Bits Security of the CDH Problems over Finite Fields

Mingqiang Wang[1], Tao Zhan[1], and Haibin Zhang[2]

[1]School of Mathematics, Shandong University
wangmingqiang@sdu.edu.cn,zhantao@moe.edu.cn
[2]Department of Computer Science, University of California Davis
hbzhang@ucdavis.edu

Abstract

It is a long-standing open problem to prove the existence of (deterministic) hard-core predicates for the Diffie-Hellman problem over finite fields, without resorting to the *generic* approaches for any one-way functions (*e.g.,* the Goldreich-Levin hard-core predicates). Fazio *et al.* (FGPS, Crypto '13) make important progress on this problem by defining a *weaker* Computational Diffie-Hellman (CDH) problem over $\mathbb{F}_{p^2}$, *i.e.,* Partial-CDH problem, and proving the unpredictability of every single bit of one of the coordinates of the secret Diffie-Hellman value. However, the existence of *specific* hard-core predicates for the *regular* CDH problems defined over finite fields remains unproven. This paper closes this gap and resolves all the open problems left in FGPS:

1 We prove that the Partial-CDH problem over finite fields $\mathbb{F}_{p^2}$ is as hard as the regular CDH problem over the same fields.

2 We show a much stronger and more generalized result over finite fields $\mathbb{F}_{p^2}$—not only the *regular* CDH problem over $\mathbb{F}_{p^2}$ admits hard-core predicates but *every* individual bit of the CDH value is unpredictable.

3 We extend the Partial-CDH problem to define the d-th CDH problem over finite fields $\mathbb{F}_{p^t}$ for any polynomial $t > 1$ and for any $0 \leq d \leq t-1$. We show that computing *any* single coordinate of the CDH value over $\mathbb{F}_{p^t}$ is equivalent to computing the entire CDH value.

4 We prove that over finite fields $\mathbb{F}_{p^t}$ for any polynomial $t > 1$, each d-th CDH problem except $d \neq 0$ admits a large class of hard-core predicates, including every individual bit of the d-th coordinate. Hence almost all individual bits of the CDH value of the regular CDH problem over finite fields $\mathbb{F}_{p^t}$ for $t > 1$ are hard-core.

Key words: CDH, Diffie-Hellman problem, d-th CDH problem, finite fields, hard-core bits, list decoding, multiplication code, Partial-CDH problem.

1 Introduction

Hard-core predicates [4, 12] are central to cryptography. Of particular interest is the hard-core predicate for the CDH problem, which is essential to establishing the security for Diffie-Hellman (DH) key exchange protocol [6] and ElGamal encryption scheme [8] without having to make a (potentially) much stronger DH assumption—the Decisional Diffie-Hellman (DDH) assumption.

However, despite the generic approaches for *randomized* predicates working for any computationally hard problems [11, 16], showing the existence of *deterministic* and *specific* hard-core predicates for the CDH problem over *finite fields* has proven elusive. This is in contrast to other conjectured hard problems such as discrete logs, RSA, and Rabin, whose deterministic hard-core predicates were discovered roughly three decades ago [2, 4]. Recently, Fazio, Gennaro, Perera, and Skeith III (FGPS) [9] made a significant breakthrough by introducing a relaxed variant of the CDH problem over finite fields $\mathbb{F}_{p^2}$, *i.e.,* the Partial-CDH problem and proving the unpredictability for a large class of predicates.

Partial-CDH problem. Given a prime p, there are many different fields $\mathbb{F}_{p^2}$ which are all isomorphic to each other. Fixing a monic irreducible polynomial of degree 2 in $\mathbb{F}_p$: $h(x) = x^2 + h_1 x + h_0$, it is well known that $\mathbb{F}_{p^2}$ is isomorphic to the field $\mathbb{F}_p[x]/(h)$, where $(h(x))$ is a principal ideal in the polynomial ring $\mathbb{F}_p[x]$ and elements of $\mathbb{F}_{p^2}$ can be written as linear polynomials. Namely, if $g \in \mathbb{F}_{p^2}$ then $g = g_1 x + g_0$ and addition and multiplication are performed as polynomial operations modulo h. Given $g \in \mathbb{F}_{p^2}$ we denote by $[g]_i$ the coefficient of the degree-i term.

