Criptografia: les matemàtiques de la informació secreta

Xevi Guitart

Departament de Matemàtiques i Informàtica
Universitat de Barcelona

Breu introducció històrica

- Criptografia ve del grec krypto (amagar) i grapho (escriure).
- L'objectiu és transmetre missatges de manera privada, que només siguin comprensibles pel destinatari autoritzat (confidencialitat).
- Durant molt de temps, pràcticament els únics usuaris d'aquestes tècniques eren els militars i els governs.
  - Els grecs ja feien servir tècniques de xifrat (segle III a.C.)
  - Juli Cèsar ($\sim$ 45 a.C.) feia servir un mètode que avui en dia es coneix com "el xifrat del Cèsar"
  - Al-Kindi, un matemàtic àrab (segle VII d.C.) fou el primer a "trencar" el xifrat del Cèsar.
  - Va jugar un paper molt important en la segona guerra mundial (màquina Enigma).

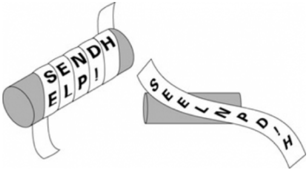
Breu introducció històrica

- Als anys 70 del segle passat hi ha una revolució en la criptografia:
  - Ús creixent de la informàtica i les comunicacions digitals:
    - demanda de serveis criptogràfics per part de la societat civil
    - entitats financeres i empreses en general
  - 1976: invenció de la criptografia de clau pública per Whitfield Diffie i Martin Hellman, matemàtic i enginyer electrònic americans.

  - 1978: Rivest, Shamir i Adelman inventen el criptosistema RSA
    - Àmpliament utilitzat avui en dia a internet

Alguns xifrats històrics: l'escítal

- Utilitzat pels espartants $\sim$ 400 a.C.
- Un bastó on s'hi enrotlla una tira de paper, i s'hi escriu el missatge



- Quan desenrotllem el paper, les lletres queden desordenades (es fa una permutació) i no s'entén el missatge
- Per a desxifrar-lo, l'enrotllem en un bastó...del mateix diàmetre!
- Emissor i receptor han d'haver acordat abans el diàmetre
    - Diem que el diàmetre és la clau privada

Un altre xifrat de transposició senzill

- Escrivim el missatge en una matriu, omplint per files

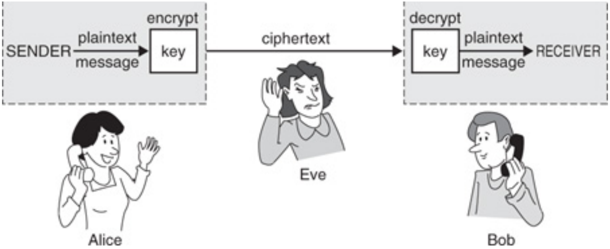
$$\begin{matrix} A & T & A & Q \\ U & E & U & A \\ L & A & T & A \\ R & D & A & \end{matrix}$$

- i llegim per columnes; el missatge xifrat seria

AULRTEADAUTAQAA

- Per a desxifrar, omplim la matriu per columnes i llegim per files.
- La clau privada és la mida de la matriu (4×4 en aquest cas)
  - També l'han d'acordar prèviament, i mantenir en secret.

El xifrat del Cèsar



- L'Alice i en Bob es reuneixen prèviament i acorden substituir cada lletra per la que està 3 posicions més endavant a l'alfabet
- Així si el missatge és HOLA, l'Alice el xifra i transmet KROD.
- En Bob rep KROD, i per desxifrar-lo substitueix cada lletra per la que està 3 posicions més enrera a l'alfabet, obtenint HOLA.
- L'Eve intercepta KROD, i no sap el seu significat.
- La clau k és el nombre de posicions que tirem a la dreta.
- Diuen que aquest sistema, amb $k = 3$, era utilitzat per Juli Cèsar.
- Aquest és un xifrat de substitució.

