TIN 085 - Vybrané kapitoly z výpočetní složitosti I **zima 2006/2007**

2. domácí úlohy - hrátky s hašováním

Vzorové řešení

do 14. listopadu 2006

1. Nechť $n > 1$ je celé číslo a $A \subseteq \{0,1\}^n$. Pro vektor $x \in \{0,1\}^n$, označme jako $A \oplus x = \{x \oplus y; y \in A\}$, kde $x \oplus y$ je xor x a y po složkách.

   (a) Ukažte, že když $|A| > \frac{3}{4}2^n$, pak existují vektory $r_1, r_2, \ldots, r_n \in \{0,1\}^n$ takové, že $\{0,1\}^n \subseteq \bigcup_{i \in \{1,\ldots,n\}} A \oplus r_i$. (Hint: Použijte náhodné vektory r_i.)

   (b) Ukažte, že když $|A| < 2^n/n$, pak pro každou n-tici vektorů $r_1, r_2, \ldots, r_n \in \{0,1\}^n$ existuje $x \in \{0,1\}^n \setminus \bigcup_{i \in \{1,\ldots,n\}} A \oplus r_i$.

   (Toto je podstata důkazu, že $BPP \subseteq \Sigma_2 = NP^{NP}$. Pro daný vstup w se za množinu A berou náhodné řetízky, na kterých BPP algoritmus odpoví 1. Rozmyslete si!)

   Řešení:

   (a) Vezměme si nějaké $x \in \{0,1\}^n$. Tvrdíme, že $\Pr_{r_1,r_2,\ldots,r_n \in \{0,1\}^n}[x \notin \bigcup_{i \in \{1,\ldots,n\}} A \oplus r_i] < 4^{-n}$. Předně, pro libovolné i, $\Pr_{r_i \in \{0,1\}^n}[x \notin A \oplus r_i] = \Pr_{r_i \in \{0,1\}^n}[x \oplus r_i \notin A] = \Pr_{r_i \in \{0,1\}^n}[r_i \notin A \oplus x] = 1 - |A|/2^n < 1/4$. Jelikož r_i jsou nezávislá, $\Pr_{r_1,r_2,\ldots,r_n \in \{0,1\}^n}[x \notin \bigcup_{i \in \{1,\ldots,n\}} A \oplus r_i] < 4^{-n}$, tudíž $\Pr_{r_1,r_2,\ldots,r_n \in \{0,1\}^n}[\exists x \in \{0,1\}^n, x \notin \bigcup_{i \in \{1,\ldots,n\}} A \oplus r_i] < 2^n \cdot 4^{-n} = 2^{-n} < 1$. Tedy hledané vektory r_i existují.

   (b) Pro libovolné $r \in \{0,1\}^n$, $|A \oplus r| = |A|$, tudíž když $|A| < 2^n/n$, pak $\left|\bigcup_{i \in \{1,\ldots,n\}} A \oplus r_i\right| < n \cdot 2^n/n = 2^n = |\{0,1\}^n|$. Tvrzení plyne okamžitě.

   Ohledně $BPP \subseteq NP^{NP}$. Pokud S je BPP algoritmus který má chybu menší než $1/n$, pak NP^{NP} algoritmus pro ten samý jazyk na vstupu w uhodne $r_1, r_2, \ldots, r_n$a poté se zeptá svého NP orakula, zda pro každý řetízek y, S přijme w alespoň na jednom z $y \oplus r_1, y \oplus r_2, \ldots, y \oplus r_n$ použitých jako náhodné bity.

2. Nechť $m, n > 1$ jsou celá čísla a $H \subseteq \{h : \{0,1\}^n \to \{0,1\}^m\}$ je 2-univerzální hašovací systém. Nechť $A \subseteq \{0,1\}^n$.

   (a) Pro $h \in H$ definujme množinu kolizí $C_h = \{(x_1, x_2) \in A^2; h(x_1) = h(x_2)\}$. Jaká je očekávaná velikost C_h, to jest $E_{h \in H}[|C_h|]$?

   (b) Dokažte, že když $2^m \geq 4|A|$, pak s pravděpodobností alespoň $1/2$ pro náhodné $h \in H$, alespoň polovina množiny A se zobrazí prostě při zobrazení h, to jest nekoliduje s jiným prvkem z A.

   (c) Dokažte, že když $2^m \geq |A|^2$, pak s pravděpodobností alespoň $1/2$ pro náhodné $h \in H$, množina A se zobrazí prostě při zobrazení h, to jest nenastává žádná kolize prvků z A. (Tomuto jevu se říká, že h *hašuje A perfektně*.)

