A Fuzzy Sketch with Trapdoor*

Julien Bringer[1], Hervé Chabanne[1], Quoc Dung Do[2]

[1]SAGEM Défense Sécurité,
[2]Ecole Polytechnique, ENST Paris.

Abstract

In 1999, Juels and Wattenberg introduce an effective construction of Fuzzy Sketch, i.e. a way of handling errors into string verification. This allows them to consider data varying into time, such as, for instance, answers to a list of subjective questions. To this end, they utilize an Error Correcting Code.

We here show how to embed a trapdoor into Fuzzy Sketches, reducing to authorized people the ability to correct errors and thus to verify the fuzzy equality to the Fuzzy Sketch.

Keywords. Fuzzy Sketch, Cryptosystem of McEliece.

1 Introduction

Handling errors into verification comes from [8]. In [14], Juels and Wattenberg give an effective construction of Fuzzy Sketches for the Hamming space. I.e. denoting $Fsk(w)$ the Fuzzy Sketch of w, from $Fsk(w)$ and $w \oplus e$ where e symbolized some errors, one can recover w. Later on, [13] describes a Fuzzy Vault which can be interpreted as a Fuzzy Sketch for the set difference metric. Finally, [9] sets a general formal framework and defines several techniques to be used for those distances as well as for the edit distance. We will here only consider binary Hamming space.

Fuzzy Sketches are often associated to biometric information but some other data can also be considered. For instance, [11] proposes a fault tolerant scheme allowing very long passwords and some errors. Previous works [3, 4] make the assumption that they are dealing with secret information. Here, we do not take this hypothesis as granted and consider rather that we are working with public fuzzy data. In our setting, one can, for instance, store

*The work presented in this paper has been exclusively supported by SAGEM Défense Sécurité.

fuzzy sketches inside a database to enforce control access to some assets. We here stress that letting an attacker the correcting capability of the underlying error correcting code may lead to awkward situations. For example, if he has access to the database, he will be able to check the membership of a given person, which could be unacceptable for privacy reasons.

We show that a slight modification of the Fuzzy Sketch of Juels and Wattenberg allows to embed a trapdoor in this construction restricting the ability to check the fuzzy equality. Starting from the cryptosystem of McEliece [16], we exploit the fact that there is today a gap between the errors which can be corrected, for a given security parameter, by an attacker and the error correction capacity of those who have knowledge of the trapdoor and of the specialized decoding algorithms for the underlying Goppa codes. In fact, the cryptosystem of McEliece resists quite well through ages as in more or less twenty years, the work factor of the best known attack of the original scheme has decreased from 2^{81} [16] to 2^{49} [12] but the complexities of the attacks are still exponential.

2 Preliminaries and notations

2.1 Coding theory

Let w be a word of $\{0,1\}^n$, the *Hamming weight* $\mathrm{wt}_H(w)$ of w is the number of coordinates of w which are non-zero. The *Hamming distance* over $\{0,1\}^n$ is the canonical metric distance defined as the number of differences between two words. Hence, for two words w_1, w_2, the Hamming distance $\mathrm{d}_H(w_1, w_2)$ is equivalent to the weight of $w_1 \oplus w_2$ for a binary alphabet. For some integer n, the set $\{0,1\}^n$ equipped with the Hamming distance d_H is an *Hamming space* $\mathcal{H}$.

An *error correcting code* in $\mathcal{H}$ is a subset C of $\mathcal{H}$, the elements of C are called codewords, the *minimum distance* d_{min} of C is the smallest distance between two distinct codewords. This means that one can detect up to $d_{min}-1$ errors in a codeword. The *capacity of correction* t of C is the radius of the largest ball for which for any $w \in \mathcal{H}$ there is at most one codeword in the ball of radius t centered on w. In fact for the Hamming distance d_H, $t = \lfloor (d_{min}-1)/2 \rfloor$. As $\{0,1\}$ can be equipped with the structure of the finite field $\mathbb{F}_2$, $\mathcal{H}$ is also a vector space $\mathbb{F}_2^n$ and we define on it a linear code C as a vector subspace of $\mathcal{H}$. If k is the dimension of the vector space C over $\mathbb{F}_2$, we denote C as a $[n, k, d_{min}]$ code over $\mathbb{F}_2$. Due to linearity property, the minimum distance d_{min} is then equal to the minimum weight of the codewords $w \in C - \{0\}$. Moreover, C can be described by a matrix G, call a *generator matrix* of C, with k lines, n rows and a rank k, such that the lines of G is a basis of C: we obtain $C = \{mG \mid m \in \mathbb{F}_2^k\}$.