Let g denote a random generator of the multiplicative group of $\mathbb{F}_{p^2}$. FGPS defined the following Partial-CDH problem over $\mathbb{F}_{p^2}$ [9]: the Partial-CDH problem is hard over $\mathbb{F}_{p^2}$ if given random inputs $g, A = g^a$, $B = g^b \in \mathbb{F}_{p^2}$, it is computationally hard to output $K = [g^{ab}]_1 \in \mathbb{F}_p$ (*i.e.,* the coefficient of the degree 1 term of g^{ab}), for any representation of $\mathbb{F}_{p^2}$.

Assuming the hardness of the Partial-CDH problem, FGPS developed the idea of randomizing the problem representation originally suggested by Boneh and Shparlinski [5] and proved a large class of hard-core predicates over a *random representation* of the finite field $\mathbb{F}_{p^2}$. Namely, given an oracle that predicts any bit of $K = \left[g^{ab}\right]_1$ over a random representation of $\mathbb{F}_{p^2}$ with non-negligible advantage, one can recover K with non-negligible probability.

Open problems. The Partial-CDH problem is clearly weaker than the regular CDH problem. Given a CDH oracle, one can easily solve the Partial-CDH problem. Note that the reason why we need hard-core predicates is exactly that we do not want to make stronger assumptions. Without characterizing the hardness of the Partial-CDH problem, the FGPS result can hardly be based on a firm foundation. Thus, the most natural open question that FGPS raised [9, Section 6] is how strong the Partial-CDH assumption is. FGPS also asked if it is possible to reduce the Partial-CDH problem over $\mathbb{F}_{p^t}$ to the regular CDH problem over $\mathbb{F}_p$.

The third open question which FGPS raised seems particularly challenging (too): Can their results be extended to *general* finite fields $\mathbb{F}_{p^t}$ for $t > 1$? As argued by FGPS [9, Section 6], their original techniques could not be used directly and "must be augmented somehow."

Of course, the most fundamentally important open question remains the existence of hard-core predicates for the regular CDH problems defined over finite fields without using the generic approaches. Indeed, according to their paper [9, p. 5], FGPS originally hoped to prove the existence of hard-core predicates for the regular CDH problem over $\mathbb{F}_{p^2}$.

1.1 Our contributions

In this paper, we prove the existence of the hard-core predicates for the regular CDH problem over finite field $\mathbb{F}_{p^t}$ for polynomial $t > 1$ over a random representation of the finite field. Moreover, we show that all the individual bits of the CDH problem over $\mathbb{F}_{p^2}$ and *almost* all the individual bits of the CDH problem over $\mathbb{F}_{p^t}$ for $t > 1$ are hard-core. To achieve these results, we have developed new and useful techniques for characterizing the hardness of cryptographic assumptions over finite fields and proving hard-core predicates for hard problems. Our results also resolved all the open problems left in FGPS.

Hardness equivalence between the Partial-CDH and Regular CDH problems. We prove that somewhat surprisingly, over finite fields $\mathbb{F}_{p^2}$ the Partial-CDH assumption is *no* stronger than the regular CDH assumption over the same fields. Therefore, the Partial-CDH problem is as hard as the CDH problem over $\mathbb{F}_{p^2}$. With this result, the Partial-CDH assumption defined in FGPS can be regarded as a *reasonable* one. Correspondingly, our result provides a solid foundation for the main result in FGPS that the hard-core predicates can be acquired from the Partial-CDH problem.

All bits security of CDH problems over $\mathbb{F}_{p^2}$. Following the *list decoding* approach for hard-core predicates developed by Akavia, Goldwasser, and Safra [1] and extended by Morillo and Ràfols [15] as well as Duc and Jetchev [7], FGPS were able to construct a *multiplication* code for the Partial-CDH problem that is accessible, concentrated, and recoverable. Up to now, the list decoding approach has only been proven successful for multiplicative codes [1, 7, 15]. It is unclear if the approach can work more generally. In this paper, we will work *directly* on a non-multiplicative code. Assuming the hardness of the Partial-CDH problem, we are still able to prove the unpredictability of every single bit of the *other* coordinate (*i.e.,* the coefficient of the lower degree term) of the secret CDH value, by using a careful analysis of the Fourier coefficients of the function. To the best of our knowledge, this is the first positive result that list decoding approach can be applied to a non-multiplicative code, a result of independent interest.

Combining all the above-mentioned results, we are able to prove our main result for the regular CDH problem over $\mathbb{F}_{p^2}$: given an oracle $\mathcal{O}$ that predicts any bit of the CDH value over a random representation of the field $\mathbb{F}_{p^2}$ with non-negligible advantage, we can solve the *regular* CDH problem over $\mathbb{F}_{p^2}$ with non-negligible probability.

