Problématiques liées à l'analyse du train binaire

Pierre Loidreau

DGA
Centre d'électronique de l'Armement

10 juin 2008

Plan de l'exposé

- Présentation du problème
- Algorithmes existants
- Résultat théorique
- Perspectives et conclusion

Présentation du problème

La chaîne de communication

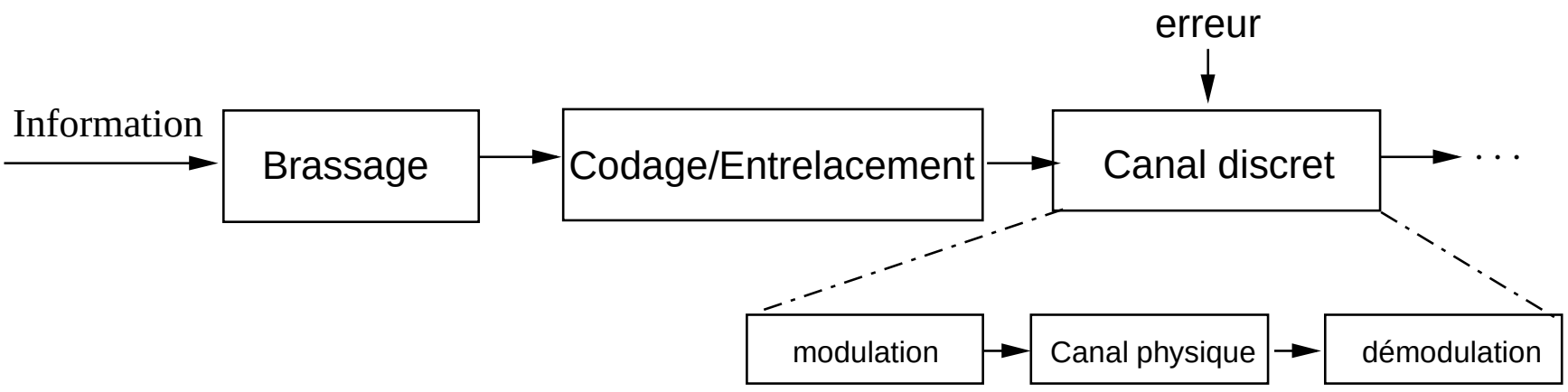

Figure: Chaîne de communication

Données du problème

- N symboles consécutifs provenant d'un code linéaire issus de la chaîne de communication en *contexte non-coopératif*
- Modèle de canal connu
  Exemple : BSC de probabilité de transition p

Hypothèse : La forme d'onde n'est pas connue.

Objectifs

- Idéal : Récupérer l'information facilement
  $\Rightarrow$ Pb. difficile sans information additionnelle

- Pragmatique :
  - Reconstruire le code
  - Avec des informations supplémentaires :
    - Retrouver le codeur
    - Retrouver le brasseur

La reconstruction de code

But : Retrouver une matrice génératrice ou de parité du code

Étapes de reconstruction :

1. Famille de code (Bloc, convolutif, turbo. . .)
2. Paramètres du code
3. Le code

A priori, la complexité de résolution des problèmes est croissante

Algorithmes existants

Principe

1. Choix de paramètres n et ℓ t.q. $\ell = \lfloor N/n \rfloor$
2. Mise des N bits sous forme d'une matrice $\mathcal{H}$ de taille $\ell \times n$ bits interceptés
3. Tests pour déterminer les bons paramètres

- H : On connaît la famille de codes

Cas des codes en blocs

- Si $p = 0$ et bons paramètres, $\mathcal{H}$ est de rang $n - k < n$
- Si $p \neq 0$ et *suffisamment petit*, et bons paramètres la répartition des vecteurs de petit poids t.q. $\mathcal{H}c = 0$ n'est pas aléatoire ([Valembois01])
- H : Sinon la répartition est aléatoire

$\Rightarrow$ Test :

$$\text{Distrib. Mots de petits poids} \stackrel{?}{=} \text{Distrib. Aléatoire}$$

Mode opératoire

On recherche *suffisamment* de vecteurs de petit poids, c t.q.