Remark 1 *For a random linear code given by its generator matrix, we then easily encode the codewords but the decoding process is not a computational effective function in general. We will see further how this fact is used in the cryptosystem of McEliece.*

2.2 Fuzzy Sketches

Let X and Y be two random discrete variables. We recall the definition of the *entropy* of X, $\mathbf{H}(X) = \mathbf{E}_{x \leftarrow X}(-\log_2 \mathbb{P}(X = x))$, and the definition of the *conditional entropy* of X given Y, $\mathbf{H}(X \mid Y) = \mathbf{E}_{y \leftarrow Y}\mathbf{H}(X \mid Y = y) = \mathbf{E}_{x,y \leftarrow X,Y}(-\log_2 \mathbb{P}(X = x \mid Y = y))$. We also introduce the *min-entropy* of X defined as $\mathbf{H}_\infty(X) = -\log_2 \max_x \mathbb{P}(X = x)$ and the *average min-entropy* for X given Y defined as $\overline{\mathbf{H}}_\infty(X \mid Y) = -\log_2 \mathbf{E}_{y \leftarrow Y}(2^{-\mathbf{H}_\infty(X|Y=y)}) = -\log_2 \mathbf{E}_{y \leftarrow Y}(\max_x \mathbb{P}(X = x \mid Y = y))$, which are more relevant for cryptographic use. Namely, the *average min-entropy* of X given Y is the logarithm of the average probability of the most likely value of X given Y. This notion is useful to measure the difference from uniform distributions. For example, if Y has values in $\{0,1\}^n$ then, $\overline{\mathbf{H}}_\infty(X \mid Y) \geq \mathbf{H}_\infty(X) - n$.

This notion allows us to introduce the definition of fuzzy sketches:

Definition 1 *A $(\mathcal{H}, m, m', t)$-fuzzy sketch is a pair of functions $(Fsk,\ Cor)$ where:*

- *Fsk is a (typically randomized thanks to a randomizer x) sketching function that on input $w \in \mathcal{H}$ outputs a sketch or redundancy data $P \in \{0,1\}^*$, such that for all random variable W over $\mathcal{H}$ with min-entropy $\mathbf{H}_\infty(W) \geq m$, the average min-entropy of W given $Fsk(W)$ is at least m'. That is, $\overline{\mathbf{H}}_\infty(W \mid Fsk(W)) \geq m'$.*

- *Cor is a correction function which allows to recover w from its sketch and any vector close to w: given a word $w' \in \mathcal{H}$ and a sketch P, it outputs a word $w'' \in \mathcal{H}$, such that for any $P = Fsk(w)$ and $\mathrm{d}_H(w, w') \leq t$, it holds that $w'' = w$.*

In practice, the sketching and correction functions, Fsk and Cor, have to be efficiently computable, that is they run in polynomial time in n. Here, we denote the parameter t as the *capacity correction of the fuzzy sketch.*

The following construction has been considered by Juels and Wattenberg:

Claim 1 (Code-offset construction [14]) *Given C a binary linear code of length n, dimension k and correction capacity t, the Fuzzy Sketch of Juels and Wattenberg is given by $Fsk_{JW}(w;x) = w \oplus c(x)$ where*

- *x, the randomizer of Fsk_{JW}, is a random vector of length k,*
- *$c(x)$ is taken at random from C.*

This yields a $(\mathcal{H}, m, m+k-n, t)$-fuzzy sketch.

Remark 2 *The value $m - m' = n - k$ represents the entropy loss of the fuzzy sketch. When m is not big enough, the amount $m+k-n$ may not be sufficient to ensure the security of the sketch from an information theory point of view. We here show how to strengthen this construction in order to add a "computational security" which stands even when m is small (see also Remark 5).*

3 Our construction

3.1 Cryptosystem of McEliece

We place ourselves into a hard instance of the McEliece cryptosystem, i.e. let $d = c \oplus e$ where

- c is a word taken at random from a concealed Goppa code,
- e stands for errors of sufficient weight,

such that there is no computational effective way, for a given security parameter Σ, to recover c from d without the knowledge of the trap. We would want then to determine the capacity to handle additional errors using the original Goppa code.

