6.897
L10.5
Mar. 8, 2005

Approximate sketch(x) (on word RAM)

- don't need sketch to pack the b_i bits right next to each other
- can be spread out in predictable pattern (indep. of x) of size $O(w^{4/5})$

Idea: use multiplication after masking $x' = x$ AND $\left(\sum_{i=0}^{r-1} 2^{b_i}\right)$

$$x' \cdot m = \left(\sum_{i=0}^{r-1} x_{b_i} 2^{b_i}\right) \cdot \left(\sum_{j=0}^{r-1} 2^{m_j}\right)$$
$$= \sum_{i=0}^{r-1} \sum_{j=0}^{r-1} x_{b_i} 2^{b_i + m_j}$$

Claim: can choose $m_0, m_1, \ldots, m_{r-1}$ (in terms of $b_0, b_1, \ldots, b_{r-1}$) such that
- $b_i + m_j$ are all distinct ($\Rightarrow$ no collision)
- $b_0 + m_0 < b_1 + m_1 < \cdots < b_{r-1} + m_{r-1}$ (preserve order)
- $(b_{r-1} + m_{r-1}) - (b_0 + m_0) = O(r^4) = O(w^{4/5})$ (small span)

$\Rightarrow$ approx. sketch$(x) = \left((x \cdot m) \text{ AND } \sum_{i=0}^{r-1} 2^{b_i + m_i}\right) \gg (b_0 + m_0)$

(discard $i \neq j$)

Proof: ① choose $m'_0, m'_1, \ldots, m'_{r-1} < r^3$ such that $b_i + m'_j$ are all distinct modulo r^3
- pick $m'_0, m'_1, \ldots, m'_{i-1}$ by induction
- m'_i must avoid $m'_i + b_j - b_k \quad \forall i, j, k$

(m'_i: t choices; b_j: r; b_k: r)

$tr^2 < r^3$ possibilities

$\Rightarrow$ choice for m'_i exists

② let $m_i = m'_i + (w - b_i$ rounded down to previous multiple of $r^3) + ir^3$

$\equiv m'_i \pmod{r^3}$

$\Rightarrow m_i + b_i$ in r^3 interval after $\left(\frac{w}{r^3} + i\right)$th multiple of r^3

$\Rightarrow m_0 + b_0 < m_1 + b_1 < \cdots < m_{r-1} + b_{r-1}$

$m_0 + b_0 \approx w$; $m_{r-1} + b_{r-1} \approx w + r^4$ ← difference $= O(r^4)$