$$\mathcal{H}c = 0$$

pour discriminer la distribution d'une distribution aléatoire

- Algorithme Canteaut-Chabaud ([Cluzeau/Finiasz08])
- Algorithme Gauss randomisé ([Sicot/Houcke05, Barbier/Sicot/Houcke06])
- Recherche exhaustive si beaucoup de mots de très petit poids (≤ 6) ([Cluzeau/Finiasz08])

Cas des codes convolutifs

Retrouver une matrice génératrice polynomiale canonique, ([Barbier06, Filiol01, Planquette96])

Si $p = 0$

1. Estimer les paramètres n, k, d (degré interne)
2. Construire une matrice Y à partir des bits interceptés
3. Déterminer le noyau de Y
4. Reconstruire la matrice polynomiale

Si $p \neq 0$

- Modification de l'étape 3 en utilisant les résultats sur les codes en bloc

Méthodologies de test adoptées

1. Choix d'un code $\mathcal{C}$
2. Choix d'un canal BSC de probabilité de transition p
3. Construction de N bits issus de mots de codes bruités
4. Appliquer l'algorithme de reconstruction
   - Si passe le test et correspond à $\mathcal{C} \Rightarrow$ Succès, sinon fausse alarme
   - Sinon arrêter l'algorithme au bout d'un certain temps et échec

Évaluation des algorithmes

Arrêt des algorithmes

- S'arrêtent quand le test est passé
- Si le test ne passe pas on arrête l'algorithme au bout d'un nombre défini d'itérations

Difficultés

- Evaluation de la probabilité d'échec
- Evaluation de la probabilité de fausse alarme

$\Rightarrow$ difficile de définir proprement un critère d'arrêt

Quelques exemples

- Codes en bloc aléatoires : $N = 5n^2$, $n = 128 = 2k$, $p = 10^{-2}$.
- Codes LDPC : $N = 512n$, $n = 1000 = 2k$, $p = 10^{-2}$

Mais

- Reconstruisent bien en deçà de la capacité de correction des codes.
- Peu de résultats permettant d'extrapoler les résultats obtenus sur des simulations à des codes utilisés en pratique.

Résultat théorique

Résultat théorique - [Cluzeau/Tillich08]

- $\mathcal{C} \in \mathcal{E}, n, R = k/n$
- Observation : $Y_i = X_i + E_i$, $X_i \in \mathcal{C}$, $i = 1, \ldots, M$

Pour espérer retrouver $\mathcal{C} \in \mathcal{E}$,

$$N/n = M \geq (1 + o(1)) \frac{\log_2(|\mathcal{E}|)}{I(X;Y) - I(X;Y|C)}$$

Résultat théorique suite

Exemples dans certains cas

- Codes de rendement fixé $R = k/n$.

$$M \geq (1 + o(1))k \frac{1 - R}{I(U;V) - R}$$

- Codes LDPC réguliers (t, s), sur BSC de transition p

$$M \geq (1 + o(1)) \frac{s(1 - 1/t)}{1 - h(p) - R}$$

Perspectives et conclusion

Pistes exploratoires

1. Reconstruire d'autres familles de codes
2. Analyser les algorithmes
   - Décider rapidement de l'échec des algorithmes
   - Évaluer les cas de fausses alarmes
3. Utiliser des propriétés de structures pour favoriser la reconstruction

A terme

- Signature de codes
- Trouver les paramètres sans reconstruire
- Reconstruire à la capacité de correction
- Que faut-il pour reconstruire ?

Conclusion

- Utiliser d'autres techniques
- Utiliser de l'information souple
- Utiliser de l'information supplémentaire
- S'attaquer aux autres parties de la chaîne de communication
- Étudier le problème pour les *nouveaux* canaux

- [Valembois01] : *Détection, reconnaissance et décodage de codes linéaires binaires*, thèse de doctorat 2001.
- [Cluzeau/Finiasz08] : Recovering a code-s length and synchronization from a noisy intercepted Bitstream, preprint 2008.
- [Sicot/Houcke05] : Blind detection of interleaver parameters, *ICASSP 2008*.
- [Barbier/Sicot/Houcke06] : Algebraic approach for the reconstruction of linear and convolutional error correcting codes, *CISE 2006*.
- [Planquette96] : *Identification de train binaires codés*, Thèse de doctorat.
- [Filiol01] : Reconstruction of punctured convolutional encoders, *Cryptography and Coding 2001*.
- [Cluzeau/Tillich08] : On the code reverse engineering problem, *ISIT 2008*.