Claim 2 (Goppa codes) *Corresponding to each irreducible polynomial of degree t over $GF(2^m)$, there exists a binary irreducible Goppa code of length $n = 2^m$, dimension $k \geq n - tm$, capable of correcting any pattern of t or fewer errors. The decoding of these codes can be done in $O(nt)$ operations.*

The cryptosystem of McEliece is an asymmetric cryptosystem defined as follows:

Claim 3 (Cryptosystem of McEliece [16]) *We consider the two codes defined by the matrices G and G_{pub} such that*

- *G is a $k \times n$ generator matrix of a Goppa code,*
- *$G_{pub} = SGP$ where S is a $k \times k$ random invertible dense matrix and P a random $n \times n$ permutation matrix.*

In the McEliece cryptosystem, G_{pub} is a public data and the matrices G, S and P are kept secret, they constitute the underlying trapdoor.

Let $d = c \oplus e$ where $c = xG_{pub}$ is a codeword with x a random binary vector of length k, and e stands for some errors ($\mathrm{wt}_H(e) \leq t$), then, given the properties of the cryptosystem, we have these facts:

- *Knowledge of the trapdoor allows to recover c from d in polynomial running time using the decoding algorithm of the underlying Goppa code,*
- *without the trapdoor, a basic attack is to perform $o(1)\binom{n}{k}/\binom{n-\mathrm{wt}_H(e)}{k}$ guesses to find k columns of G_{pub} with no errors enabling the recovery of c. For each guess, k^3 binary operations must be performed.*

Improvements of the basic attack are numerous, see [1, 2, 15, 17, 5, 12], leading to a complexity $C_{alg} = N_{alg} \times C_{iter}$ in the basic algorithm of [12] where

$$N_{alg} = \left(\sum_{j=1}^{M} (-1)^{j+1} \binom{M}{j} \left(\sum_{i=0}^{p} \frac{\binom{k}{i}\binom{n-k-l}{r-i}}{\binom{n}{r}} \right)^{j} \right)^{-1},$$

$$C_{iter} = \frac{nk}{2}(k+M) + Ml + \sum_{i=1}^{p} \binom{k}{i} il + \frac{\sum_{j=0} p\binom{k}{j} Mn(j+1)}{2^l}.$$

This complexity is obtained for an $[n, k]$ linear code, M received codewords with at most r errors (below the correction capacity) and for two other algorithm parameters p, l. For example, some optimal complexities with some modifications and well chosen parameters M, p and l are given below [12].

M	C_{alg} for $n = 1024,\ k = 524,\ t = 50$
2^{15}	$2^{56.2}$
2^{30}	$2^{50.2}$
2^{40}	$2^{49.0}$

Usually, instances of McEliece's cryptosystem are made such that the error weight is taken equal to t, at the limit of the correction capacity of the underlying Goppa code. However, if it is not the case, one can note that the complexity of known attacks is a growing function of this error weight. For a given security parameter, and considering an increasing length n, there is some place for correcting additional errors to e.

The following table gives for $\Sigma = 80$, i.e. a work factor greater than 2^{Σ}, the corresponding minimum weight of errors for the cryptosystem of McEliece according to the best known attack [12] with $M = 2^{20}$. In other words and for example, for $n = 2048$, $k = 1024$, when $\mathrm{wt}_H(e) \geq 78$, an attacker must perform more than 2^{80} operations to decode. For $n \leq 1024$, for all parameters n and k, the attacker gets a work factor smaller than 2^{80}.

k/n	$n = 2048$	$n = 4096$	$n = 8192$
0.2	–	235	232
0.4	105	104	101
0.5	78	77	74
0.6	60	59	56
0.8	35	34	33

Table 1: Errors minimum weight for a work factor greater than 2^{80}

Note that when n increases, for a given k/n, $\mathrm{wt}_H(e)$ slowly decreases.

3.2 Fuzzy sketch with a trapdoor

We use the fuzzy sketch construction of Juels and Wattenberg (see Claim 1) in order to introduce a trapdoor. We point out that utilizing Fsk_{JW} with a trapdoor is indeed possible:

Definition 2 (Fuzzy Sketch with Trapdoor) *Let G_{pub} be as in Claim 3 and C the corresponding code of dimension k, then a Fuzzy Sketch with trapdoor is given by $Fsk^{\Sigma}_{trap}(w; x) = w \oplus c(x) \oplus e(x)$ where*

- *x is a random vector of length k,*
- *$c(x) = xG_{pub}$ is a codeword of C,*
- *Σ is a security parameter,*
- *$e(x)$ stands for a well chosen error of weight $\mathrm{wt}_H(e) = \varpi$ with ϖ computed from n, k and Σ.*

In the sequel, we consider that $e(x)$ is taken in a pseudo random way leaded by x from the set of words of length n and weight ϖ.