This result strengthens the FGPS result in two fundamental ways: first, our result is based on the standard CDH assumption over $\mathbb{F}_{p^2}$ instead of the Partial-CDH assumption; it also generalizes the FGPS hard-core predicates, including every single bit of the entire secret CDH value, not merely every single bit of one of the coordinates.

The d-th CDH problems. We extend the Partial-CDH problem by defining the d-th CDH problems over $\mathbb{F}_{p^t}$. Specifically, fixing an l-bit prime p and an integer t where $t > 1$ and $t \in poly(l)$, there are many different fields $\mathbb{F}_{p^t}$, but they are all isomorphic to each other. Let $h(x)$ be a monic irreducible polynomial of degree t in $\mathbb{F}_p$. It is well known that $\mathbb{F}_{p^t}$ is isomorphic to the field $\mathbb{F}_p[x]/(h)$, where $(h(x))$ is a principal ideal in the polynomial ring $\mathbb{F}_p[x]$ and elements of $\mathbb{F}_{p^t}$ can be written as polynomials of degree $t-1$. Namely, if $g \in \mathbb{F}_{p^t}$ then $g = g_{t-1}x^{t-1}+g_{t-2}x^{t-2}+\cdots+g_1x+g_0$. Addition and multiplication of the elements in $\mathbb{F}_{p^t}$ are performed as polynomial operations modulo h. In the following, given $g \in \mathbb{F}_{p^t}$ we denote by $[g]_i$ the coefficient of the degree-i term, *i.e.,* $g_i = [g]_i$.

Let g be a random generator of the multiplicative group of $\mathbb{F}_{p^t}$ and d be an integer such that $0 \leq d \leq t-1$. Informally we say that the d-th CDH problem is hard in $\mathbb{F}_{p^t}$ if given $g, g^a, g^b \in \mathbb{F}_{p^t}$, no efficient algorithm can compute $[g^{ab}]_d$, for any representations of $\mathbb{F}_{p^t}$.

We are able to prove a general equivalence result (with more involved techniques): all the d-th CDH problems over finite fields $\mathbb{F}_{p^t}$ (with $t > 1$) are as hard as the regular CDH problem over the same fields. Namely, we prove that over finite fields $\mathbb{F}_{p^t}$ computing any individual coordinate of secret CDH value is equivalent to computing the entire CDH value.

Almost all bits security of the CDH problems over $\mathbb{F}_{p^t}$ $(t > 1)$. We go on to prove that assuming the hardness of the d-th CDH problem, every single bit of the d-th CDH coordinate for $d \neq 0$ is hard-to-compute. FGPS [9, Section 6] found that their technique was not powerful enough to solve the generalized problem. To overcome the difficulty, we identify a *general* yet *simplified* class of isomorphisms. The isomorphisms identified generalize that of finite field $\mathbb{F}_{p^2}$ in FGPS to the case of general finite fields $\mathbb{F}_{p^t}$ for any $t > 1$. More importantly, they simplify that of FGPS by adopting a more restrictive class of isomorphisms. We comment that it is the simplicity that is essential to overcoming the original technical difficulty and establishing the bits security for general finite fields. To achieve this result, we also use another idea of Boneh and Shparlinski [5] on d-th residue modulo p.

Together with the equivalence result between all the d-th CDH problems over $\mathbb{F}_{p^t}$ (with $t > 1$) and the regular CDH problem, we obtain another main result of the paper: all bits except the bits of the degree-0 term of the usual CDH problem over a random representation of the finite field $\mathbb{F}_{p^t}$ are hard-core. It is legitimate to ask whether the above result can be applied to the case of the 0-th CDH coordinate. The question of closing this gap is left open.

2 Preliminaries

2.1 Notations

We use the standard symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{C}$ to denote the natural numbers, the integers, the real numbers and the complex numbers, respectively. Let $\mathbb{Z}_+$ and $\mathbb{R}_+$ stand for the positive integers and reals, respectively. A function $\nu(l)\colon \mathbb{N} \to \mathbb{R}$ is *negligible* if for every constant $c \in \mathbb{R}_+$ there exists $l_c \in \mathbb{N}$ such that $\nu(l) < l^{-c}$ for all $l > l_c$. A function $\rho(l)\colon \mathbb{N} \to \mathbb{R}$ is *non-negligible* if there exists a constant $c \in \mathbb{R}_+$ and $l_c \in \mathbb{N}$ such that $\rho(l) > l^{-c}$ for all $l > l_c$. For a Boolean function $f\colon \mathcal{D} \to \{\pm 1\}$ over an arbitrary domain $\mathcal{D}$, denote by $\mathsf{maj}_f = \mathsf{max}_{\{b=\pm 1\}} \Pr_{\alpha\in\mathcal{D}}[f(\alpha) = b]$ the *bias* of f toward its majority value.

