AUTHORIZED WRITING FOR "WRITE-ONCE" MEMORIES

Philippe Godlewski and Gérard D. Cohen
ENST, Département SYC,
46 rue Barrault, 75013 PARIS, France.

Abstract

We describe a method for storing information on a "write-once" memory with the following feature : reading is easy, whereas writing is difficult, except for the designer.

1. Introduction

We start from the following problem, (see [1] for details and terminology) : how to reuse a "Write-Once" memory. That is, we have a storage medium, called **wom** , consisting of n binary positions or **wits** , initially containing a "0". At some step, a wit can be irreversibly overwriten with a "1" (e.g. by some laser beam in digital optical disks or burning microscopic fuses in PROMS).

We consider a coding technique, which we call **coset coding** ([2], [3]), based on error-correcting codes, enabling many rewritings on a wom. We denote by C[n,k] a binary linear code with length n and dimension k. It is used to encode r=n-k bits on a wom as follows : every message $s \in F_2^{n-k}$ is one-to-one associated with a coset of C, say x+C, having for syndrome s. That is $s=x.H^t=s(x)$ where H is a generator matrix for $C^*[n,n-k]$, the dual of C. To encode s (or "write" or "update") envolves finding a vector y in x+C (i.e. with syndrome s), and writing it in the wom. Then writing thriftly (using a minimum number of wits) needs a complete decoding algorithm in the sense of error-correcting codes. Reading on the wom (or decoding the womcode) is simply a syndrome computation retrieving s from y. Notice that, whereas the writing procedure is NP-hard for general codes, see [4], reading only takes $O(n^2)$ mod 2-additions. This is a desirable feature for many applications, where the most frequent operation is reading, but updating is exceptionnal. Our method also has a cryptographic flavour, inspired by McEliece ([5]) which we describe now.

F. Pichler (Ed.): Advances in Cryptology - EUROCRYPT '85, LNCS 219, pp. 111-115, 1986.

2. An easy reading – reserved writing procedure

A disk designer D proposes a data storing system which enables him to keep "economic" control on updatings. This "authorized writer" chooses a code possessing a simple complete decoding scheme which allows him easy updating. He keeps secret H', the parity matrix of the code, and publishes H, a random permutation of the columns of an equivalent matrix MH'. Then

$H = M H' P$,

where M is a $(n-k)x(n-k)$ invertible matrix, and P a nxn permutation matrix. A guaranteed number of possible updatings is also made public. Everybody can use H to read in the wom by syndrome computation. However, anyone, apart from D, willing to write syndrome s is faced with the following alternative :

- find y, with minimum or upperbounded weight, representing s, by use of H : reputed untractable ([4]).
- find any y representing s, hereby losing updating possibilities ; in that case the weight of y is likely to be large ($\sim (n-k)/2$, see section 4).

3. The updating problem

The designer needs a complete decoding algorithm or a maximum likelihood decoding algorithm (a difference with McEliece scheme). He must solve the following :

Problem :

Given a message $s \in F_2^{n-k}$

and a set I of written wits $I \subset \{1,2,\ldots,n\}$

Find a word y with minimum weight satisfying

(i) $yH^t = s$

(ii) $I \subset supp(y)$.

The updating procedure is depicted in figure 1.

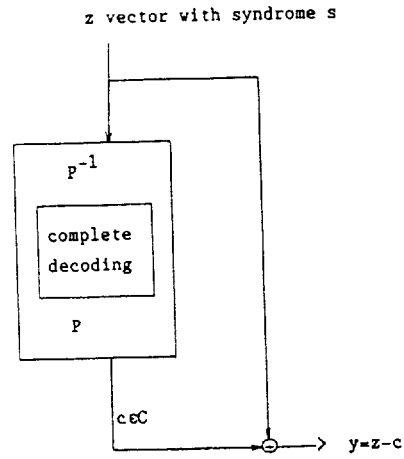


Figure 1

In fact, because of point (ii), D needs a little bit more than a complete decoding, that is, a kind of weigthed decoding. This does not bring additional complexity for Viterbi type or treilis decoding ; just consider for instance some metric modification in the treillis when $i \in I$:

$d(z_i, c_i) = 0$ if $z_i \neq c_i$

$= \infty$ if $z_i = c_i$.

We now give a list of block codes with "practical" complete decoding algorithms.

- block codes with k or n-k not too large,
- some high rate BCH codes.

Such codes may not be used isolated in the scheme but can be part of the following constructions :

- product and concatenated codes ([6]),
- block codes constructed from time-varying convolutional codes ([7]).