Xifrat de substitució i aritmètica modular

- L'alfabet català té 26 lletres, identifiquem lletres amb nombres:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

- Un missatge és una cadena de nombres: HOLA $\rightarrow$ 7 14 11 0
- L'Alice i en Bob acorden una clau secreta, per exemple $d = 20$
- Per xifrar sumem 20 al nombre associat a cada lletra
  - Compte! Si el resultat dóna ≥ 26, hem de restar-li 26
  - 7 14 11 0 $\rightarrow$ 1 8 5 20
  - HOLA $\rightarrow$ BIFU
- Per desxifrar, a cada lletra li hem de restar 20.
  - Compte! Si el resultat dóna ≤ 0, aleshores li hem de sumar 26
- A aquesta manera de sumar i restar, en què volem portar els resultats a un nombre entre 0 i 25, se li diu aritmètica mòdul 26
- No és res nou per a nosaltres:
  - Quan sumem o restem hores ho fem mòdul 12
  - Quan sumem o restem angles, ho fem mòdul 360

Com podem trencar el xifrat del Cèsar?

1. Força bruta: hi ha 26 claus possibles, les podem provar totes...
2. Hi ha una manera millor, que és la que va descobrir Al-Kindi
    - Interceptem el missatge YRF HEARF NEEVORA N YRF FRG
    - No totes les lletres són igual de freqüents: la més freqüènt és la E.
    - En un text xifrat llarg, és probable que la lletra que aparegui més cops sigui la que es correspon a la E.
    - La R apareix 5 cops: intuïm que la $E = 4$ s'ha xifrat com $R = 17$.
    - Si desxifrem amb la clau $17 - 4 = 13 = N$ obtenim:
      LES URNES ARRIBEN A LES SET
    - Altres pistes: Dígrafs habituals (RR, SS, QU, etc.)

- Aquests atacs es poden solucionar, amb els anomenats sistemes de subsitució polialfabètica. Ara en veurem un exemple.
- El problema encara hi és: Alice i Bob han d'acordar una clau

El xifrat de Vigenere

- Com la substitució simple, però utilitzem més d'una clau
  - podem pensar que triem una paraula clau
- Exemple: paraula clau "COSA" (2 14 18 0)
- Missatge: Ataqueu a les dotze

text clar	0	19	0	16	20	4	20	0	11	4	18	4	14	19	25
clau repetida	2	14	18	0	2	14	18	0	2	14	18	0	2	14	18
text xifrat	2	7	18	16	22	18	12	0	13	18	10	4	16	7	17

- Missatge xifrat: chsqwsmanskdqhre
- Dificulta l'atac per força bruta: hi ha $26^4 = 456\,976$ possibles claus
- Dificulta l'anàlisi de freqüència: la primera *A* es xifra sumant 2, i la segona *A* es xifra sumant 18.
- El problema encara hi és: Alice i Bob han d'acordar una clau

Criptosistemes de clau secreta

Fenomen general

Com més llarga és la clau, més segur és el criptosistema.

- La màquina Enigma, o altres màquines de rotors, utilitzades durant la segona guerra mundial, eren dispositius mecànics que realitzaven substitució polialfabètica de període molt llarg.

Criptosistemes de clau secreta

- Avui en dia en comunicacions digitals
  - criptosistemes de clau secreta (combinen substitució i permutació)
  - AES: Advanced Encryption Standard (utilitzat a internet)
- Com es fa per a compartir de manera segura la clau secreta?
  - Aquí és on entren en joc els criptosistemes de clau pública, descoberts el 1976 per Diffie–Helman.
  - Veurem el criptosistema RSA, el més utilitzat avui en dia

El criptosistema RSA

- Idea: hi ha dues claus, una per a xifrar i l'altra per a desxifrar
  - La clau per a xifrar la sap tothom, és pública
  - La clau per a desxifrar, només la sap qui rep el missatge
- Treballem amb aritmètica modular, mòdul un nombre que sigui el producte de dos primers molt grans
  - nombres primers: $2, 3, 5, 7, 11, 13, 17, \ldots$
  - Quants nombres primers hi ha? N'hi ha infinits!
  - 1235324553999999981554151456773 és un primer de 30 xifres
- Per al sistema RSA cal:
  - Escollir dos primers molt grans: p i q
  - Calcular $N = p \cdot q$
  - Fer tots els càlculs amb aritmètica mòdul N
- Farem un exemple de joguina amb $p = 5$, $q = 11$, $N = 55$

