NTIN066 - solutions 8

David Mareček

April 05, 2023

1. Show that the family of all constant functions from $\mathcal{U}$ to $[m]$ is 1-independent.

2. Prove that the family of linear functions is not 3-independent.

```
m = p
keys: x, y, z
buckets: i, j, k

(ax + b) mod p = r
(ay + b) mod p = s

(a,b) maps to (r,s) (bijection)

(az + b) mod p = t

for (r,s) we can compute a,b,t

Pr(h(x) = i) = Pr(r = i) = 1/m
Pr(h(y) = j) = Pr(s = j) = 1/m
Pr(h(z) = k) = Pr(t = k) = 1 or 0

For i,j,k satisfying the equations, the probability is 1/m^2.
For i,j,k not satisfying the equations, the probability is 0.
```

3. Prove that tabulation hashing is 3-independent but not 4-independent.

If tabulation uses only one table, it is perfectly random, so it is k-independent for each k not exceeding the size of the universe.

Otherwise, we have $t \geq 2$ tables, each indexed by bits. Now consider x_1, x_2, x_3 and a_1, a_2, a_3 from the definition of 3-independence. At the same time, we imagine x_i as ordered k-tuples of b-bit values (parts indexing individual tables).

(a) If there is a position i in which all x_1, x_2, x_3 differ, then regardless of the values in the other positions, we can always fill in the table corresponding to this position so that the XORs come out a_1, a_2, a_3. Since T_i is perfectly random, the probability of generating a_1, a_2, a_3 is $1/m^3$.

(b) Or there exists (WLOG) a position i, position j, and some values A, B, C, D, for which:

$$T_i[A] \oplus T_j[C] \oplus v_1 = a_1$$
$$T_i[B] \oplus T_j[C] \oplus v_2 = a_2$$
$$T_i[A] \oplus T_j[D] \oplus v_3 = a_3,$$

where v1, v2, and v3 are XORed values from all the other tables. For each $T_j[C]$ and each v_1, v_2, v_3, this set of equations has just one solution. So the probability that filling of the tables leads to a solution is exactly $1/m^3$, as we need for 3-independence.

Counter-example for 4-independence: Imagine four different values A, B, C, D and four following keys x_1, x_2, x_3, x_4:

$$x_1 = AB,\ x_2 = AC,\ x_3 = DB,\ x_4 = DC.$$

Then for each a_1, a_2, a_3, a_4, we get the following four equations:

$$T_1[A] \oplus T_2[B] = a_1$$
$$T_1[A] \oplus T_2[C] = a_2$$
$$T_1[D] \oplus T_2[B] = a_3$$
$$T_1[D] \oplus T_2[C] = a_4$$

If we sum the equations, we get $a_1 \oplus a_2 \oplus a_3 \oplus a_4 = 0$. Thus, given the hash values of any of three of the keys, we can uniquely determine the fourth hash value. So when $a_4 = a_1 \oplus a_2 \oplus a_3$, then, if e.g. $T_1[A]$ is given, there is just one option how to fill $T_2[B]$, $T_2[C]$, and $T_1[D]$. These were filled independently, so $P = 1/m^3$, which is always higher than c/m^4 for $m > c$.