2.2 Fourier Transform

Let $\mathbb{G}$ be a finite abelian group. For any two functions $f, g\colon \mathbb{G} \to \mathbb{C}$, their *inner product* is defined as $\langle f, g\rangle = 1/|\mathbb{G}| \sum_{x\in\mathbb{G}} \overline{f(x)}g(x)$. The l_2-norm of f on the vector space $\mathbb{C}(\mathbb{G})$ is defined as $\|f\|_2 = \sqrt{\langle f, f\rangle}$. A *character* of $\mathbb{G}$ is a homomorphism $\chi\colon \mathbb{G} \to \mathbb{C}^*$, *i.e.,* $\chi(x + y) = \chi(x)\chi(y)$ for all $x, y \in \mathbb{G}$. The set of all characters of $\mathbb{G}$ forms a *character group* $\widehat{\mathbb{G}}$, whose elements form an orthogonal basis (the *Fourier basis*) for the vector space $\mathbb{C}(\mathbb{G})$. One can then describe any function $f \in \mathbb{C}(\mathbb{G})$ via its *Fourier expansion* $\sum_{\chi\in\widehat{\mathbb{G}}} \widehat{f}(\chi)\chi$, where $\widehat{f}\colon \widehat{\mathbb{G}} \to \mathbb{C}$ is the *Fourier transform* of f and we have $\widehat{f}(\chi) = \langle f, \chi\rangle$. The coefficients $\widehat{f}(\chi)$ in the Fourier basis $\{\chi\}_{\chi\in\widehat{\mathbb{G}}}$ are the *Fourier coefficients* of f. The *weight* of a Fourier coefficient is denoted by $|\widehat{f}(\chi)|^2$. When $\mathbb{G} = \mathbb{Z}_n$ (*i.e.,* the additive group of integers modulo n) and $\widehat{\mathbb{G}} = \widehat{\mathbb{Z}}_n$, for each $\alpha \in \mathbb{Z}_n$, the α-character is defined as a function $\chi_\alpha\colon \mathbb{Z}_n \to \mathbb{C}$ such that $\chi_\alpha(x) = \omega_n^{\alpha x}$, where $\omega_n = e^{2\pi i/n}$. If Γ is a subset of $\mathbb{Z}_n$ then it is natural to consider the projection of f in set Γ, *i.e.,* $f_{|\Gamma} = \sum_{\alpha\in\Gamma} \widehat{f}(\alpha)\chi_\alpha$, where $\widehat{f}(\alpha) = \langle f, \chi_\alpha\rangle$. Since the characters are orthogonal, we have $\|f\|_2^2 = \sum_{\alpha\in\mathbb{Z}_n} |\widehat{f}(\alpha)|^2$ and $\|f_{|\Gamma}\|_2^2 = \sum_{\alpha\in\Gamma} |\widehat{f}(\alpha)|^2$.

Definition 1 (Fourier concentrated function [1]). *A function $f\colon \mathbb{Z}_n \to \mathbb{C}$ is Fourier ϵ-concentrated if there exists a set $\Gamma \subseteq \mathbb{Z}_n$ consisting of poly$(\log n, 1/\epsilon)$ characters, so that*

$$\|f - f_{|\Gamma}\|_2^2 = \sum_{\alpha\notin\Gamma} |\widehat{f}(\alpha)|^2 \le \epsilon.$$

A function f is called Fourier concentrated if it is Fourier ϵ-concentrated for every $\epsilon > 0$.

Definition 2 (τ-heavy characters [1]). *Given a threshold $\tau > 0$ and an arbitrary function $f\colon \mathbb{Z}_n \to \mathbb{C}$, we say that a character $\chi_\alpha \in \mathsf{Heavy}_\tau(f)$ is τ-heavy if the weight of its corresponding Fourier* *coefficient is at least τ. The set of all heavy characters is denoted by*

$$\mathsf{Heavy}_\tau(f) = \{\chi_\alpha : |\widehat{f}(\alpha)|^2 \geq \tau\}.$$

2.3 Error Correcting Codes: Definitions and Properties

Error correcting codes can encode messages into codewords by adding redundant data such that it can be recovered even in the presence of noise. The code to be discussed here encodes each element $\alpha \in \mathbb{Z}_n$ into a codeword C_α of length n. Each codeword C_α can be represented by a function $C_\alpha : \mathbb{Z}_n \to \{\pm 1\}$. We now recall a number of definitions and lemmata [1, 7] about codes over $\mathbb{Z}_n$.

