Complexity of Decoding Positive-Rate Primitive Reed-Solomon Codes

Qi Cheng and Daqing Wan

***Abstract*—It has been proved that the maximum likelihood decoding problem of Reed-Solomon codes is NP-hard. However, the length of the code in the proof is at most polylogarithmic in the size of the alphabet. For the complexity of maximum likelihood decoding of the primitive Reed-Solomon code, whose length is one less than the size of alphabet, the only known result states that it is at least as hard as the discrete logarithm in some cases where the information rate unfortunately goes to zero. In this paper, it is proved under a well known cryptography hardness assumption that**

1) **There does not exist a randomized polynomial time maximum likelihood decoder for the Reed-Solomon code family $[q, k(q)]_q$, where $k(x)$ is any function in $\mathbf{Z}^+ \to \mathbf{Z}^+$ computable in time $x^{O(1)}$ satisfying $\sqrt{x} \le k(x) \le x - \sqrt{x}$.**
2) **There does not exist a randomized polynomial time bounded-distance decoder for primitive Reed-Solomon codes at distance $\frac{2}{3} + \epsilon$ of the minimum distance for any constant $0 < \epsilon < \frac{1}{3}$.**

In particular, this rules out the possibility of a polynomial time algorithm for maximum likelihood decoding problem of primitive Reed-Solomon codes of any rate under the assumption.

***Index Terms*—Computational complexity, Maximum likelihood decoding, Reed-Solomon codes.**

I. Introduction

Let $\mathbf{F}_q$ be a finite field of q elements and of characteristic p. A linear error-correcting $[n, k]_q$ code is defined to be a linear subspace of dimension k in $\mathbf{F}_q^n$. Let $D = \{x_1, \cdots, x_n\} \subseteq \mathbf{F}_q$ be a subset of cardinality $|D| = n > 0$. For $1 \le k \le n$, let f run over all polynomials in $\mathbf{F}_q[x]$ of degree at most $k-1$. The vectors of the form

$$(f(x_1), \cdots, f(x_n)) \in \mathbf{F}_q^n$$

constitute a linear error-correcting $[n, k]_q$ code, which is called a Reed-Solomon code. If $D = \mathbf{F}_q^*$, it is famously known as a primitive Reed-Solomon code. If $D = \mathbf{F}_q$, it is known as an extended primitive Reed-Solomon code. We denote them by $RS_q[q-1, k]$ and $RS_q[q, k]$ respectively. A generalized Reed-Solomon code $[n, k]_q$ is defined to be

$$\{(y_1 f(x_1), \cdots, y_n f(x_n)) | f \in \mathbf{F}_q[x], deg(f) < k\},$$

where $y_1, y_2, \cdots, y_n$ are nonzero elements in $\mathbf{F}_q$.

The preliminary version of this paper appeared in the Proceedings of the 35th International Colloquium on Automata, Languages and Programming (ICALP), volume 5125 of Lecture Notes in Computer Science, Springer-Verlag, 2008.

Qi Cheng is with School of Computer Science, University of Oklahoma, Norman, OK73019, Email: qcheng@cs.ou.edu. His research is partially supported by NSF grant CCF-0830524 and CCF-0830522.

Daqing Wan is with Department of Mathematics, University of California, Irvine, CA 92697-3875, Email: dwan@math.uci.edu. His research is partially supported by NSF.

The minimal distance of a generalized Reed-Solomon $[n, k]_q$ code is $n - k + 1$ because a non-zero polynomial of degree at most $k-1$ has at most $k-1$ zeroes. The ultimate decoding problem for an error-correcting $[n, k]_q$ code is the maximum likelihood decoding: given a received word $u \in \mathbf{F}_q^n$, find a codeword v such that the Hamming distance $d(u, v)$ is minimal. When the number of errors is reasonably small, say, smaller than $n - \sqrt{nk}$, then the list decoding algorithms of Guruswami-Sudan [6] gives a polynomial time algorithm to find all the codewords.

When the number of errors increases beyond $n - \sqrt{nk}$, it is not known whether there exists a polynomial time decoding algorithm. The maximum likelihood decoding of a Reed-Solomon $[n, k]_q$ code is known to be **NP**-complete [4]. The proof explores the combinatorial complication of the subset D, thus requires that n is at most polylogarithmic in q. In fact, there is a straightforward way to reduce the subset sum problem in D to the deep hole problem of a Reed-Solomon code, which can then be reduced to the maximum likelihood decoding problem [2]. Note that the subset sum problem for $D \subseteq \mathbf{F}_q$ is hard only if $|D|$ is much smaller than q. See [8] for an in-depth discussion of the subset sum problem when $|D|$ is close to q.