A correction function is allowed for the possessor of the trapdoor as from $w' = w \oplus f$ and $Fsk^{\Sigma}_{trap}(w;x)$, we get $w' \oplus Fsk^{\Sigma}_{trap}(w;x) = c(x) \oplus e(x) \oplus f$ which allows to compute $c(x)$ whenever $e(x) \oplus f$ has a Hamming weight smaller than the decoding capacity of the underlying Goppa code. From $c(x)$, we get x and $e(x)$ and finally w. Thus, the same randomizer x is used for computing c and e.

We will denote Cor_{trap} the corresponding procedure. Note that this function Cor_{trap} is efficient only for those who know the trapdoor. For the others, our Fsk^{Σ}_{trap} is made such that no correction function operates in less than 2^{Σ} steps, with an overwhelming probability. Indeed, the weight of errors $e(x) \oplus f$ needs to be very small to be corrected without the knowledge of the trapdoor.

We have $\varpi - \mathrm{wt}_H(f) \leq e(x) \oplus f \leq \varpi + \mathrm{wt}_H(f)$. Let $\tau = t - \varpi$. For given n, k and Σ, we compute ϖ such that $\varpi - \tau$ errors leads to a work factor for an attacker greater than 2^{Σ} (as we have done for $\Sigma = 80$ during Section 3.1), see Table 2.

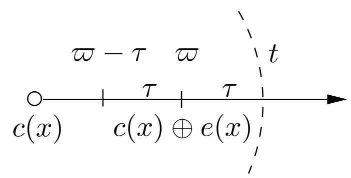

Figure 1: Choice of ϖ and τ

Remark 3 *The error $e(x)$ is chosen such that $\mathrm{wt}_H(e(x) \oplus w) \geq \varpi - \tau$ in order to avoid the fuzzy sketch to be invertible without a vector w' close to w. Whenever this inequality does not hold, we simply choose a new randomizer x.*

With our construction, the possessor of the trapdoor can use the correction function Cor_{trap} for an error of weight $\mathrm{wt}_H(e(x) \oplus f) \leq t$, as for instance it always happens whenever $\mathrm{wt}_H(f) \leq \tau$. Letting its capacity correction be τ, our Fuzzy Sketch with trapdoor has nearly the same behaviour in terms of min-entropy than those of code-offset construction (see Claim 1).

Remark 4 *Note that even the possessor of the trapdoor can not recover w from $Fsk^{\Sigma}_{trap}(w;x)$ without w' close to w, if w has not a weight too small.*

Remark 5 *In [9, Lemma 3.1 (Fuzzy Extractors from Sketches)], Fuzzy Extractors are introduced to alleviate the anticipated poor behaviour of w and then $Fsk(w)$ in terms of min-entropy. Our construction has the same goal. Nevertheless, if needed, it is still possible to apply Fuzzy Extractor to our scheme because the involved technique of [9] is quite general and needs only to operate to have Fuzzy Sketches as defined in Definition 1.*

3.3 Correction capacity

We here place ourselves in a specific security setting, as Fsk_{trap}^{Σ} is designed to resist to known attacks of McEliece up to a given security parameter. Using best known cryptanalysis algorithm [12], we compute τ for a security parameter $\Sigma = 80$:

k/n	τ for $n = 2048$	τ for $n = 4096$	τ for $n = 8192$
0.2	–	19	136
0.4	3	50	139
0.5	7	47	121
0.6	7	39	98
0.8	1	17	47

Table 2: τ for the security parameter $\Sigma = 80$

Note that for a given k/n, τ grows together with n.

In Section 3.2, we have seen how the capacity correction of our fuzzy sketch depends on the capacity correction of the underlying code and on the security parameter Σ. Indeed, the capacity correction of this fuzzy sketch with trapdoor is τ and it has been compute according to Σ (see Table 2).

