Information, Calcul et Communication

Correction d'erreurs : principes de base

Olivier Lévêque

École polytechnique fédérale de Lausanne

Comment transmettre des données ?

Il existe deux moyens de transmission :

- A travers le **temps** : enregistrement sur un support
  (et les données peuvent lues plus tard)
- A travers **l'espace** :
  - par câble électrique ou fibre optique (téléphonie, internet)
  - dans l'air (téléphonie sans fil, wifi)

Problème commun : des erreurs surviennent régulièrement !

- lors de l'écriture ou de la lecture de données
- lors de la transmission par câble électrique, fibre optique ou onde électromagnétique

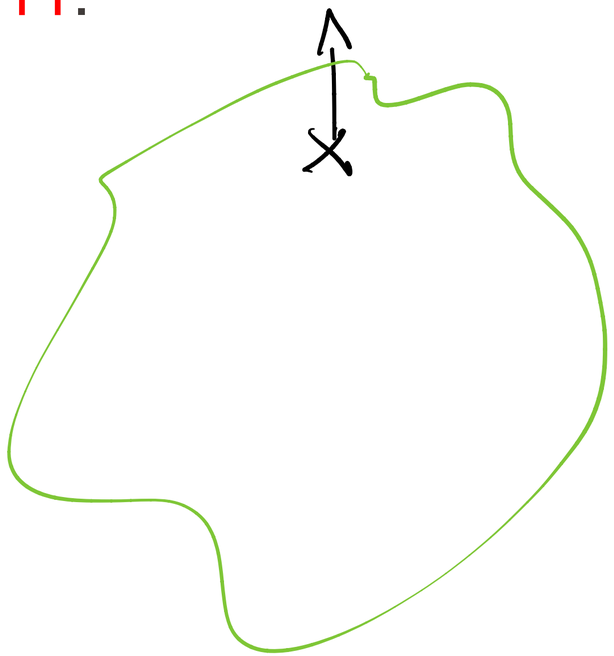
Exemple

- Vous désirez communiquer la direction à prendre (N,S,E,W) à un ami perdu.

Code binaire utilisé :

N	S	E	W
11	10	01	00

- Vous indiquez à votre ami d'aller au nord en envoyant 11.



Exemple

- Vous désirez communiquer la direction à prendre (N,S,E,W) à un ami perdu.

Code binaire utilisé :

N	S	E	W
11	10	01	00

- Vous indiquez à votre ami d'aller au nord en envoyant 11.
- Si votre ami reçoit :

11 Tout se passe bien : votre ami se dirige vers le nord.

1? Un **effacement** survient : votre ami ne sait pas s'il doit se diriger vers le nord ou vers le sud.

10 Une **erreur** survient : votre ami se dirige alors vers le sud !

Exemples de redondance dans la vie courante

- Un père à ses enfants :

« Mettez vos chaussures… mettez vos chaussures, j'ai dit ! »

- Vous à votre ami :

« Pour venir chez moi, c'est simple :
tu tournes dans la 2e rue à droite après le feu rouge, juste après la station-service ;
j'habite au numéro 3, dans l'immeuble bleu avec les grands balcons »

Les informations en rouge sont redondantes (donc inutiles en théorie…).

Corriger un effacement

- Codage par répétition

N	S	E	W
1111	1100	0011	0000

  - 11?1 → N
  - Simple mais coûteux : il faut envoyer 4 bits au lieu de 2 !

2e idée :

N S E W

1 1 0 1 0 1 0 0 1 1 0 0 0

somme modulo 2

Si on reçoit

1?0 → N

?01 → S

1 ?? → N/S ?

Corriger un effacement

- Codage par répétition

N	S	E	W
1111	1100	0011	0000

  - 11?1 → N
  - Simple mais coûteux : il faut envoyer 4 bits au lieu de 2 !

- Bit de parité

N	S	E	W
110	101	011	000

  - 1?0 → N
  - Le dernier bit correspond à la somme modulo 2 des deux premiers.

Pour envoyer un message de n bits et corriger un effacement, il faut ajouter :

$$n\ bits \quad | \quad 1\ bit\ de\ parité$$

La taille finale du message est :

$$2n\ bits \quad | \quad n+1\ bits$$

Corriger une erreur

- Doubler chaque bit
  - 1101 → N ou S ??