In practical applications, one rarely uses the case of arbitrary subset D. The most widely used case is when $D = \mathbf{F}_q^*$, where the rich algebraic structure of the field facilitates a concise representation of alphabet and a fast encoding algorithm. This case is essentially equivalent to the case $D = \mathbf{F}_q$. For simplicity, we focus on the extended primitive Reed-Solomon code $RS_q[q, k]$ in this paper, all our results can be applied to the Reed-Solomon code $RS_q[q-1, k]$ with little modification. The maximum likelihood decoding problem of $RS_q[q, k]$ is considered to be hard, but the attempts to prove its **NP**-completeness have failed so far. The methods in [4][2] can not be specialized to $RS_q[q, k]$ because we have lost the freedom to select D. The only known complexity result [3] in this direction says

Proposition 1: Let $\delta > 0$ be a constant. Let q be a prime power. Suppose h and k are positive integers satisfying

$$h \le \sqrt{q} - k, h \le q^{\frac{1}{2+\delta}} + 1 \quad \text{and} \quad h \le \frac{k - \frac{4}{\delta} - 2}{\frac{4}{\delta} + 1}.$$

The discrete logarithm in $\mathbf{F}_{q^h}^*$ can be solved in randomized time $q^{O(1)}$ with oracle access to a maximum likelihood decoder of $RS_q[q, k]$.

The main weakness of this result is that for the discrete logarithm over $\mathbf{F}_{q^h}^*$ to be hard, $\sqrt{q}$ has to be greater than k, which implies that the information rate k/q goes to zero. But

in the real world, we tend to use the primitive Reed-Solomon codes of high rates.

II. Our results on hardness of decoding

Our main result of this paper is to remove the restriction on rate. The starting point of our results is the following lemma which we proved in [3]. Let $h \geq 2$ be a positive integer. Let $\mathbf{h}(x)$ be a monic irreducible polynomial in $\mathbf{F}_q[x]$ of degree h. Let α be a root of $\mathbf{h}(x)$ in an extension field of $\mathbf{F}_q$. Then, $\mathbf{F}_q[\alpha] = \mathbf{F}_{q^h}$ is a finite field of q^h element. We have

Lemma 1: If every element of $\mathbf{F}^*_{q^h}$ can be written as a product of exactly g distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_q$, then the discrete logarithm over $\mathbf{F}^*_{q^h}$ can be efficiently solved in random time $q^{O(1)}$ with oracle access to either a bounded distance decoder of $RS_q[q, g-h]$ at distance $q-g$, or a maximum likelihood decoder of $RS_q[q, g-h]$.

Two simple observations are crucial for us to obtain the new results in this paper.

- If every element of $\mathbf{F}^*_{q^h}$ can be written as a product of exactly g distinct elements in $\alpha + \mathbf{F}_q$, then every element of $\mathbf{F}^*_{q^h}$ can be written as a product of exactly $q-g$ distinct elements in $\alpha + \mathbf{F}_q$.
- Let α be an element in $\mathbf{F}_{q^{mh}}$ such that $\mathbf{F}_q[\alpha] = \mathbf{F}_{q^{mh}}$. If every element in $\mathbf{F}^*_{q^{mh}}$ can be written as a product of g_1 many distinct elements in $\alpha + \mathbf{F}_q$, then for any nonnegative integer $g_2 \leq q^m - q$, every element in $\mathbf{F}^*_{q^{mh}}$ can be written as a product of $g_1 + g_2$ many distinct elements in $\alpha + \mathbf{F}_{q^m}$.

Our main theorem states:

Theorem 1: Let $\delta > 0$ be a constant. Let q be a prime power. Let $m > 1$ be an integer. Suppose h and k are positive integers satisfying

$$h \leq \frac{q^{\frac{1}{2+\delta}}}{m} + \frac{1}{m}, h \leq \frac{\sqrt{q}}{m(\frac{4}{\delta}+2)} - \frac{1}{m}$$

and

$$q \leq k \leq q^m - q.$$

The discrete logarithm in $\mathbf{F}^*_{q^{mh}}$ can be solved in randomized time $(q^m)^{O(1)}$ with oracle access to a maximum likelihood decoder of $RS_{q^m}[q^m, k]$.

