI \in IP$

Proof. Uvažme dvě kola uvedené komunikace. Efektivita je zřejmá. Pokud $(G_0, G_1) \in GraphNI$, pokud $\varphi(H) = G_j$, pak platí $i = j$ – B přijme vstup s pravděpodobností 1 (úplnost).

$(G_0, G_1) \notin GraphNI$, tj. grafy jsou izomorfní. Definujme multiset

$$M_i := \{\pi(G_i) | \pi \in S_n\}, \ i = 0, 1.$$

Zřejmě v tomto případě $M_0 = M_1$, tj. A odpoví nezávisle na volbě i. Z uniformity volby i je tedy pr. přijetí vstupu $(\frac{1}{2})^2$, čímž je dokázána spolehlivost.



□

Theorem 7 (Shamir). $IP = PSPACE$

Definition 12. *Interaktivní důkazový systém (A, B) se nazývá důkaz s nulovou znalostí (zero-knowledge proof) pro jazyk L, pokud rozhoduje jazyk ve L ve smyslu Definice 9 a navíc pro každý interaktivní stroj B^* existuje poly-time PTM M takový, že pro každé $x \in L$ platí:*

- *$Pr[M(x) = \perp] \leq 1/2$ (výpočet M selhal)*
- *$M(x) \neq \perp \Rightarrow M(x) \sim \sigma_F(A, B^*)(x)$ pro $\sigma_F(A, B^*)(x)$ závěrečný snímek $B*$ po výpočtu (A, B^*) na x.*

Symbol $\sim$ odpovídá jednomu ze stupňů blízkosti ($\{M(x)\}_{x \in L}$ i $\{\sigma_F(A, B^*)(x)\}_{x \in L}$ soubory náhodných veličin):

- *$\sim$ rovnost souborů – důkaz s dokonale nulovou znalostí (třída PZK),*
- *$\sim$ statistická blízkost – důkaz s téměř dokonalou nulovou znalosti (třída SZK),*
- *$\sim$ výpočetní nerozlišitelnost – důkaz s výpočetně nulovou znalosti (třída CZK).*

Poznamenejme, že M neumí rozhodovat jazyk – simulátor je úspěšný jen pro $x \in L$, pro $x \notin L$ se může chovat libovolně.

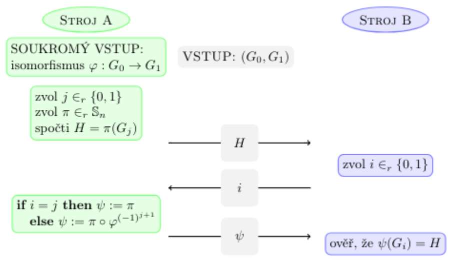
Claim 8. $GraphNI \in PZK$

Proof. Popíšeme simulátor M ověřovatele B^*. Stroj M se může chovat jako B, jen si sám musí generovat zprávy od A. Pokud $\{G_0, G_1\} \in GraphNI$ je simulace A snadná – A vždy pošle j, které je rovnou i, což je hodnota stroji M známá z předchozí simulace. □

Definition 13 (Grafový isomorfismus)**.** *GraphISO*=$\{(G_1, G_2) | G_1 \cong G_2\}$

Claim 9. $GraphISO \in PZK$

Proof. Důkaz spočívá v dvojím opakování popsané komunikace. Stroj B přijme, jestliže v obou kolech $\phi(G_i) = H$. Systém je zřejmě efektivní a úplný.



Předpokládejme nyní, že $(G_0, G_1) \notin GraphISO$, pak ale zpráva H je, bez ohledu na postup A^*, isomorfní nejvýše jednomu z grafů. S pravděpodobností $1/2$ B zvolí i, pro které neexistuje ψ splňující $\phi(G_i) = H$. Ve dvou kolech tedy odmítne s pravděpodobností $3/4$ – spolehlivost.

Nyní je třeba dokázat nulovou znalost. M může simulovat celý výpočet s výjimkou situace $\psi = \pi \circ \varphi^{-1}$, tj. $i \neq j$. S pr. $1/2$ (pr. volby $i \neq j$) simulace selže hned v prvním kole. Při dvou kolech je tato pravděpodobnost $1/4$ – stačí tedy tři opakování pokusu o simulaci, aby se pravděpodobnost dostala nad $1/2$. M tedy opakuje celý postup třikrát a vytiskne $\perp$, pokud všechny pokusy selhaly. Jinak vytiskne výstupní snímek B^* po úspěšném pokusu. □

Hašovací funkce

Definition 14 (Hašovací funkce)**.** *Funkci* $h : \{0,1\}^* \to \{0,1\}^n$ *nazýváme hašovací funkcí. Obraz* $h(x)$ *nazýváme otisk, hash nebo digest prvku* x*. Pro* $x \neq y : h(x) = h(y)$ *řekneme, že pár* (x, y) *je kolize funkce* h*.*

Definition 15 (Kryprografická hašovací funkce)**.** *Kryptografická hašovací funkce je hašovací funkce, která navíc splňuje 3 vlastnosti:*