N	S	E	W
1111	1100	0011	0000

- Tripler chaque bit
  - 111101 → N

N	S	E	W
111111	111000	000111	000000

règle de la majorité : 111 | 101

1 1 → N

Corriger une erreur

- Doubler chaque bit
  - 1101 → N ou S ??

N	S	E	W
1111	1100	0011	0000

- Tripler chaque bit
  - 111101 → N

N	S	E	W
111111	111000	000111	000000

On applique ici la **règle de la majorité** : 111101 → N 111100 → S

Mais à nouveau, cette solution est très coûteuse :

Chaque bit est triplé, on aura donc besoin de $3n$ bits pour envoyer un message de n bits à l'origine !

$$C = \{\overset{S}{00000}, \overset{N}{11111}\}$$

message reçu $\longrightarrow$ 10001

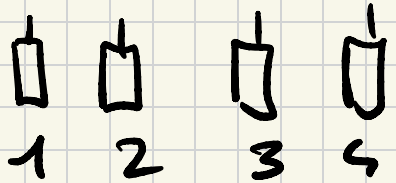
règle de la majorité

$\longrightarrow$ message envoyé = S

ceci fonctionne dans ce cas

s'il n'y a pas plus de 2 erreurs.

rats et bouteilles



rat A rat B

rat A goûte un mélange de 1 et 2

rat B goûte un mélange de 1 et 3

Corriger une erreur

Un meilleure méthode : ajouter *plusieurs* bits de parité

Voici un essai :

- Lorsque $n = 4$: on ajoute 2 bits de parité
  - le premier indique la parité de la somme des bits 1 et 2
  - le deuxième indique la parité de la somme des bits 1 et 3

$(1+1) \bmod 2$

Exemple : Pour envoyer 1101, on envoie 110101.

$(1+0) \bmod 2$

- Si on reçoit 100101 :
  - bit 1 et 2 : $1 + 0 = 1 \neq 0$
  - bit 1 et 3 : $1 + 0 = 1$

} On déduit que le bit 2 est faux :
Le message d'origine est donc 110101

Corriger une erreur (suite)

Un meilleure méthode : ajouter *plusieurs* bits de parité

Voici un essai :

- Lorsque $n = 4$: on ajoute 2 bits de parité
  - le premier indique la parité de la somme des bits 1 et 2
  - le deuxième indique la parité de la somme des bits 1 et 3

Problème : Pour envoyer 1101, on envoie toujours 110101.

- Si l'erreur survient sur le bit 4, on reçoit alors 110001, mais les deux bits de parité sont corrects → erreur indétectée !

• si une erreur survient sur un des bits de parité (5 ou 6)

Corriger une erreur (suite)

Un meilleure méthode : ajouter *plusieurs* bits de parité

Voici un essai :

- Lorsque $n = 4$: on ajoute 2 bits de parité
  - le premier indique la parité de la somme des bits 1 et 2
  - le deuxième indique la parité de la somme des bits 1 et 3

Problème : Pour envoyer 1101, on envoie toujours 110101.

- Si l'erreur survient sur le bit 4, on reçoit alors 110001, mais les deux bits de parité sont corrects → erreur indétectée !

- Pour réparer ce problème, *3 bits de parité* sont en fait nécessaires → code de Hamming

Code de Hamming

16 messages possibles

message à envoyer $x_1 x_2 x_3 x_4$ 4 bits

→ on rajoute 3 bits de parité :

$$x_5 = x_1 \oplus x_2 \oplus x_3$$

$$x_6 = x_1 \oplus x_2 \oplus x_4$$

$$x_7 = x_1 \oplus x_3 \oplus x_4$$

→ Ce code corrige une erreur.

Corriger plusieurs effacements ou erreurs ?