The discrete logarithm problem over finite fields is well studied in computational number theory. It is not believed to have a polynomial time algorithm. Many cryptographic protocols base their security on this assumption. The fastest general purpose algorithm [7] solves the discrete logarithm problem over finite field $\mathbf{F}^*_{q^h}$ in conjectured time

$$exp(O((\log q^h)^{1/3}(\log\log q^h)^{2/3})).$$

Thus, in the above theorem, it is best to take h as large as possible in order for the discrete logarithm to be hard. If $h = q^{\Theta(1)}$, this complexity is superpolynomial on q. The above theorem rules out a polynomial time algorithm for the maximum likelihood decoding problem of Reed-Solomon code of any rate under a cryptographic hardness assumption. Interestingly our computational lower bound for decoding Reed-Solomon codes is not sensitive to their dimensions. To obtain some intuition from the theorem, we set $m = 2$ and $\delta = 0.1$ and conclude:

Corollary 1: Assume that there is no randomized algorithm solving in time $q^{O(1)}$ the discrete logarithm over $\mathbf{F}^*_{q^{2h}}$ for all $h \leq q^{0.4}$. Let $k(x)$ be a function in $\mathbf{Z}^+ \to \mathbf{Z}^+$ computable in time polynomial in $x^{O(1)}$ and

$$\sqrt{x} \leq k(x) \leq x - \sqrt{x}.$$

Then there is no polynomial time maximum likelihood decoder for the code family $RS_q[q, k(q)]$.

In other words, no polynomial algorithm exists to solve the maximum likelihood decoding of $RS_q[q, k(q)]$ if $\sqrt{q} \leq k(q) \leq q - \sqrt{q}$, under well-studied cryptographic hardness assumption. In particular, under the assumption, for any constant $0 < c < 1$, there is no polynomial time maximum likelihood decoder for $RS_q[q, \lfloor cq \rfloor]$. Furthermore, no algorithm is known which can solve the discrete logarithm over $\mathbf{F}^*_{q^{2h}}$ for infinitely many q and all $h \leq q^{0.4}$ in time $q^{O(1)}$. Under the reasonable assumption that such algorithm does not exist, there does not exist a polynomial time algorithm to solve the maximum likelihood decoding of $RS_{q^2}[q^2, k(q^2)]$ for infinitely many q.

It is well known that Reed-Solomon codes possess a polynomial time unique decoder, which is a bounded-distance decoder at distance $\frac{1}{2}$ of the minimum distance. We prove however under a cryptographic hardness assumption that there does not exist an efficient bounded-distance decoder for primitive Reed-Solomon codes at distance $\frac{2}{3} + \epsilon$ of the minimum distance.

Theorem 2: Let ϵ be a positive constant less than $1/3$. There does not exist a randomized polynomial time bounded-distance decoder at distance $(2/3 + \epsilon)d$ for the Reed-Solomon code $RS_q[q, k]$, where $d = q - k + 1$ is the minimum distance, unless the discrete logarithm problem over $\mathbf{F}_{q^h}$ can be solved in randomized time $q^{O(1)}$ for any $h \leq q^{0.8\epsilon}$.

We comment that the above theorem does not contradict to the efficient list decoding algorithm in [6], since the code in our proof has rate approaching one.

In [4], the authors asked whether one can establish NP-hardness of maximum-likelihood decoding for a nontrivial family of binary codes. Though we do not solve the problem, we can establish cryptographic hardness of maximum-likelihood decoding of binary codes, obtained from concatenation of Reed-Solomon codes $RS_{2^m}(2^m, k)$ with $(2^m, m)$-Hadamard codes, denoted by $RSH_2(m2^m, k)$.

Corollary 2: Let ϵ be a positive constant less than $1/3$. There does not exist a randomized polynomial time bounded-distance decoder at distance $(2/3 + \epsilon)d$ for $RSH_2[m2^m, k]$, where $d = 2^{m-1}(2^m - k + 1)$ is the minimum distance, unless that the discrete logarithm in $\mathbf{F}_{2^{mh}}$ can be solved in randomized time $(2^m)^{O(1)}$ for any $h \leq 2^{0.8\epsilon m}$.

