Hashing

- In CS there are many applications where a large set S needs to be mapped to a small image-set T, by a map φ. (easy to compute?)
  e.g. shortening of long URLs,
  mapping long addresses to shorter virtual address,
  labelling files, encryption/signature, etc

- Think of φ as a rnd function in $\{\varphi_r \mid r \in R\}$.
  Denote random variable by φ_R.

- **Defn:** Hashing $\phi_R : S \to T$ is called pairwise independent (p.i.) if:

(i) Rnd. variables $\{\phi_R(s) \mid s \in S\}$ are p.i.

(ii) $\forall s \in S$, $\phi_R(s)$ is uniformly distributed in T.

- **Exercise:** (i) $\iff \forall s \neq s' \in S$, $\forall t, t' \in T$,

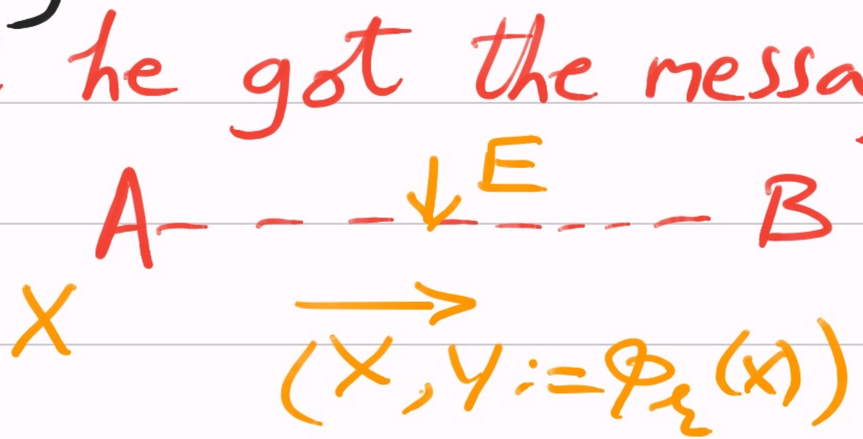
$$P(\phi_R(s) = t \wedge \phi_R(s') = t') = (1/|T|)^2.$$

(ii) $\iff \forall s \in S$, $t \in T$, $P(\phi_R(s) = t) = 1/|T|$.

- Intuitively, rnd. ϕ_R maps any subset $A \in \binom{S}{|T|}$ to T in "a 1-1 way"!

▷ Clearly, it fails if $|A| > |T|$. (Collisions)

- Let's see a cryptographic eg.
Suppose A(lice) & B(ob) communicate in a channel that's hijacked by E(ve).
How can B be sure that he got the message from A (& not by E)?



Protocol:

0) A & B keep a secret:
rnd. key $r \in R$. (i.e. hash fn. Φ_r)

1) A sends msg X with signature $Y := \Phi_r(X)$.

2) B accepts it, only after checking: $Y \stackrel{?}{=} g_2(X)$

Analyse: • Suppose E steals (X, Y) & instead sends (X', Y') on the channel to B.

• Let $\mathcal{E}$ be the event: $Y' = g_2(X')$; in which case B wrongly accepts msg $X' \neq X$.

(Fix X', Y'!)

$$\cdot P(\mathcal{E}) = \sum_{\substack{s \in S \\ t \in T}} P(X=s \wedge Y=t \wedge \mathcal{E}) = \sum_{s,t} P(X=s) \cdot P(g_2(s)=t \wedge g_2(X')=Y' \wedge X' \neq s)$$

$$\leq \sum_t \sum_s P(X=s) \cdot \frac{1}{|T|^2} = \sum_t \frac{1}{|T|^2} = \frac{1}{|T|}.$$

$\Rightarrow$ Chances of an evil collision are very small (if T is large)!

Implementation of Hash

- Idea – φ is a linear transformation from vector space S to space T.
  ($\mathbb{F}_2$:= field on $\{0,1\}$.)
- Let $S := \mathbb{F}_2^n$, $T := \mathbb{F}_2^m$ & $n > m \geq 1$.
- φ_R is rnd. matrix $R \in \mathbb{F}_2^{m \times n}$ s.t. $\varphi_R : S \to T$, $s \mapsto R \cdot s =: t$ (s: column vecs.)

Exercise: Show that φ_R is p.i. hashing!