Tout l’art de la théorie du codage (70 ans d’histoire…) consiste à choisir parcimonieusement les bits de parité pour corriger un nombre maximum d’erreurs…

Correction d'erreurs

Déf : • code binaire $C = \{c^{(1)}, \ldots, c^{(M)}\}$

$c^{(j)}$ = mot de code de longueur n

• distance de Hamming : (tous)

$$d(c, c') = \#\{1 \leq i \leq n : c_i \neq c_i'\}$$

• distance minimale d'un code binaire C :

$$d = \min_{c, c' \in C ; c \neq c'} d(c, c')$$

3 paramètres importants :

- M = nombre de mots de code dans C
  ($\leftrightarrow$ quantité d'information)
- n = longueur des mots de code
  ($\leftrightarrow$ quantité de redondance à ajouter)
- d = distance minimale de C
  ($\leftrightarrow$ capacité de correction)

N S E W

Ex 1 : $C = \{11, 10, 01, 00\}$

$M = 4$, $n = 2$, $d = 1$ (NB : $M \leq 2^n$)

capacité de correction nulle

Ex 2 : $C = \{1111, 1100, 0011, 0000\}$

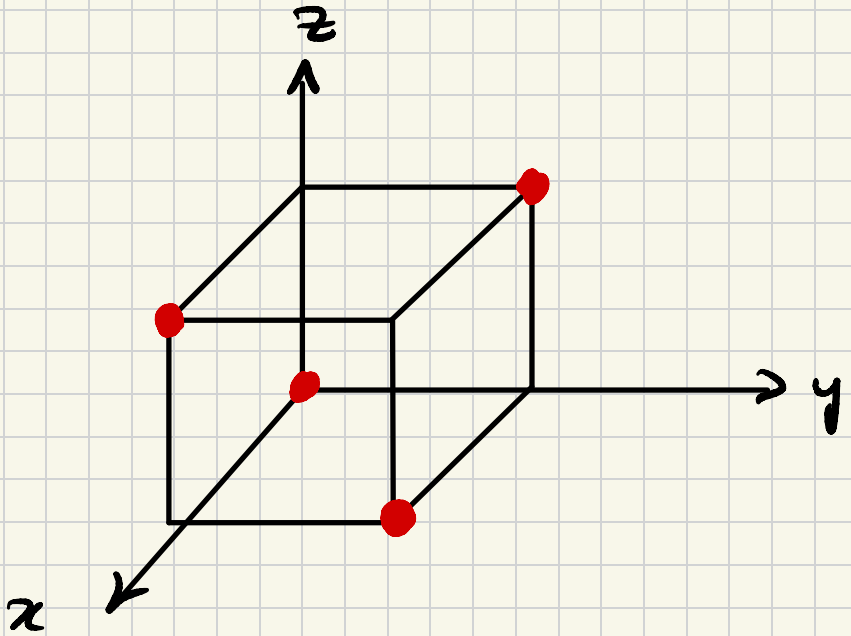
$M = 4$, $n = 4$, $d = 2$

↳ corrige au plus un effacement

ne corrige pas d'erreur

Ex 3 : $C = \{\overset{xyz}{110}, 101, 011, 000\}$

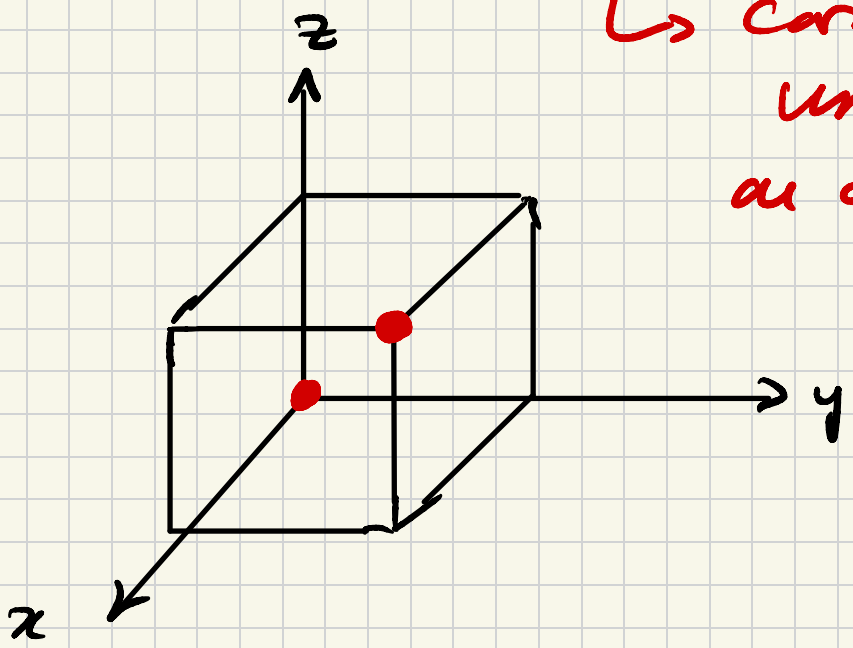
$M = 4,\ n = 3,\ d = 2$



Ex 4 : $C = \{\underbrace{000}_{N}, \underbrace{111}_{S}\}$

$M = 2$, $n = 3$, $d = 3$

↳ corrige jusqu'à une erreur ou deux effacements



But de la théorie du codage :

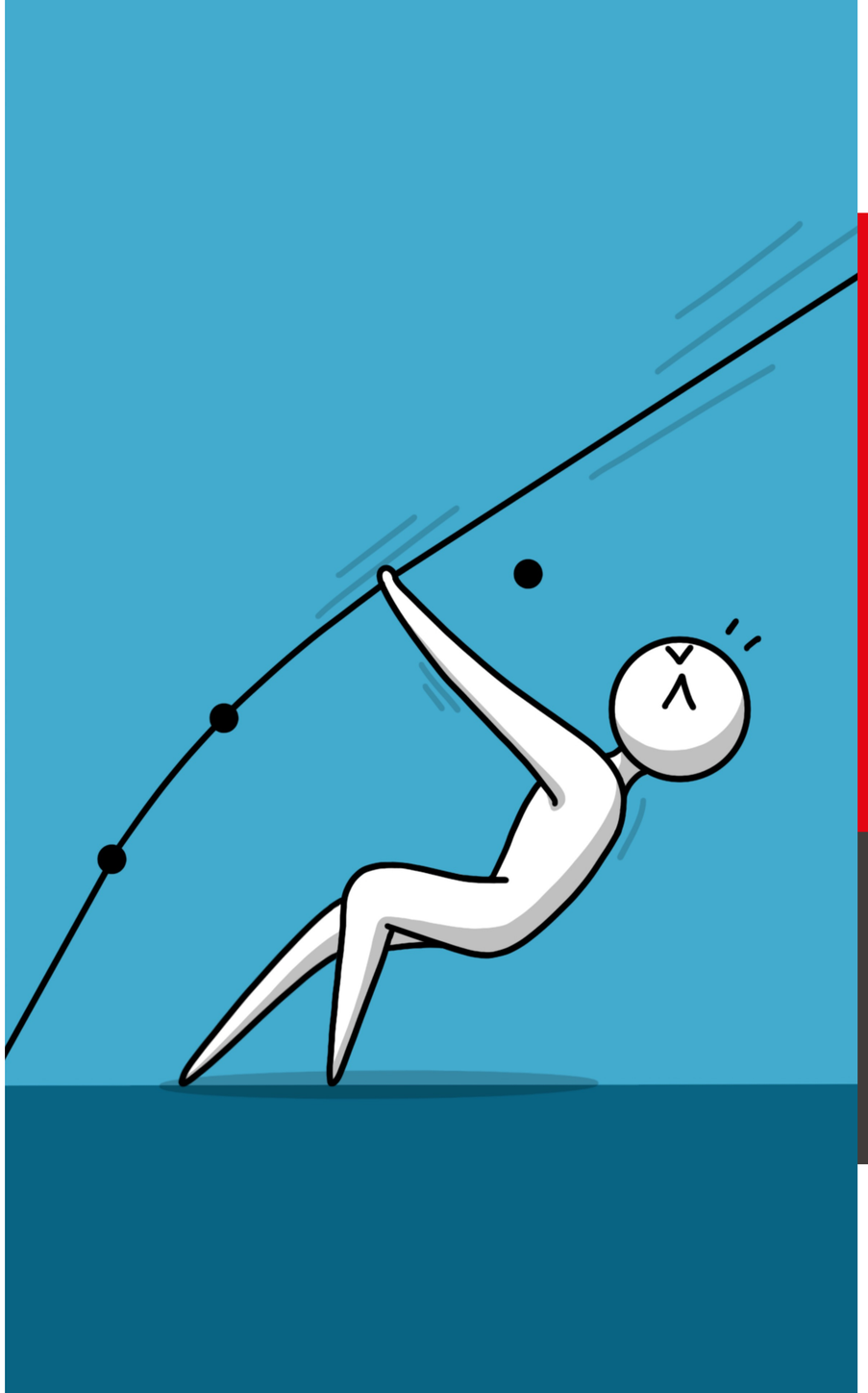
créer des codes avec M

et d les plus grands possibles

et n le plus petit possible

Mais : . $M \leq 2^n$

et : . $M \leq 2^{n-d+1}$ (borne de Singleton)