A. Our results on finding Hamming balls with many codewords

By a direct counting argument, for any positive integer $r < q - k$, there exists a Hamming ball of radius r containing at least $\binom{q}{r}/q^{q-r-k}$ many codewords in Reed-Solomon code $RS_q[q, k]$. Thus, if $k = \lfloor cq \rfloor$ for a constant $0 < c < 1$, we

set $r = \lfloor q - k - q^{1/4} \rfloor$ and the number of codewords in the Hamming ball will be exponential in q. However, finding such a Hamming ball deterministically is an open problem. There is some progress on the problem [5][1], but all the results are for codes of diminishing rates. Our contribution to this problem is to remove the rate restriction.

Theorem 3: Let $0 < c_1 < c_2 < 1$ be two real numbers. There exists a deterministic algorithm that given a prime power q and an integer k satisfying $c_1 q^2 \leq k \leq c_2 q^2$, runs in time $q^{O(1)}$, outputs a vector $v \in \mathbf{F}_{q^2}^{q^2}$ such that the Hamming ball centered at v and of radius $q^2 - k - q^{0.4}$ contains $exp(\Omega(q^2))$ many codewords in $RS_{q^2}[q^2, k]$.

Our construction allows the information rate to be positive. On the other hand, the ratio between the Hamming ball radius $q^2-k-q^{0.4}$ and the minimum distance q^2-k+1 is approaching 1, as is in [5][1]. The following result shows that we can decrease the radius of Hamming ball so that it is smaller than the minimum distance by a constant factor less than 1 if we work with codes with information rates going to one.

Theorem 4: For any real number $\rho \in (2/3, 1)$, there is a deterministic algorithm that, given a prime power q, outputs a positive integer $k = q - o(\sqrt{q})$ and a vector $v \in \mathbf{F}_q^q$ such that the Hamming ball centered at v and of radius $\lfloor \rho(q - k + 1) \rfloor$ contains at least $exp(q^{0.8(\rho - 2/3)})$ many codewords in $RS_q[q, k]$. The algorithm has time complexity $q^{O(1)}$. Note that the information rate is $1 - o(1)$.

It would be interesting for future research to extend the result to all $\rho \in (1/2, 1)$ and to prove a similar result with both positive information rate and the ratio between the Hamming ball radius and minimum distance less than 1.

III. Proof of Lemma 1

For readers' convenience, in this section, we sketch the main ideas in our earlier paper [3]. This will be the starting point of our new results in the present paper. In [3], the result was stated only for weaker bounded distance decoding. See that paper for a full proof.

Proof of Lemma 1. Let $\mathbf{h}(x)$ be a monic irreducible polynomial of degree $h > 1$ in $\mathbf{F}_q[x]$. We shall identify the extension field $\mathbf{F}_{q^h}$ with the residue field $\mathbf{F}_q[x]/(\mathbf{h}(x))$. Let α be the class of x in $\mathbf{F}_q[x]/(\mathbf{h}(x))$. Then, $\mathbf{F}_q[\alpha] = \mathbf{F}_{q^h}$. Consider the Reed-Solomon code $RS_q[q, g-h]$. For a polynomial $f(x) \in \mathbf{F}_q[x]$ of degree at most $h-1$, let u_f be the received word

$$u_f = (\frac{f(a)}{\mathbf{h}(a)} + a^{g-h})_{a \in \mathbf{F}_q}.$$

By assumption, we can write

$$f(\alpha) = \prod_{i=1}^{g} (\alpha + a_i),$$

where $a_i \in \mathbf{F}_q$ are distinct. It follows that as polynomials, we have the identity

$$\prod_{i=1}^{g} (x + a_i) = f(x) + t(x)\mathbf{h}(x),$$

where $t(x) \in \mathbf{F}_q[x]$ is some monic polynomial of degree $g-h$. Thus,

$$\frac{f(x)}{\mathbf{h}(x)} + x^{g-h} + (t(x) - x^{g-h}) = \frac{\prod_{i=1}^{g}(x + a_i)}{\mathbf{h}(x)},$$

where $t(x) - x^{g-h} \in \mathbf{F}_q[x]$ is a polynomial of degree at most $g-h-1$ and thus corresponds to a codeword. This equation implies that the distance of the received word u_f to the code $RS_q[q, g-h]$ is at most $q-g$. If the distance is smaller than $q-g$, then one gets a monic polynomial of degree g with more than g distinct roots. Thus, the distance of u_f to the code is exactly $q-g$.