4. Strategy for the unauthorized writer

Hereafter could be a strategy of the unauthorized writer (UW) :

- Pick any set S of r=n-k columns in the parity-check matrix H.
- Check whether S has full rank r : this is easy, and the answer is "yes" with probability $\sim 1-2^{r}$ ("no" means S is the support of a codeword in C, an event of probability $\sim 2^{r}$).
- Write syndrome s as a linear combination of elements in S.

The average number of wits used in this operation is r/2. More precisely, the probability that this number would be essentially smaller, that is, λr, for some fixed λ, $0 \leq \lambda \leq 1/2$ is

$$2^{-r} \sum_{i=0}^{\lambda r} \binom{r}{i} \sim 2^{r(h_2(\lambda)-1)},$$

where $h_2(.)$ is the binary entropy function. In other words, when r becomes large, the unauthorized writer almost surely uses effectively r/2 wits to write s.

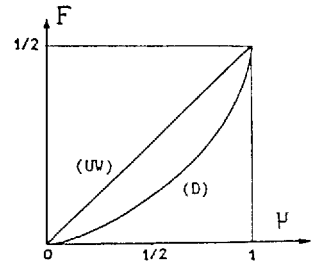
Let us now compare this with the situation of the designer (D).

We call $t(C)$ the **covering radius** of C, i.e. the maximal weight of a coset leader, or equivalently, the greatest possible distance between vectors in F_2^n and C. Then it is clear that the coset coding procedure uses at most $t(C)$ wits. For most codes, it is known (cf. e.g. [8]) that, for $r/n = \mu$,

$$t(C) < n h_2^{-1}(\mu).$$

where h_2^{-1} is the inverse of h_2 on $[0,1/2]$. Let us summarize these facts

	UW	D
F : fraction of the wom used by coset coding with C[n,n(1-μ)]	$\frac{1}{2}\mu$	$h_2^{-1}(\mu)$



Finally, we give a cleverer strategy for UW which gives slightly (but not essentially) better results :

- Transform H into a equivalent matrix H"=M".H containing as a submatrix a rxr identity matrix ; M" is an invertible rxr matrix. Let V be the set of the n-r other columns of H".
- For given parameters λ and j, compute the Hamming distances between the transformed syndrome $s"=sM"^t$ to be written and the sums of at most j columns of V. Call the algorithm successful if one of these distances is at most λr.
- In case of success, take the associated set of columns in V, and add the appropriate columns from the identity submatrix (at most λr of them), getting s as a linear combination of at most $\lambda r + j$ columns in H" and then in H.

The number of basic operations to perform is

$$\sum_{i=0}^{j} \binom{n-r}{i} < (n-r)^j.$$

For λ fixed, the probability of success is upperbounded by

$$1 - \left[1-2^{-r(1-h_2(\lambda))}\right]^{(n-r)^j} \sim (n-r)^j\, 2^{-r(1-h_2(\lambda))},$$

i.e. goes to zero for $\lambda < 1/2$ and fixed r/n when r goes to infinity.
For $\lambda=(1/2)-\theta.r^{-1/2}$, θ constant, the probability of success is non vanishing, but the average number of wits written is $\lambda r \sim r/2$, like with the previous strategy.

References

[1] R.L.Rivest and A.Shamir, "How to Reuse a "Write-Once" Memory", Inform. and Control 55, 1-19 (1982).

[2] Heegard, "An efficient Encoder for Algebraic Optical Disk Codes", Preprint.

[3] G.D. Cohen, P. Godlewski, F. Merkx, "Linear Binary Codes for Write-Once Memories", submitted to IEEE-IT.

[4] E.R. Berlekamp, R.J. McEliece and H.C.A. Van Tilborg, "On the Inherent intractability of certain Coding Problems", IEEE Trans. Inform. Theory, vol. IT-24, pp.384-386, May 1978.

[5] R.J. McEliece, "A Public-Key Cryptosystem based on Algebraic Coding Theory", DSN Progress Report, Jan 1978.

[6] J.K. Wolf, "Efficient Maximum Likelihood Decoding of Linear Block Codes Using a Trellis", IEEE Trans. Inform. Theory, vol. IT-24, No. 1, pp. 76-80, Jan. 1978.

[7] J.L. Massey, "Error Bounds for Tree codes, Treillis Codes and Convolutional Codes with Encoding and Decoding Procedures", in Coding and Complexity, G. Longo Ed., C.I.S.M. Courses and Lectures No. 216, New York, Springer-Verlag, 1974, pp. 1-57.

[8] G.D. Cohen, "A Nonconstructive Upper Bound on Covering Radius", IEEE Trans. Inform. Theory, vol. IT-29, pp. 352-353, May 1983.