Information, Calcul et Communication

Les codes de Reed-Solomon

Olivier Lévêque

École polytechnique fédérale de Lausanne

Introduction

- Deux personnes (disons Alice et Bob) cherchent à communiquer, mais une proportion non-négligeable des informations transmises par Alice sont effacées/bruitées avant d'être reçues par Bob:

- Les codes de Reed-Solomon sont un bon moyen de gérer une telle situation.
- Pour simplifier leur description, nous allons supposer qu'Alice et Bob travaillent avec des moyens de communication capables de manipuler des *nombres réels* (et non des bits).

Protocole de communication : Envoi

- *Avant de communiquer*, Alice et Bob se mettent d'accord sur un ensemble de n nombres réels *tous différents* :

$$\{t_1, t_2, t_3, \dots, t_n\}$$

- Alice désire envoyer un message x composé d'une suite de nombres réels :

$$x = \{x_1, x_2, x_3, \dots, x_k\} \quad \text{avec} \quad k < n$$

- Elle définit le polynôme suivant :

$$P(t) = \sum_{i=1}^{k} x_i \cdot t^{i-1} \qquad \deg(P) \leq k - 1$$

Ex ($k=3$) : $P(t) = x_1 + x_2 t + x_3 t^2$

Protocole de communication : Envoi

- *Avant de communiquer*, Alice et Bob se mettent d'accord sur un ensemble de n nombres réels *tous différents* :

$$\{t_1, t_2, t_3, \dots, t_n\}$$

- Alice désire envoyer un message x composé d'une suite de nombres réels :

$$x = \{x_1, x_2, x_3, \dots, x_k\} \quad \text{avec} \quad k < n$$

- Elle définit le polynôme suivant :

$$P(t) = \sum_{i=1}^{k} x_i \cdot t^{i-1} \qquad \deg(P) \leq k - 1$$

- et envoie finalement le message :

$$y = \{y_1 = P(t_1)\,,\ \ y_2 = P(t_2)\,,\ \ y_3 = P(t_3)\,, \dots,\ \ y_n = P(t_n)\}$$

Protocole de communication : Réception

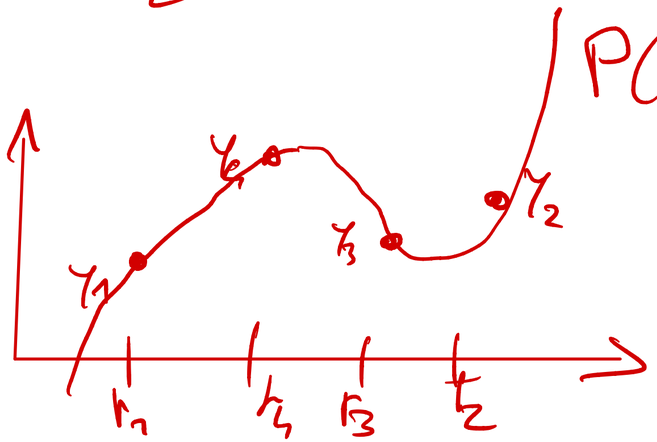
- Bob reçoit le message y. On suppose que $n - k$ nombres du message sont effacés; il reçoit donc que k nombres de y. Pour simplifier, admettons de plus que Bob ne reçoive que les k premiers nombres $\{y_1, y_2, y_3, \ldots, y_k\}$.