Let C_f be the set of codewords in $RS_q[q, g-h]$ that have distance exactly $q-g$ to the received word u_f. The cardinality of C_f is then equal to $\frac{1}{g!}$ times the number of ordered ways that $f(\alpha)$ can be written as a product of exactly g distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_q$. For error radius $q-g$, the maximum likelihood decoding of the received word u_f is the same as finding a solution to the equation

$$f(\alpha) = \prod_{i=1}^{g} (\alpha + a_i),$$

where $a_i \in \mathbf{F}_q$ being distinct.

To show that the discrete logarithm in $\mathbf{F}_{q^h}^*$ can be reduced to the decoding of the words of the type u_f, we apply the index calculus algorithm. Let $b(\alpha)$ be a primitive element of $\mathbf{F}_{q^h}^*$. Taking $f(\alpha) = b(\alpha)^i$ for a random $0 \leq i \leq q^h - 2$, the maximum likelihood decoding of the word u_f gives a relation

$$b(\alpha)^i = \prod_{j=1}^{g} (\alpha + a_j(i)),$$

where $a_j(i) \in \mathbf{F}_q$ are distinct for $1 \leq j \leq g$. This gives the congruence equation

$$i \equiv \sum_{j=1}^{g} \log_{b(\alpha)}(\alpha + a_j(i)) \pmod{q^h - 1}.$$

Repeating the decoding and let i vary, this would give enough linear equations in the q variables $\log_{b(\alpha)}(\alpha + a)$ $(a \in \mathbf{F}_q))$. Solving the linear system modulo $q^h - 1$, one finds the values of $\log_{b(\alpha)}(\alpha + a)$ for all $a \in \mathbf{F}_q$. To compute the discrete logarithm of an element $v(\alpha) \in \mathbf{F}_{q^h}^*$ with respect to the base $b(\alpha)$, one applies the decoding to the element $v(\alpha)$ and finds a relation

$$v(\alpha) = \prod_{j=1}^{g} (\alpha + b_j),$$

where the $b_j \in \mathbf{F}_q$ are distinct. Then,

$$\log_{b(\alpha)} v(\alpha) \equiv \sum_{j=1}^{g} \log_{b(\alpha)}(\alpha + b_j) \pmod{q^h - 1}.$$

In this way, the discrete logarithm of $v(\alpha)$ is computed. The detailed analysis can be found in [3]. □

In order to use the above theorem, one needs to get good information on the integer g satisfying the assumption of the theorem. This is a difficult theoretical problem in general. It can be done in some cases, with the help of Weil's character

sum estimate together with a simple sieving. In particular, the following result was proved in [3].

Theorem 5: Let $h < g$ be positive integers. Let

$$N(g,h) = \frac{1}{g!}\left(\frac{q^g - \binom{g}{2}q^{g-1}}{q^h - 1} - (1+\binom{g}{2})(h-1)^g q^{g/2}\right).$$

Then every element in $\mathbf{F}_{q^h}^*$ can be written in at least $N(g,h)$ ways as a product of exactly g distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_q$.

If for some constant $\delta > 0$, we have

$$q \geq \max(g^2, (h-1)^{2+\delta}), \quad g \geq (\frac{4}{\delta}+2)(h+1),$$

then

$$N(g,h) \geq q^{g/2}/g! > 0.$$

The main draw back of the above theorem is the condition $q \geq g^2$, which translates to the condition that the information rate $(g-h)/q$ goes to zero in applications.

IV. The proof of Theorem 2 and Theorem 4

To prove Theorem 2, we start with a lemma.

Lemma 2: Let g, h be positive integers such that for some constant $\delta > 0$, we have

$$q \geq \max(g^2, (h-1)^{2+\delta}), \quad g \geq (\frac{4}{\delta}+2)(h+1).$$

1) Every element in $\mathbf{F}_{q^h}^*$ can be written in at least $N(g,h)$ ways as a product of exactly $q-g$ distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_q$.
2) Let $\mathbf{h}(x)$ be an irreducible polynomial of degree h over $\mathbf{F}_q$ and let $f(x)$ be a nonzero polynomial of degree less than h over $\mathbf{F}_q$. Then in Reed-Solomon code $RS_q[q, q-g-h]$, the Hamming ball centered at $(\frac{f(a)}{\mathbf{h}(a)} + a^{q-g-h})_{a\in\mathbf{F}_q}$ of radius g contains at least $\frac{q^{g/2}}{g!}$ many codewords.

To prove this lemma, we observe that the map that sends $\beta \in \mathbf{F}_{q^h}^*$ to $\prod_{a\in\mathbf{F}_q}(\alpha+a)/\beta$ is one-to-one from $\mathbf{F}_{q^h}^*$ to itself.