Definition 3 (Fourier concentrated code). *A code $\mathcal{C} = \{C_\alpha : \mathbb{Z}_n \to \{\pm 1\}\}$ is concentrated if each of its codewords C_α is Fourier concentrated.*

Definition 4 (Recoverable code). *A code $\mathcal{C} = \{C_\alpha : \mathbb{Z}_n \to \{\pm 1\}\}$ is recoverable, if there exists a recovery algorithm that, given a character $\chi \in \widehat{\mathbb{Z}}_n$ and a threshold τ, returns in time poly$(\log n, 1/\tau)$ a list of all elements α associated with codewords C_α for which χ is a τ-heavy coefficient (i.e., $\{\alpha \in \mathbb{Z}_n : \chi \in \mathsf{Heavy}_\tau(C_\alpha)\}$).*

Lemma 1 below shows that in a concentrated code $\mathcal{C}$, any corrupted ("noisy") version $\widetilde{C}_\alpha$ of codeword C_α share at least one heavy coefficient with C_α. Lemma 2 shows that when given query access to any function f one can efficiently learn all its heavy characters.

Lemma 1 ([1, Lemma 1]). *Let $f, g : \mathbb{Z}_n \to \{\pm 1\}$ such that f is concentrated and for some $\epsilon > 0$,*

$$\Pr_{\alpha \in \mathbb{Z}_n} [f(\alpha) = g(\alpha)] \geq \mathsf{maj}_f + \epsilon.$$

There exists a threshold τ such that $1/\tau \in$ poly$(1/\epsilon, \log n)$, and there exists a non-trivial character $\chi \neq 0$ heavy for f and g: $\chi \in \mathsf{Heavy}_\tau(f) \cap \mathsf{Heavy}_\tau(g)$.

Lemma 2 ([1, Theorem 6]). *There is a probabilistic algorithm that, given query access to $w : \mathbb{Z}_n \to \{\pm 1\}$, $\tau > 0$ and $0 < \delta < 1$, outputs a list L of $O(1/\tau)$ characters containing $\mathsf{Heavy}_\tau(w)$ with probability at least $1 - \delta$, whose running time is $\widetilde{O}\left(\log(n) \cdot \ln^2 \frac{(1/\delta)}{\tau^{5.5}}\right)$.*

2.4 Review of List Decoding Approach for Hard-Core Predicates

Informally, a cryptographic one-way function $f : \mathcal{D} \to \mathcal{R}$ is a function which is easy to compute but hard to invert. Given a one-way function f and a predicate π, we say π is hard-core if there is an efficient probabilistic polynomial-time (PPT) algorithm that given $\alpha \in \mathcal{D}$ computes $\pi(\alpha)$, but there is no PPT algorithm $\mathcal{A}$ that given $f(\alpha) \in \mathcal{R}$ predicts $\pi(\alpha)$ with probability $\mathsf{maj}_\pi + \epsilon$ for a non-negligible ϵ.

Goldreich and Levin [11] showed hard-core predicates for general one-way functions by providing a general list decoding algorithm for Hadamard code. Akavia, Goldwasser, and Safra (AGS) [1] formalized the list decoding methodology and applied it to a broad family of conjectured one-way functions. In particular, they proved the unpredictability of *segment predicates* [1] for any one-way function f with the following *homomorphic* property: given $f(\alpha)$ and λ, one can efficiently compute $f(\lambda\alpha)$. This includes discrete logarithms in finite fields and elliptic curves, RSA, and Rabin. Morillo and Ràfols [15] extended the AGS result to prove the unpredictability of every individual bit for these functions. Duc and Jetchev [7] showed how to extend to elliptic curve-based one-way functions which do not necessarily enjoy the homomorphic property. Their result instead requires introducing a random description of the curve, an idea originally developed by Boneh and Shparlinski [5]. In their paper, Boneh and Shparlinski proved for the elliptic curve Diffie-Hellman problem that the least significant bit of each coordinate of the secret CDH value is hard-core over a random representation of the curve. Recently, FGPS extended the Boneh and Shparlinski idea to prove every individual bit (not merely the least significant bit) of the elliptic curve Diffie-Hellman problem is hard-core. By extending the same idea to the case of finite fields $\mathbb{F}_{p^2}$, FGPS also proved for a weak CDH problem (*i.e.* Partial-CDH problem) the unpredictability of every single bit of one of the coordinates of the secret CDH value.