El criptosistema RSA: un exemple de joguina

- Alice vol enviar un missatge a Bob. Aleshores Bob fa el següent:
  - Escull dos primers, per exemple $p = 5$ i $q = 11$. Calcula $N = 55$.
  - Fa públic el nombre 55, per ex. posant-lo a la seva web.
- L'Alice veu que la clau pública d'en Bob és $N = 55$.
  - Un missatge són nombres mòdul 55, és a dir nombres entre 0 i 54.
  - Suposem que el missatge és $m = 6$.
  - El missatge xifrat és $m^3 \pmod{55}$.
    - ⋆ $6^3 = 216$, restant 55 tants cops com calgui perquè caigui entre 0 i 54.
    - ⋆ Fem la divisió entera: $216 = 3 \cdot 55 + 51 \rightsquigarrow$ envia $c = 51$.
- Bob rep $c = 51$. Per a desxifrar, ha de fer l'arrel cúbica mod 55.
  - En aritmètica modular, es pot calcular l'arrel cúbica elevant a un altre nombre. Pel cas de $N = 55$, cal elevar a $d = 27$.
  - Si fem 51^{27} i restem múltiples de 55 fins a caure entre 0 i 54 obtenim 6, que és el missatge original.
- Per què això és segur?
  - Per desxifrar el missatge, cal saber calcular arrels cúbiques mod N. Què evita que Eve pugui prendre una arrel cúbica mod N?
  - Si N és molt gran, només es coneixen mètodes ràpids per a calcular arrels cúbiques si en coneixem els dos factors primers.

RSA: exemple real

$p = 5311520355088381732434640236485900354436072788334240607719207689566489176525$
$9588109185123498599491128573613382194379700775829505261007423053350140642582 31$

$q = 5009685181411005100125133414437627030624530352244501699703092725010366500853$
$599762826993291441120385661185616114031477999669207674977753225026374348245 1$

- $N = p \cdot q$ té unes 300 xifres ($N \simeq 10^{300}$).
- Quant de temps trigaríem a factorizar-lo?
- Hauríem provar unes 10^{150} operacions...
- Un ordinador pot fer unes 10^{11} operacions per segon...
- Trigariem uns 10^{139} segons, que són $\sim 10^{131}$ anys...
- L'edat de l'únivers és de 13×10^9 anys!
- Això seria fent el mètode naïf, però hi ha mètodes més ràpids:
    - Récord actual: factoritzar un producte de dos primers de 232 dígits (equivalent a 2000 anys de càlculs en un processador a 2.2GHz)

Conclusions

- La seguretat a internet es basa en què no es coneix cap mètode per a factoritzar enters grans en un temps raonable.
- Se sabria fer amb un ordinador quàntic potent.
    - S'està treballant en construir ordinadors quàntics, però poden passar anys abans no es tinguin.
- Els matemàtics busquen nous xifrats que no es puguin trencar ni tan sols amb un ordinador quàntic...

Per a saber-ne més

- Los códigos secretos, Simon Singh.
  - Llibre d'història de la criptografia. Cobreix tot el que hem vist i molt més de manera amena i entenedora.
- https://www.simonsingh.net/The_Black_Chamber/
  - Pàgina web on podeu jugar a encriptar i desencriptar missatges. Té molts xifrats, els que hem vist i alguns que no...
- The imitation game (Descifrando Enigma)
  - pel·lícula del 2014 on s'explica la història d'Alan Turing i com va trencar el xifrat d'enigma.
- Videos sobre la màquina enigma de James Grime (googlejar: numberphile enigma)
- És un bon tema per a un treball de recerca!