Proof: Note that

$$\prod_{a\in\mathbf{F}_q}(\alpha + a) \neq 0.$$

Given an element $\beta \in \mathbf{F}_{q^h}^*$, from Theorem 5, we have that $\prod_{a\in\mathbf{F}_q}(\alpha+a)/\beta$ can be written in at least $N(g,h)$ ways as a product of exactly g distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_q$, hence β can be written in at least $N(g,h)$ ways as a product of exactly $q-g$ distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_q$.

To prove the second assertion, we follow an argument similar to the proof of Lemma 1. Observe that the number of codewords in the Hamming ball centered at $(\frac{f(a)}{\mathbf{h}(a)} + a^{q-g-h})_{a\in\mathbf{F}_q}$ of radius g is exactly $\frac{1}{g!}$ times the number of ordered ways that $f(\alpha)$ can be written as a product of exactly $q-g$ distinct linear factors of the form $\alpha+a$ with $a \in \mathbf{F}_q$, which is at least $N(g,h) > q^{g/2}/g!$. □

Now we are ready to prove Theorem 2:

Proof of Theorem 2: Set $\delta = \frac{1}{2\epsilon}$ and $g = \frac{2+3\epsilon}{1-3\epsilon}(h+1)$. We can verify that

$$q \geq \max(g^2, (h-1)^{2+\delta}), \quad g \geq (\frac{4}{\delta}+2)(h+1)$$

hold for q big enough, since $h \leq q^{0.8\epsilon}$. Thus it follows from Lemma 1 that the bounded distance decoding of $RS_q[q, q-g-h]$ at distance

$$q - (q-g) = g = (2/3+\epsilon)(g+h+1) = (2/3+\epsilon)d$$

is at least as hard as the discrete logarithm over the finite field $\mathbf{F}_{q^h}^*$. Note that the rate $(q-g-h)/q$ approaches 1 as q increases.

Proof of Theorem 4: We set

$$\epsilon = \rho - 2/3, h = q^{0.8\epsilon}, \delta = \frac{1}{2\epsilon}, \text{ and } g = \frac{2+3\epsilon}{1-3\epsilon}(h+1).$$

One can verify that

$$q \geq \max(g^2, (h-1)^{2+\delta}), \quad g \geq (\frac{4}{\delta}+2)(h+1)$$

hold for q big enough. We find an irreducible polynomial $\mathbf{h}(x)$ of degree h over $\mathbf{F}_q$ using the algorithm in [9]. It follows from the second assertion in the above lemma that the number of codewords in the Hamming ball centered at

$$v = (\frac{1}{\mathbf{h}(a)} + a^{q-g-h})_{a\in\mathbf{F}_q}$$

of radius $g = (2/3+\epsilon)d$ is

$$\frac{q^{g/2}}{g!} > (\sqrt{q}/g)^g > exp(h) = exp(q^{0.8\epsilon}).$$

□

V. The proof of Theorem 1 and Theorem 3

We now consider the case where the rate is positive less than one. The main new idea for this case is to exploit the role of subfields. For this purpose, we take a positive integer $m \geq 2$. Let α be an element in $\mathbf{F}_{q^{mh}}$ with $\mathbf{F}_q[\alpha] = \mathbf{F}_{q^{mh}}$. Since

$$\mathbf{F}_q[\alpha] \subseteq \mathbf{F}_{q^m}[\alpha] \subseteq \mathbf{F}_{q^{mh}},$$

we also have $\mathbf{F}_{q^{mh}} = \mathbf{F}_{q^m}[\alpha]$.

Theorem 6: Let g_1 and g_2 be non-negative integers with $g_2 \leq q^m - q$. Let

$$N'(g_1, g_2, h, m) = N(g_1, mh)\binom{q^m - q}{g_2}.$$

Then, every element in $\mathbf{F}_{q^{mh}}^*$ can be written in at least $N'(g_1, g_2, h, m)$ ways as a product of exactly $g_1 + g_2$ distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_{q^m}$.