However, not only the errors f of weight τ can be handle, i.e. much more than $\sum_{i=0}^{\tau} \binom{n}{i}$ errors can be corrected. Actually, if f has a weight $\tau + 1 \leq \mathrm{wt}_H(f) \leq 2\varpi + \tau$ and has enough intersection with e, than it can be corrected. Hence the total number of errors f, for which the correction function works correctly, increases to

$$\sum_{i=0}^{\tau} \binom{n}{i} + \sum_{i=1}^{2\varpi} \binom{\varpi}{\lceil i/2 \rceil} \binom{n - \lceil i/2 \rceil}{\tau + \lfloor i/2 \rfloor}. \tag{1}$$

As an example, for $n = 4096$, $R = k/n = 0.2$, the designed capacity correction $\tau = 19$ of our fuzzy sketch with trapdoor is about 0.5% of n, but according to (1), the correction function can correct the same number of errors f as a function with a capacity correction 273, which is nearly 6.7% of n. So, Cor_{trap} succeeds in managing correctly a large amount of errors.

k/n	$n = 2048$	$n = 4096$	$n = 8192$
0.2	–	6.7%	6.1%
0.4	5.4%	5%	4.6%
0.5	4.5%	4.1%	3.8%
0.6	3.6%	3.3%	3.1%
0.8	1.8%	1.7%	1.5%

Table 3: Equivalent rate of number of errors f correctly processed

4 Conclusion

We show how to include a trapdoor into the Fuzzy Sketches of Juels and Wattenberg.

This renewal in the utilization of the cryptosystem of McEliece can also be viewed as a way of encrypting fuzzy data. And we hope that this will incite to retain more attention on public-key cryptosystem based on error-correcting codes. In particular, the correction capacity of – what we call – McEliece channel (errors that can be added to a hard instance of the cryptosystem of McEliece and corrected) has to be improved.

References

[1] C.M. Adams and H. Meijer, *Security-Related Comments Regarding McEliece's Public-Key Cryptosystem*, Advances in cryptology – CRYPTO 1987, pp. 224–228.

[2] C.M. Adams and H. Meijer, *Security-related comments regarding McEliece's public-key cryptosystem*, IEEE Transactions on Information Theory, vol. 35(2), 1989, pp. 454–455.

[3] X. Boyen, *Reusable cryptographic fuzzy extractors*, ACM Conference on Computer and Communications Security 2004, pp. 82–91.

[4] X. Boyen, Y. Dodis, J. Katz, R. Ostrovsky and A. Smith, *Secure Remote Authentication Using Biometric Data*, Advances in cryptology – EUROCRYPT 2005, pp. 147–163.

[5] A. Canteaut and F. Chabaud, *A New Algorithm for Finding Minimum-Weight Words in a Linear Code: Application to McEliece's Cryptosystem and to Narrow-Sense BCH Codes of Length 511*, IEEE Transactions on Information Theory, vol. 44(1), 1998, pp. 367–378.

[6] G. Cohen and G. Zémor, *The wire-tap channel applied to biometrics*, ISITA2004.

[7] G. Cohen and G. Zémor, *Generalized coset schemes for the wire-tap channel: application to biometric*, ISIT2004.

[8] G. Davida, Y. Frankel, and B. Matt, *On enabling secure applications through offline biometric identification*, In Proc. IEEE Symp. on Security and Privacy, 1998, pp. 148–157.

[9] Y. Dodis, L. Reyzin and A. Smith, *Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data*, Advances in cryptology – EUROCRYPT 2004, pp. 523–540.

[10] I. Dumer, D. Micciancio and M. Sudan, *Hardness of approximating the minimum distance of a linear code*, IEEE Transactions on Information Theory, vol. 49(1), 2003, pp. 22–37.

[11] N. Frykholm and A. Juels, *Error-tolerant password recovery*, ACM Conference on Computer and Communications Security 2001, pp. 1–9.

[12] T. Johansson and F. Jönsson, *On the complexity of some cryptographic problems based on the general decoding problem*, IEEE Transactions on Information Theory, vol. 48(10), 2002, pp. 2669–2678.

[13] A. Juels and M. Sudan, *A Fuzzy Vault Scheme*, In IEEE International Symp. on Information Theory, 2002.

[14] A. Juels and M. Wattenberg, *A Fuzzy Commitment Scheme*, ACM Conference on Computer and Communications Security 1999, pp. 28–36.

[15] P. J. Lee and E. F. Brickell, *An Observation on the Security of McEliece's Public-Key Cryptosystem*, Advances in cryptology – EUROCRYPT 1988, pp. 275–280.

[16] R.J. McEliece, *A public-key cryptosystem based on algebraic coding theory*, JPL DSN Progress Report, 1978, pp. 114–116.

[17] J. van Tilburg, *On the McEliece Public-Key Cryptosystem*, Advances in cryptology – CRYPTO 1988, pp. 119–131.