# Solution 9

1、**The Matrix Method**

- 假设输入key有$u$个bit，即$u = \lceil \log |U| \rceil$
- 哈希表的大小$M = 2^m$是2的整次幂
- 随机选择一个$m \times u$的$0-1$矩阵$A$（$A$中的每一个元素等可能取值$0/1$）
  - 定义哈希函数$h(\vec{x}) = A\vec{x}$，其中输入$\vec{x}$是一个$u$-bit向量，输出$h(\vec{x})$是一个$m$-bit向量
  - 将$h(\vec{x})$每一位对2取模，并转换回整数

证明：**The Matrix Method是Universal Hashing**（Hint：可以先固定$A$矩阵的一些位置，思考一下两个不同向量$\vec{x}$和$\vec{y}$的哈希值$h(\vec{x})$和$h(\vec{y})$是否相同由$A$矩阵的哪些位置决定；可以先考虑$m=1$的简化版本）

先考虑$m=1$的情况，对于任意$\vec{x} \neq \vec{y}$，$A\vec{x} = A\vec{y} \Leftrightarrow A(\vec{x} \oplus \vec{y}) = 0 \overset{\vec{z}=\vec{x}\oplus\vec{y}}{\Leftrightarrow} A\vec{z} = 0$

任取一位$\vec{z}_i \neq 0$上式等价于$A_i\vec{z}_i = \sum_{j\neq i} A_j\vec{z}_j$，在除了$A_i$以外的位置全部固定以后，等式右边固定，等式两边是否相等只取决于$A_i$的取值，概率为$\frac{1}{2}$，即$\Pr[h(\vec{x}) = h(\vec{y})] = \frac{1}{2}$

$\vec{x}$和$\vec{y}$有$m$个bits，每一个bit不同的概率都是$\frac{1}{2}$，于是有至少有一个bit不同的概率
$\Pr[h(\vec{x}) = h(\vec{y})] = \frac{1}{2^m} = \frac{1}{M}$

2、**The dot-product Method**

假定哈希表的大小$M$是一个质数

将每个输入$\vec{x}$写成一个$M$进制数：
$(x_1, x_2, \cdots, x_k)^T = \vec{x}, \text{ w/ each } x_i \in \{0, 1, \cdots, M-1\} \text{ and } k = \log_M |U|$

关于哈希函数，我们选择$k$个随机数$\vec{r} = (r_1, r_2, \cdots, r_k)^T \in \{0, 1, \cdots, M-1\}^k$，定义
$h(\vec{x}) = \langle \vec{x}, \vec{r} \rangle \mod M = r_1x_1 + r_2x_2 + \cdots + r_kx_k \mod M$

证明：**The dot-product Method是Universal Hashing**（Hint：思考一下**The dot-product Method**和**The Matrix Method**的相似之处，这可能可以为你的证明提供一些思路）

证明中所有运算都在$\mod M$意义下

对于任意$\vec{x} \neq \vec{y}$，$h(\vec{x}) = h(\vec{y}) \overset{\vec{z}=\vec{x}-\vec{y}}{\Leftrightarrow} \sum_{i=1}^k r_iz_i = 0$

任取一位$z_i \neq 0$，上式等价于$r_iz_i = \sum_{j\neq i} r_jz_j$，在除了$r_i$以外的位置全部固定以后，等式右边固定（记作$a$），等式两边是否相等只取决于$r_iz_i$的取值

$r_iz_i = a \Leftrightarrow r_i = az_i^{-1}$，$r_i$等可能取任意$[0, M-1]$，则$\Pr[h(\vec{x}) = h(\vec{y})] = \frac{1}{M}$。$M$是质数保证了$z_i$的乘法逆元$z_i^{-1} = z_i^{M-2}$存在（费马小定理）。

3、**2-wise Universal Hashing**

**2-wise Universal Hashing是k-wise Universal Hashing**在$k=2$时的特例

**The Matrix Method**稍作修改可以满足**2-wise Universal**

- 在**The Matrix Method**的基础上额外随机选择一个$m$维$0-1$向量$\vec{b} \in \{0,1\}^m$
- 定义哈希函数$h(\vec{x}) = A\vec{x} + \vec{b}$

(a) 证明：这种**The Matrix Method**的变种是**2-wise Universal Hashing**

还是先考虑$m=1$的情况，先固定$A$，可以得出：对于任意$x_1 \in U$，$\forall v_1 \in \{0, 1, \cdots, M-1\}$，有$\Pr[h(x_1) = v_1] = \frac{1}{2}$，这里的$\frac{1}{2}$完全来自于$b$的随机性，与$A$的取值无关。

$\Pr[h(x_1) = v_1 \wedge h(x_2) = v_2] = \Pr[h(x_1) = v_1] \cdot \Pr[h(x_2) = v_2 | h(x_1) = v_1] = \frac{1}{2} \Pr[h(x_2) = v_2 | h(x_1) = v_1]$

再使用第一题的结论，有$\Pr[h(x_2) = v_2 | h(x_1) = v_1] = \frac{1}{2}$

因此对于任意$x_1, x_2 \in U$，$\forall v_1, v_2 \in \{0, 1, \cdots, M-1\}$，有$\Pr[h(x_1) = v_1 \wedge h(x_2) = v_2] = \frac{1}{4}$

$m > 1$，利用行之间的独立性，有$\Pr[h(x_1) = v_1 \wedge h(x_2) = v_2] = \frac{1}{M^2}$

(b) **The Matrix Method**使用了$O(um)$个随机的$0-1$bit，更多的随机bit意味着哈希函数需要更多空间存储。一种变种可以减少随机bit的使用

- 用随机的$0-1$bit填充$A$矩阵的第一行和第一列，然后对于任意位置$i, j$，$(i > 1 \text{ and } j > 1)$，令$A_{i,j} = A_{i-1,j-1}$。
- 这样我们只需要使用$u + m - 1$个随机bit
- $\vec{b}$的生成方式不变

证明：这个变种依然是满足**2-wise Universal**

只考虑证明$h(\vec{x}) = A\vec{x}$是Universal Hash，剩下的证明和(a)相同。

取$z = x \oplus y$的第一个1，从$A$的第一行开始，每一行固定这一位右边的部分，那么每一行都有$\Pr[h(\vec{x}) = h(\vec{y})] = \frac{1}{2}$，总共$\Pr[h(\vec{x}) = h(\vec{y})] = \frac{1}{2^m} = \frac{1}{M}$