

Birthday Attack

Messaggi P_i di lunghezza fissa b-bit, sono in tutto $|P_i| = 2^b$

Codici Hash H_i di lunghezza fissa a-bit, sono in tutto $|H_i| = 2^a$

Che probabilita' c'e' di $H_i = H_j$, e cioe' di "collisione" per $P_i \neq P_j$, al variare di a?

Devo calcolare i codici hash H_i per un numero grande di prove k , $1 \leq i \leq k$, metterli in ordine ascendente (decrescente) e verificare se ce ne sono due uguali corrispondenti a due prove diverse: $H_i = H_j$ per $i \neq j$, $1 \leq i, j \leq k$.

Analogia del calcolo combinatorio delle "disposizioni" (conta l'ordine) con e senza ripetizione di n oggetti presi a gruppi di k .

I "gruppi" sono le k prove che faccio $(1, 2, 3, \dots, k)$ ordinate nel tempo.

Gli "oggetti" sono $n=2^a$ e cioe' il numero totale dei possibili codici hash. In genere $k \leq n$.

Quindi, n sono i vari codici hash, pensiamo a delle "caselle" (box) ciascuna marcata col suo codice hash, k sono le prove che faccio, sono delle palle numerate da 1 a k che lancio a caso nel mucchio di caselle. La "collisione" avviene se alla fine del lancio delle k palle, trovo almeno due palle nella stessa casella.

Tutte le possibili disposizioni delle palle nelle varie caselle alla fine dei k lanci sono:

- n^k , disposizioni con ripetizione, se ci sono collisioni (posso avere una, due o piu' palle nella stessa casella).
- $n!/(n-k)!$, disposizioni senza ripetizione, se non ci sono collisioni, ovvero tutte le k palle finiscono in una casella diversa.

Quindi, posso scrivere agevolmente:

$$P \{ \text{nessuna collisione} \} = \frac{\# \text{ eventi senza collisione}}{\# \text{ eventi totali}} = \frac{n!}{(n-k)!} / n^k = \frac{(n-1)!}{n^{k-1} (n-k)!} = (1-1/n)(1-2/n)(1-3/n) \dots [1-(k-1)/n] =$$

$$= \prod_{i=1}^{k-1} (1-i/n) < \prod_{i=1}^{k-1} e^{-i/n} = e^{-k(k-1)/2n} \quad (\text{infatti: } e^{-x} > 1-x, \text{ e cioe': } 1-x < e^{-x})$$

$$\text{E cioe': } P \{ \text{almeno una collisione} \} = [1 - P \{ \text{nessuna collisione} \}] > 1 - e^{-k(k-1)/2n}$$

Se voglio, ad esempio, $P \{ \text{almeno una collisione} \} > 0,5$ (50%), allora deve essere

$$k \approx 1,17 \sqrt{n} \quad (*)$$

ove $n=2^a$, e se $n=2^{56}$, e cioe' se il codice hash e' lungo come la chiave DES 56 bit, si ha $k \approx 2^{28} \approx 10^8$, e cioe' un numero di prove (codifiche hash di messaggi) uguale a 100 milioni, un numero piccolo!

Pertanto, la lunghezza dei codici hash deve essere perlomeno di 128 bit per avere un numero di prove $k \approx 2^{64} \approx 10^{19}$. E' il caso di MD5, mentre in SHA-1 si tratta di 160 bit con un numero di prove $k \approx 2^{80} \approx 10^{24}$.

(*) Nel "paradosso del compleanno" si ha $n=365$ [gg/anno], e si ricava $k \approx 23$ [invitati al birthday party], per avere una probabilita' $> 50\%$ di trovare almeno due invitati nati nello stesso giorno.