If for some constant $\delta > 0$, we have

$$q \geq \max(g_1^2, (mh-1)^{2+\delta}), \quad g_1 \geq (\frac{4}{\delta}+2)(mh+1)$$

then

$$N'(g_1, g_2, h, m) \geq \frac{q^{g_1/2}}{g_1!}\binom{q^m - q}{g_2} > 0.$$

Proof. Since $g_2 \leq q^m - q$, we can choose g_2 distinct elements $b_1, \cdots, b_{g_2}$ from the set $\mathbf{F}_{q^m} - \mathbf{F}_q$. There are $\binom{q^m-q}{g_2}$

many choices. For any element $\beta \in \mathbf{F}_{q^{mh}}^*$, since $\mathbf{F}_q[\alpha] = \mathbf{F}_{q^{mh}}$, we can apply Theorem 5 to deduce that

$$\frac{\beta}{(\alpha+b_1)\cdots(\alpha+b_{g_2})} = (\alpha+a_1)\cdots(\alpha+a_{g_1}),$$

where the $a_i \in \mathbf{F}_q$ are distinct. The number of such sets $\{a_1, a_2, a_3, \cdots, a_{g_1}\} \subseteq \mathbf{F}_q$ is greater than $N(g_1, mh)$. Since $\mathbf{F}_q$ and its complement $\mathbf{F}_{q^m} - \mathbf{F}_q$ are disjoint, it follows that

$$\beta = (\alpha+b_1)\cdots(\alpha+b_{g_2})(\alpha+a_1)\cdots(\alpha+a_{g_1})$$

is a product of exactly $g_1 + g_2$ distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_{q^m}$. □

Theorem 7: Let $m \geq 2$ and $h \geq 2$ be two positive integers. Let q be a prime power and k be an integer satisfying $\sqrt{q^m} < k < q^m - \sqrt{q^m}$. Assume that

$$q \geq \max((mh-1)^{2+\delta}, (\frac{4}{\delta}+2)^2(mh+1)^2)$$

for some constant $\delta > 0$.

1) Every element in $\mathbf{F}_{q^{mh}}^*$ can be written as a product of exactly $k+h$ distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_{q^m}$.
2) Let $\mathbf{h}(x)$ be an irreducible polynomial of degree h over $\mathbf{F}_{q^m}$ whose root α satisfies that $\mathbf{F}_q[\alpha] = \mathbf{F}_{q^{mh}}$. Let $f(x)$ be a nonzero polynomial over $\mathbf{F}_{q^m}$ of degree less than h. Then in the Reed-Solomon code $RS_{q^m}[q^m, k]$, the Hamming ball centered at $(\frac{f(a)}{\mathbf{h}(a)} + a^k)_{a \in \mathbf{F}_{q^m}}$ of radius $q^m - k - h$ contains at least

$$\frac{q^{\lfloor\sqrt{q}\rfloor/2}}{\lfloor\sqrt{q}\rfloor!}\binom{q^m - q}{k+h-\lfloor\sqrt{q}\rfloor}$$

many codewords.

Proof: Take $g_1 = \lfloor q^{1/2} \rfloor$ and we have

$$g_1 \geq (\frac{4}{\delta}+2)(mh+1) \text{ and } q \geq g_1^2.$$

The conditions

$$q \geq \max(g_1^2, (mh-1)^{2+\delta}), \quad g_1 \geq (\frac{4}{\delta}+2)(mh+1)$$

hold. Furthermore we have

$$0 \leq k - g_1 + h \leq q^m - q.$$

Now take $g_2 = k - g_1 + h$. According to Theorem 6, every element in $\mathbf{F}_{q^{mh}}^*$ can be written in at least $N'(g_1, g_2, h, m)$ ways as a product of exactly

$$g_1 + g_2 = k + h$$

distinct linear factors of the form $\alpha + a$ with $a \in \mathbf{F}_{q^m}$. And in the Reed-Solomon code $RS_{q^m}[q^m, k]$, the Hamming ball centered at $(\frac{f(a)}{\mathbf{h}(a)} + a^k)_{a \in \mathbf{F}_{q^m}}$ of radius $q^m - k - h$ contains at least $N'(g_1, g_2, h, m)$ many codewords. Finally

$$N'(g_1, g_2, h, m) \geq \frac{q^{\lfloor\sqrt{q}\rfloor/2}}{\lfloor\sqrt{q}\rfloor!}\binom{q^m - q}{k+h-\lfloor\sqrt{q}\rfloor} > 0$$

□

Proof of Theorem 1 and Theorem 3. Theorem 1 follows directly from the above theorem and Lemma 1 by setting $m = 2$.