$(y_1, y_2, y_3, y_4 \cdots, y_k, \cancel{y_{k+1} \cdots y_n})$

$y_1 = P(t_1) \quad y_2 = P(t_2) \quad \cdots \quad y_k = P(t_k)$

P(t) est reconstruit

Ex (k=4) :



↳ on retrouve $y_1 \cdots y_k$

Protocole de communication : Réception

- Bob reçoit le message y. On suppose que $n - k$ nombres du message sont effacés; il reçoit donc que k nombres de y. Pour simplifier, admettons de plus que Bob ne reçoive que les k premiers nombres $\{y_1, y_2, y_3, \dots, y_k\}$.
- Le but de Bob est de retrouver $\{x_1, x_2, x_3, \dots, x_k\}$ à partir de $\{y_1, y_2, y_3, \dots, y_k\}$

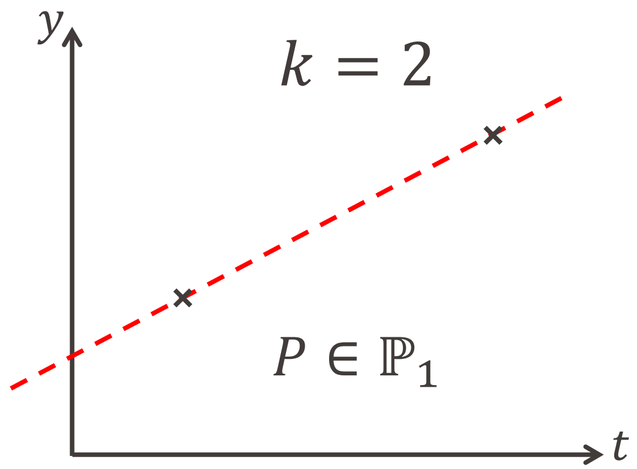
$$\text{Rappelons que } y_j = P(t_j) = \sum_{i=1}^{k} x_i \cdot {t_j}^{i-1} \quad \text{pour} \quad 1 \leq j \leq k$$

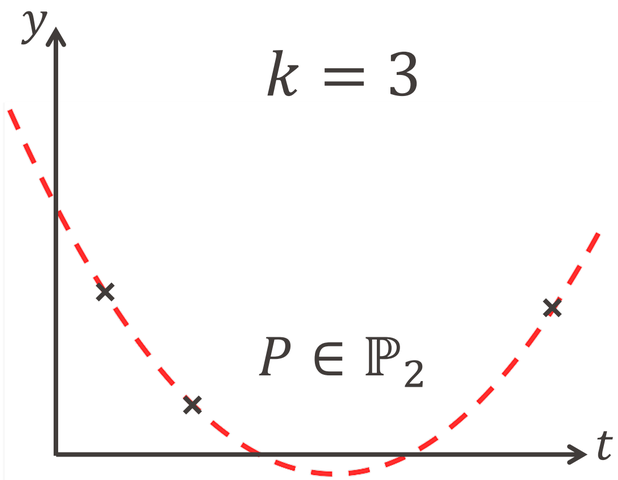
- Il s'agit donc de résoudre un système linéaire de k équations à k inconnues!

$$\text{Exemple pour } k = 3 \text{ et } t = \{1, 2, 3\} : \begin{cases} x_1 \cdot 1 + x_2 \cdot 1 + x_3 \cdot 1 = y_1 \\ x_1 \cdot 1 + x_2 \cdot 2 + x_3 \cdot 4 = y_2 \\ x_1 \cdot 1 + x_2 \cdot 3 + x_3 \cdot 9 = y_3 \end{cases}$$

Une autre façon de visualiser le problème

- En réalité, Bob reçoit les coordonnées des points (t_j, y_j) $\forall\ 1 \leq j \leq k$. Il s'agit donc de trouver l'unique polynôme P tel que $\deg(P) \leq k - 1$ passant par les k points différents.