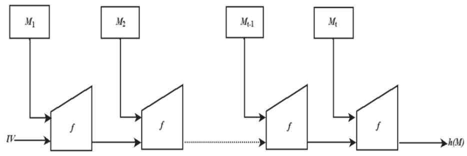
- ***Preimage-resistance:*** *je obtížné nalézt* x *pro dané* y *t.ž.* $h(x) = y$*.*
- ***2nd-preimage-resistance:*** *je obtížné nalézt* $x \neq x'$ *pro dané* y *t.ž.* $h(x) = y = h(x')$*.*
- ***Collision resistance*** *(bezkoliznost/kolizevzdornost): je obtížné nalézt* x*,* x' *t.ž.* $h(x) = h(x')$*.*

Formální definice může být zkonstruována například prostřednictvím zanedbatelné funkce. Nicméně v praxi má úskalí – např. nevíme, zda-li existuje jednosměrná funkce, kterou bychom touto konstrukcí dostali.

Obecně máme dvě konstrukce pro hašovací funkce - Merkle-Dåmgard (SHA1, SHA2, MD5) a houbovitá konstrukce (SHA3). Hlavní rozdíl mezi konstrukcemi je, že houbovité konstrukce ve výstupu nezveřejní svůj celý vnitřní stav. Toto například zabrání tzv. *length-extension attack*.

Merkle-Dåmgard

Tato konstrukce má dvě varianty, které v praxi splývají. První varianta používá padding $x_n|10...0|length_{64}(x)$. Druhá varianta, která za cenu většího nafouknutí zprávy odstaní omezení na celkovou délku zprávy. Tato varianta prefixuje první blok 1 a ostatní bloky 0. Poslední blok je vyplněn stejným způsobem, jen místo délky zprávy se použije délka paddingu. Důkaz bezpečnosti pro obě varianty je velmi podobný, proto uvedeme důkaz pouze pro první variantu.



Theorem 10. *Je-li kompresní funkce $f : \{0,1\}^s \times \{0,1\}^b \to \{0,1\}^s$ kolizivzdorná, pak h je též kolizivzdorná.*

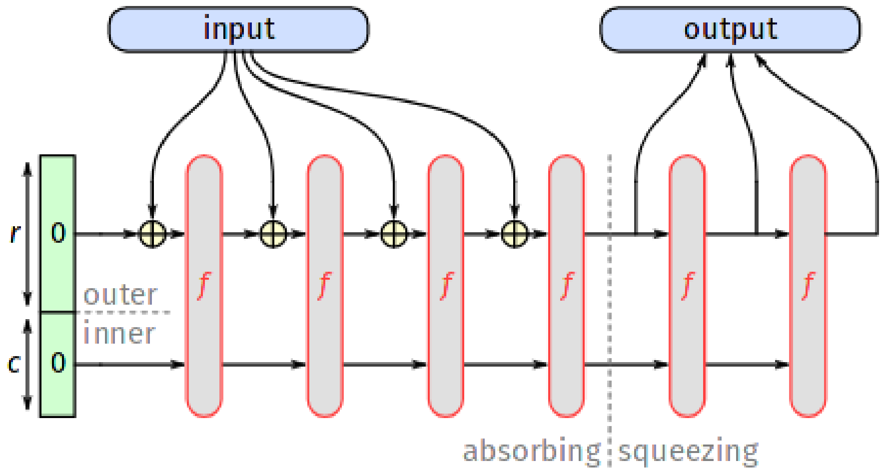
Proof. $x = B_1||\dots||B_n, x' = B'_1||\dots||B'_n$ t.ž. $h(x) = h(x'), x \neq x'$

$\Rightarrow f(S_{n-1}, B_n) = f(S'_{n-1}, B'_n)$, tj, buď $(S_{n-1}, B_n) \neq (S'_{n-1}, B'_n)$ (a máme kolizi pro $f \Rightarrow$ spor nebo kolize nastala dříve. Jelikož B_n a B'_n obsahuje informaci o délce, pak jsou tyto zprávy x a x' stejně dlouhé.

Indukčně postupujeme až k prvnímu bloku – buďto nastala kolize v průběhu (což je spor s kolizivzdorností) nebo, jelikož první bloky musí být různé (dle předpokladu $x \neq x'$ a rovnosti $B_i = B'_i$ pro $i > 1$ z indukčních kroků), nastala kolize v prvním bloku což je opět spor s kolizivzdorností f. □

Remark (Davies-Mayer schéma). Buď E šifra, pak kompresní funkci dle Davies-Mayer schéma definujeme jako $f(S_{i-1}, B_i) = E_{B_i}(S_{i-1}) + S_{i-1}$, kde + je celočíselné sčítání po slovech mod 2^{32}. Tuto konstrukci používá mj. MD5, SHA-1, SHA-2.

Houbovité konstrukce



Houbovité konstrukce mají dvě fáze – absorpce a ždímání. Existují i konstrukce, které tyto fáze střídají (tzv. duplex), vstup