Set $m = 2$, $h = q^{0.4}$, $k = q^2/2$ and $\delta = 0.1$ in the above theorem. We can verify that the conditions are satisfied. Hence the number of codewords in the Hamming ball centered at

$$v = (\frac{1}{\mathbf{h}(a)} + a^k)_{a \in \mathbf{F}_{q^2}}$$

of radius $q^2 - k - h$ contains at least

$$\frac{q^{\lfloor\sqrt{q}\rfloor/2}}{\lfloor\sqrt{q}\rfloor!}\binom{q^2 - q}{k+h-\lfloor\sqrt{q}\rfloor} = exp(\Omega(q^2))$$

many codewords in $RS_{q^2}[q^2, k]$. It remains to find an irreducible polynomial of degree h over $\mathbf{F}_{q^2}$, whose root α satisfies that $\mathbf{F}_q[\alpha] = \mathbf{F}_{q^{2h}}$. Let p be the characteristic of $\mathbf{F}_q$. We can use α such that $\mathbf{F}_p[\alpha] = \mathbf{F}_{q^{2h}}$. We need to find an irreducible polynomial of degree $h \log_p(q^2)$ over $\mathbf{F}_p$. It can be done in time polynomial in p and the degree [9]. Then we factor the polynomial over $\mathbf{F}_{q^2}$, which can be done in deterministic time $q^{O(1)}$, and take any factor to be $\mathbf{h}(x)$. □

VI. Conclusion and future research

In this paper, we show that the maximum likelihood decoding of the primitive Reed-Solomon code is at least as hard as the discrete logarithm over finite fields for any given information rate. We also prove a hardness result for the bounded-distance decoding of primitive Reed-Solomon codes at radius $2/3 + \epsilon$ of the minimum distance. It is a very interesting problem whether $2/3 + \epsilon$ can be improved to $1/2 + \epsilon$. We feel that substantially new ideas are required. Some codes in our proof are defined over finite fields of composite cardinalities. While this is not a problem in practical applications, e.g. $q = 256$ is quite popular, it would be interesting to remove this restriction, that is, allowing prime finite fields as well.

Many important questions about decoding Reed-Solomon codes remain open. For example, does there exist a Hamming ball of radius less than the minimum distance by a constant factor smaller than one that contains superpolynomially many codewords in Reed-Solomon codes of rate less than one? Another interesting problem is whether the primitive Reed-Solomon maximum likelihood decoding problem is equivalent to the discrete logarithm problem over finite fields. In other words, if we have oracle access to a discrete logarithm solver over finite fields, can we solve the maximum likelihood decoding problem for primitive Reed-Solomon codes? If so, this would imply that the problem is unlikely to be NP-hard, since discrete logarithm over finite fields are not believed to be NP-hard.

References

[1] Eli Ben-Sasson, Swastik Kopparty, and Jaikumar Radhakrishnan. Subspace polynomials and list decoding of Reed-Solomon codes. In 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 207–216, 2006.

[2] Qi Cheng and Elizabeth Murray. On deciding deep holes of Reed-Solomon codes. In Proceedings of Annual Conference on Theory and Applications of Models of Computation(TAMC), volume 4484 of Lecture Notes in Computer Science, pages 296–305. Springer-Verlag, 2007.

[3] Qi Cheng and Daqing Wan. On the list and bounded distance decodability of Reed-Solomon codes. SIAM Journal on Computing, 37(1):195–209, 2007. Special Issue on FOCS 2004.

[4] V. Guruswami and A. Vardy. Maximum-likelihood decoding of Reed-Solomon codes is NP-hard. IEEE Transactions on Information Theory, 51(7):2249–2256, 2005.

[5] Venkatesan Guruswami and Atri Rudra. Limits to list decoding Reed-Solomon codes. IEEE Transactions on Information Theory, 52(8):3642–3649, 2006.

[6] Venkatesan Guruswami and Madhu Sudan. Improved decoding of Reed-Solomon and algebraic-geometry codes. IEEE Transactions on Information Theory, 45(6):1757–1767, 1999.

[7] Antoine Joux, Reynald Lercier, Nigel Smart, and Frederik Vercauteren. The number field sieve in the medium prime case. In Advances in Cryptology - CRYPTO 2006, volume 4117 of Lecture Notes in Computer Science, pages 326–344. Springer-Verlag, 2006.

[8] Jiyou Li and Daqing Wan. On the subset sum problem over finite fields. Finite Fields and Applications, 14(4):911–929, 2008.

[9] Victor Shoup. New algorithms for finding irreducible polynomials over finite fields. Mathematics of Computation, 54:435–447, 1990.