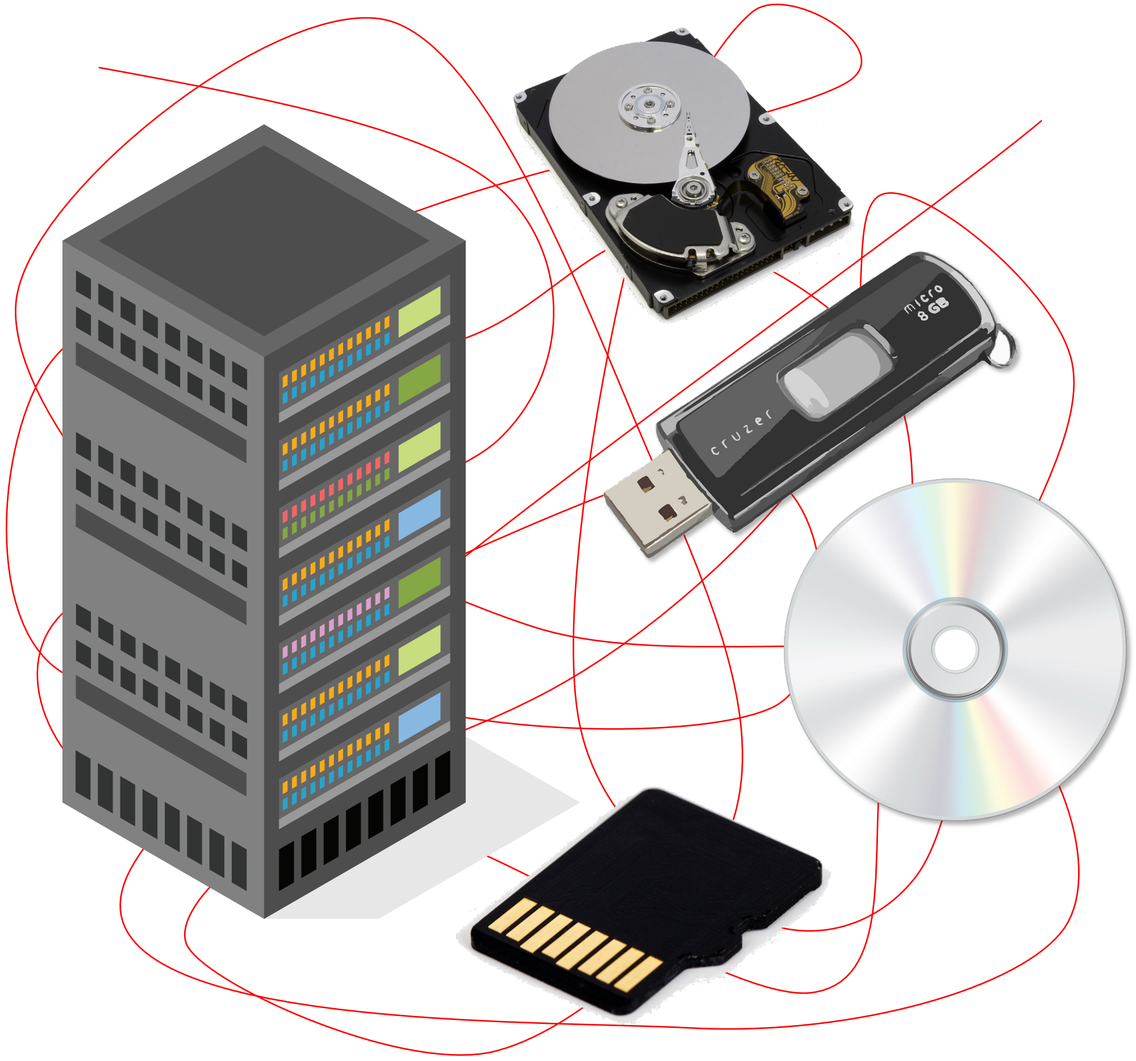


- Ainsi, Bob retrouve le message envoyé x (composé des coefficients du polynôme P).

En pratique

- On ne travaille pas avec des nombres réels, mais avec des nombres entiers modulo p (où p est un nombre premier) : $\{0,1,2,\ldots,p-1\}$, ou plus généralement des nombres appartenant à un groupe fini.
- Et les mêmes principes concernant les polynômes s'appliquent dans ce cadre.

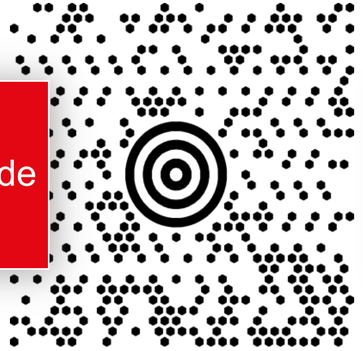
Application : Stockage de données et codes-barres 2D



QR code

Aztec Code

MaxiCode



Data Matrix

PDF-417