Korespondenční Seminář z Programování

38. ročník **KSP** **Červenec 2026**

Komentáře k páté sérii třicátého osmého ročníku KSP

38-5-S Hešování řetězců

Úkol 5: Izomorfismus – více dětí

Jedna z oblíbených hešovacích funkcí spočítala heš rodiče z kódů synů $x_1, \ldots, x_k$ jako:

$$\sum_{i=1}^{k} h(x_i) \bmod p$$

kde h je náhodná hešovací funkce do $\mathbb{Z}_p$, kterou jsme dostali v zadání.

Ačkoliv by výše popsaná hešovací funkce fungovala, tak dokázat to o ní se nikomu nepovedlo. Nabízí se totiž velmi jednoduchý (avšak falešný) důkaz jak omezit pravděpodobnost kolize:

Uvažme dva vrcholy s kódy synů $x_1, \ldots, x_k$ a $y_1, \ldots, y_\ell$:

$$\sum_{i=1}^{k} h(x_i) \equiv \sum_{i=1}^{\ell} h(y_i) \pmod p$$

$$\sum_{i=1}^{k} h(x_i) - \sum_{i=1}^{\ell-1} h(y_i) \equiv h(y_\ell) \pmod p$$

Zafixujeme-li $x_1, \ldots, x_k$ a $y_1, \ldots, y_{\ell-1}$, tak nám to dává jednu hodnotu, do které se $h(y_\ell)$ musí strefit, aby nastala kolize. A protože je h dokonale náhodná, tak pravděpodobnost kolize je $1/p$.

Nicméně tohle nefunguje. Problém je, že jednotlivé x_i a y_i nejsou nezávislé, a tím, že jsme zafixovali část z nich, jsme už mohli předurčit hodnotu $h(y_\ell)$. Např. když levý podstrom je třívrcholová cesta a pravý dvouvrcholová, tak při výpočtu heše levé cesty $h(h(0))$ jsme už použili hodnotu heše pravé $h(0)$. Obecně se jednotlivé podstromy mohou složitě překrývat, a proto není možné v důkazu používat nezávislost.

Daniel Skýpala

KSP pro vás připravují studenti Matematicko-fyzikální fakulty Univerzity Karlovy.
Realizace projektu byla podpořena Ministerstvem školství, mládeže a tělovýchovy.

Webové stránky: *https://ksp.mff.cuni.cz/*
E-mail: *ksp@mff.cuni.cz*
Organizátoři a kontakty: *https://ksp.mff.cuni.cz/kontakty/*