Задача 1-1 (30 баллов). Семейство $\mathcal{H}$ хеш-функций h, принимающих m значений, называется *C-универсальным*, если для произвольных различных ключей x и y справедливо $P[h(x) = h(h)] \leqslant C/m$. Рассмотрим семейство хеш-функций

$$h_a(x) = \lfloor ((ax) \bmod 2^w) / 2^{w-l} \rfloor,$$

определенных на w-битных ключах x ($0 \leqslant x < 2^w$) и дающих l-битные хеш-значения, где a — случайное нечетное число на отрезке $[1, 2^w]$. Покажите, что существует абсолютная константа C, такая что данное семейство является C-универсальным.

Решение. Обозначения:

- w — количество бит в ключах x, $0 \leqslant x < 2^w$.
- l — количество бит в хеш-значениях, $m = 2^l$.
- a — случайное нечетное число из множества $\{1, 3, 5, \ldots, 2^w - 1\}$ (всего 2^{w-1} чисел).

Идея доказательства:

Мы будем анализировать вероятность совпадения хеш-значений $h_a(x)$ и $h_a(y)$ при случайном выборе a. Для этого рассмотрим разность $s_x - s_y$, где $s_x = (ax) \bmod 2^w$ и $s_y = (ay) \bmod 2^w$. Заметим, что $s_x - s_y \equiv a(x - y) \pmod{2^w}$.

Совпадение хеш-значений $h_a(x) = h_a(y)$ эквивалентно тому, что верхние l битов s_x и s_y совпадают, то есть разность $s_x - s_y$ делится на 2^{w-l}.

Доказательство:

1. **Определим разность $d = x - y \neq 0$.**

   Поскольку $x \neq y$, то $d \neq 0$ и $d \in \{-2^w + 1, \ldots, 2^w - 1\}$.

2. **Выразим условие совпадения хеш-значений через a и d.**

   Хеш-значения совпадают, если

   $$h_a(x) = h_a(y) \quad \Leftrightarrow \quad \left\lfloor \frac{(ax) \bmod 2^w}{2^{w-l}} \right\rfloor = \left\lfloor \frac{(ay) \bmod 2^w}{2^{w-l}} \right\rfloor.$$

   Это эквивалентно тому, что

   $$\frac{(ax) \bmod 2^w}{2^{w-l}} - \frac{(ay) \bmod 2^w}{2^{w-l}} \in [0, 1).$$

   Таким образом,

   $$((ax) - (ay)) \bmod 2^w < 2^{w-l}.$$

3. **Преобразуем неравенство:**

   $$(ad) \bmod 2^w < 2^{w-l}.$$

   Это означает, что ad делится на 2^{w-l} при приведении по модулю 2^w, то есть

   $$2^{w-l} \mid (ad) \pmod{2^w}.$$

4. **Рассмотрим два случая в зависимости от d.**

   Случай 1: d чётное, то есть $v_2(d) \geqslant 1$, где $v_2(d)$ — показатель степени при разложении d на простые множители, соответствующий двойке.

   Пусть $v_2(d) = t \geqslant 1$, тогда $d = 2^t d'$, где d' — нечётное число.

   Случай 2: d нечётное, то есть $v_2(d) = 0$.

5. **Анализируем вероятность в обоих случаях.**

   Случай 1: $v_2(d) = t \geqslant 1$.

   Поскольку a — нечётное, $v_2(a) = 0$. Тогда $v_2(ad) = v_2(a) + v_2(d) = t$.

   Значит, ad делится на 2^t, но не на 2^{t+1}.

   Для того чтобы $2^{w-l} \mid ad$, необходимо, чтобы $w - l \leqslant t$, то есть $t \geqslant w - l$.

   Если $t \geqslant w - l$, то $v_2(ad) \geqslant w - l$, и неравенство выполняется.

   В этом случае $ad \bmod 2^w$ будет делиться на 2^{w-l}.

   Случай 2: $v_2(d) = 0$.

   Поскольку $v_2(a) = 0$, то $v_2(ad) = 0$. Следовательно, ad не делится на 2^{w-l}, и неравенство не выполняется.

6. **Вывод вероятности:**

   Случай 1: Когда $v_2(d) \geqslant w - l$, то $ad \bmod 2^w$ делится на 2^{w-l}.

   Однако, поскольку $v_2(ad) = t$, а $t \geqslant w - l$, то

   $$ad \bmod 2^w = 2^t k \pmod{2^w},$$

   где k — нечётное число.

   Количество таких a равно количеству нечётных чисел в $[1, 2^w - 1]$, то есть 2^{w-1}.

   Но поскольку d фиксировано, и a пробегает все нечётные числа, $ad \bmod 2^w$ будет принимать каждое значение, кратное 2^t, ровно один раз.

   Число значений a, для которых $ad \bmod 2^w$ делится на 2^{w-l}, равно

   $$N = \frac{2^{w-1}}{2^{t-(w-l)}} = 2^{l-1}.$$

   Здесь мы делим общее число нечётных a на количество возможных значений $ad \bmod 2^w$, которые делятся на 2^{w-l}.

   Таким образом, вероятность

   $$\mathsf{P}[h_a(x) = h_a(y)] = \frac{N}{2^{w-1}} = \frac{2^{l-1}}{2^{w-1}} = \frac{1}{2^{w-l}} = \frac{1}{2^l} = \frac{1}{m}.$$

   Случай 2: Когда $v_2(d) < w - l$.

   Тогда $ad \bmod 2^w$ не делится на 2^{w-l}, и неравенство не выполняется для любого a.

   Таким образом,

   $$\mathsf{P}[h_a(x) = h_a(y)] = 0 \leqslant \frac{1}{m}.$$