   Řešení:

   (a) Jak si mnozí správně povšimli, chceme $C_h = \{(x_1, x_2) \in A^2; h(x_1) = h(x_2) \& x_1 \neq x_2\}$. Pro $(x_1, x_2) \in A^2$, kde $x_1 \neq x_2$, definujme náhodnou indikátorovou proměnnou C_{h,x_1,x_2}, která je jedna pro náhodně vybrané $h \in H$ právě když $h(x_1) = h(x_2)$ a jinak je nula. Zjevně $C_h = \sum_{(x_1,x_2)} C_{h,x_1,x_2}$. Tedy $E_{h \in H}[|C_h|] = \sum_{(x_1,x_2)} E_{h \in H}[|C_{h,x_1,x_2}|]$. Jelikož H je 2-univerzální hašovací systém, $\Pr_{h \in H}[h(x_1) = h(x_2)] = \sum_y \Pr_{h \in H}[h(x_1) = y = h(x_2)] = 2^m \cdot 2^{-2m} = 2^{-m}$. Tudíž $E_{h \in H}[|C_h|] = |A| \cdot (|A| - 1) \cdot 2^{-m}$.

   (b) Pokud $2^m \geq 4|A|$, pak $E_{h \in H}[|C_h|] \leq |A|/4$. Podle Markovovi nerovnosti, $\Pr_{h \in H}[|C_h| > |A|/2] < E_{h \in H}[|C_h|]/(|A|/2) \leq 1/2$. Pokud $|C_h| \leq |A|/2$, pak nutně alespoň polovina množiny A je zobrazena prostě. Z toho plyne požadované tvrzení.

   (c) Pokud $|C_h| < 2$, pak $C_h = \emptyset$, neboť každý pár x_1 a x_2 je počítaný dvakrát. Opět podle Markovovi nerovnosti, $\Pr_{h \in H}[|C_h| \geq 2] < 1/2$. Tvrzení je dokázáno.

3. Splnitelnost $SAT = \{\varphi; \varphi$ je splnitelná Booleovská formule$\}$ je dobře známý NP-úplný problém. O něco méně známý avšak taktéž NP-úplný je problém $SAT' = \{(\varphi, a, b, r); \varphi$ je Booleovská formule, pro kterou existuje splňující ohodnocení x, jenž zároveň splňuje $a \circ x \oplus b = r\}$. Zde $\circ$ je konvoluce definovaná během přednášky a vektory $a, b, r \in \{0,1\}^*$ jsou libovolné velikosti vhodné pro provedení konvoluce. Vektor x splňující dané vztahy je nazýván *svědek* pro (φ, a, b, r).

   Sestrojte pravděpodobnostní algoritmus, který dostane jako vstup Booleovskou formuli φ a zkonstruuje čtveřici (φ', a, b, r) takovou, že pokud φ není splnitelná, pak $(\varphi', a, b, r) \notin SAT'$ a pokud φ je splnitelná, pak s *rozumnou* pravděpodobností $(\varphi', a, b, r) \in SAT'$ a existuje právě jeden svědek pro (φ', a, b, r). Ukažte, že váš algoritmus splňuje toto zadání. (Hint: Použijte předchozí cvičení.)

   (Tomuto algoritmu se říká Valiant-Vazirani redukce ze SAT na $unique - SAT$. Ukazuje se tak, že rozpoznávat mezi formulí, která nemá žádné splnitelné ohodnocení a která má právě jedno splnitelné ohodnocení, je v principu skoro stejně těžké jako rozhodovat SAT. Rozmyslete si.)

   Řešení: Vezměme si splnitelnou Booleovskou formuli φ a označme jako $A \subseteq \{0,1\}^n$ množinu splňujících ohodnocení pro φ. Nechť m je celé číslo takové, že $2^{m-1} < 4|A| \leq 2^m$. Pokud vezmeme $H = \{h : \{0,1\}^n \to \{0,1\}^m, h(x) = a \circ x \oplus b, a \in \{0,1\}^{n+m-1}, b \in \{0,1\}^m\}$, což je 2-univerzální hašovací systém, pak podle předchozího příkladu část b), s pravděpodobností alespoň $1/2$ pro náhodné $h \in H$, alespoň polovina A je hašována bez kolize. Tedy pro takové h, a pro alespoň $|A|/2 > 2^m/16$ různých $r \in \{0,1\}^m$, existuje právě jedno $x \in A$ takové, že $h(x) = r$. Tedy pokud vezmeme náhodné $a \in \{0,1\}^{m+n-1}$, $b, r \in \{0,1\}^m$, pak s pravděpodobností alespoň $1/32$, existuje právě jedno splňující ohodnocení x formule φ, které zároveň splňuje $a \circ x \oplus b = r$. Tudíž náš pravděpodobnostní algoritmus na vstupu φ vybere náhodné $a \in \{0,1\}^{m+n-1}$, $b, r \in \{0,1\}^m$ a vystoupí (φ, a, b, r).

   Jenže! Jenže náš algoritmus nemůže vědět, kolik je m pro danou formuli φ. Ví pouze, že m je celé číslo mezi 2 a $n+2$. Tudíž m vybere taktéž náhodně a s pravděpodobností $1/n$ se trefí do správného m. Tudíž algoritmus nejprve náhodně zvolí m, pak a, b a r příslušných velikostí a poté vypíše (φ, a, b, r). Pokud φ je splnitelná, pak s pravděpodobností $1/32n$, vypsané $(\varphi, a, b, r) \in SAT'$ a existuje právě jeden svědek pro (φ, a, b, r). Pokud φ není splnitelná, pak $(\varphi, a, b, r) \notin SAT'$. Tudíž náš algoritmus dělá to, co jsme požadovali.