CS124 Lecture 15 — Monday, March 30, 20

RSA

If we can factor → we can break RSA.

Extended Euclid's Algorithm

gcd(a,b) but also ints x, y s.t. $ax+by=d$

↳ multiplicative inverses

$ax+by=1$

```
EE(a,b):                                   (gcd, x, y)
    if b=0 : return (a, 1, 0)
    a = b*k + (a mod b)
    (d, x, y) = EE(b, a mod b)
    return (d, y, x - k*y)
```

$bx - (a \bmod b)y = d$ (−bky, +bky)

$ay + (x - ky)b = d$

$y*(a \bmod b + bk)$

a	b	k	x	y
313	37	8	−13	110
		8×37=296	6	−13
37	17	2	−1	6
17	3	5	1	−1
3	2	1	0	1
2	1	2	1	0
1	0			

17 = 313 mod 37

1 = gcd

y → x

x − ky → y

Claim: $313(-13)+37(110)=1$

Duplicates

AltaVista – problem w/ duplicate documents on the web

- people don't want the same page in search results
- storage problem: no use storing duplicates / near duplicates

64-bit hashes → look for collisions?

doesn't work for near-duplicate documents

hash value should be same or nearly the same

document = set of numbers

set A, set B

Resemblance / Jacard coefficient

$$R(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

$$0 \leq R(A,B) \leq 1$$

0 ↑ disjoint

1 ↑ exactly same set

$O(n^2)$ → almost method

$O(n \log n)$ → sorting and comparing

$O(n)$ → hash table and look for collisions

Method to Estimate Resemblance

calling card set (sketch)

$\pi: \{0,1\}^{64} \to \{0,1\}^{64}$

Black Box $(1, x) = \pi_1(x)$

Black Box $(50, y) = \pi_{50}(y)$

On 4-bit integers: $\pi : [0, 15] \to [0, 15]$

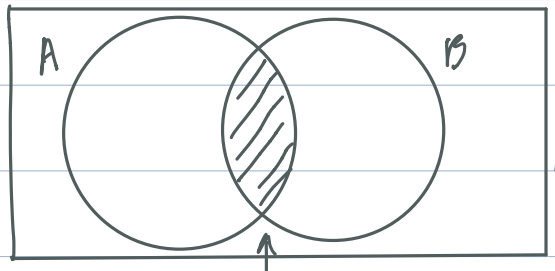
x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\pi(x)$	9	2	14	11	6	3	7	8	15	10	4	13	12	0	1	5

Set $A = \{3, 5, 11\} \longrightarrow \pi_1(A) = \{11, 3, 13\}$

[calling card: $\min(\pi_1(A))$, $\min(\pi_2(A))$, $\min(\pi_3(A))$, ...
$\min(\pi_1(B))$, $\min(\pi_2(B))$, ...

constant-size

$$\text{Prob}[\min(\pi_1(A)) = \min(\pi_1(B))] = \frac{|A \cap B|}{|A \cup B|} = R$$



$\Pr[\min \pi_1(A \cup B) = \min \pi_1(A \cap B)]$

min (union) for sum to be =

Estimate resemblance: $\frac{\text{count \# of matches}}{\text{\# permutations}} \approx R$

$E[\text{\# matches}] = R \times \text{\# permutations}$

Documents → Sets

4-shingling

Four score and seven years ago

hash hash hash

Subsample: only take hash values = 0 mod 8

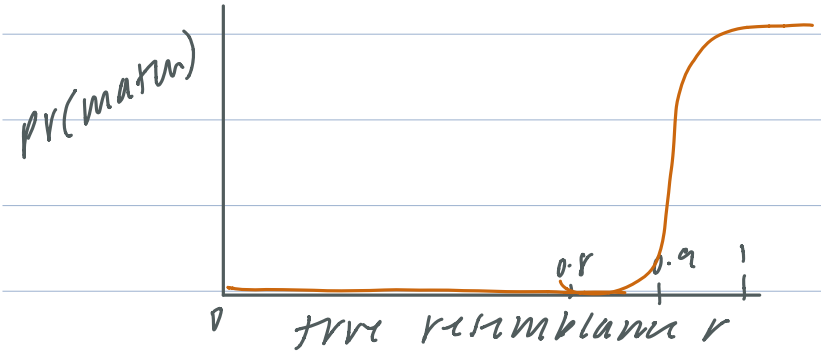
Working in Practice

similar if estimated resemblance ≥ 9

true resemblance = r

Pr[match ≥ 90 matches in 100-tn calling card]

$$= \sum_{k=90}^{100} \binom{100}{k} r^k (1-r)^{100-k}$$



r = 0.05 → 10^{-18}

r = 0.95

78.8% time similar

2-SAT

$$(x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_3) \wedge (x_1 \vee x_2) \wedge (x_4 \vee \bar{x}_3) \wedge (x_4 \vee \bar{x}_1)$$

SAT: normally NP-complete

2SAT: poly-time

Randomized Algo (slower than deterministic)

start w/ randomized assignment

while ∃ unsatisfied clause until tired

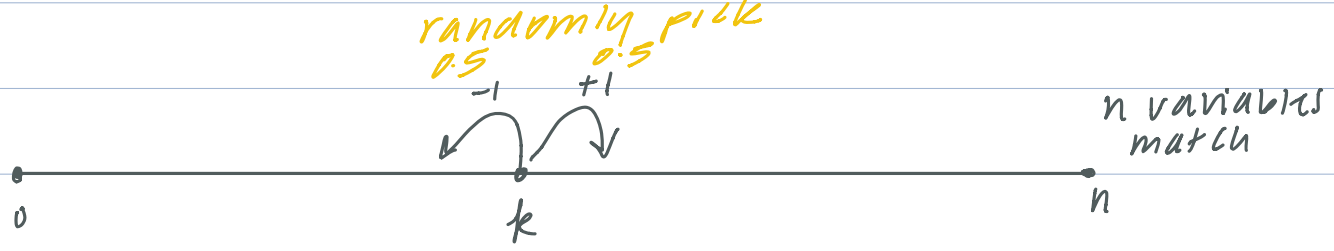
flip a variable at random in an unsatisfied clause

return "unsatisfiable" or truth assignment

A = assignment

S = "star" - truth assignment

matches S, A



Random Walk on a Line

- Gambler's Ruin Player