List decoding approach overview. Given a one-way function $f\colon \mathcal{D} \to \mathcal{R}$ and a predicate π, one would have to identify an error-correcting code $\mathcal{C}^\pi = \{C_\alpha\colon \mathcal{D} \to \{\pm 1\}\}_{\alpha \in \mathcal{D}}$ such that every input α of the one-way function is associated with a codeword C_α. In particular, the code needs to satisfy the following properties:

(1) *Accessibility.* One should be able to obtain a corrupted ("noisy") version $\widetilde{C}_\alpha$ of the original codeword C_α. Such a corrupted codeword must be close to the original codeword, *i.e.,* $\Pr_\lambda[C_\alpha(\lambda) = \widetilde{C}_\alpha(\lambda)] > \mathsf{maj}_\pi + \epsilon$ for a non-negligible ϵ.

(2) *Concentration.* Each codeword C_α should be a Fourier concentrated function, *i.e.,* each codeword can be approximated by a small number of heavy coefficients in the Fourier representation.

(3) *Recoverability.* There exists a *poly*$(\log n, \tau^{-1})$ algorithm that on input a Fourier character χ and a threshold τ outputs a short list L_χ which contains all the values $\alpha \in \mathcal{D}$ such that χ is τ-heavy for the codeword C_α.

We now show how to invert $y = f(\alpha)$ with the prediction oracle Ω. Querying Ω will allow one to have access to a corrupted codeword $\widetilde{C}_\alpha$ that is close to C_α. According to Lemma 1, we know that there should exist a threshold τ and at least one Fourier character that is τ-heavy for both $\widetilde{C}_\alpha$ and C_α. Applying the learning algorithm in Lemma 2, we can find the set of all τ-heavy characters for $\widetilde{C}_\alpha$. Due to the recovery property, we are able to produce for each heavy character a polynomial size list containing possible α. Note that one can identify the correct α since f is efficiently computable.

List decoding via multiplication code. The crux of list decoding approach is to identify the "right" code which is accessible, concentrated, and recoverable. To this end, AGS and subsequent work either define a multiplication code, or transform the original code to an equivalent multiplication code. Such a multiplication code is of the form $C_\alpha(\lambda) = \pi(\lambda\alpha)$. Indeed, as argued in [1, 7], this is at the basis of their proofs: multiplication codes can be proven to satisfy concentration and recoverability.

In Section 4, we will directly work on a code that is *not* multiplicative. Not surprisingly, this makes it hard to prove code concentration and recoverability. To our knowledge, we are the first to apply the list decoding approach to the case of a non-multiplicative code.

3 The Partial-CDH Assumption is Equivalent to the CDH Assumption over $\mathbb{F}_{p^2}$

In this section, we show that over finite fields $\mathbb{F}_{p^2}$ the Partial-CDH problem [9] is as hard as the regular CDH problem. Fix an l-bit prime p. Let g be a random generator of the multiplicative group of $\mathbb{F}_{p^2}$. The Partial-CDH problem over $\mathbb{F}_{p^2}$ is a relaxed variant of the conventional CDH problem over $\mathbb{F}_{p^2}$, which we formally state as follows:

Assumption 1 (The CDH assumption over $\mathbb{F}_{p^2}$). *We say that the CDH problem is hard in $\mathbb{F}_{p^2}$ if for any PPT adversary $\mathcal{A}$, his CDH advantage*

$$\mathbf{Adv}^{\mathrm{cdh}}_{\mathcal{A},\mathbb{F}_{p^2}} = \Pr\left[\mathcal{A}(p,g,h,g^a,g^b) = g^{ab} \middle| a,b \overset{\$}{\leftarrow} \left\{1,\cdots,p^2-1\right\}\right]$$

is negligible in l.

Let $I_2(p)$ be the set of monic irreducible polynomials of degree 2 in $\mathbb{F}_p$. Informally we say that the *Partial-CDH* problem [9] is hard in $\mathbb{F}_{p^2}$ if for all $h \in I_2(p)$ no efficient algorithm given $g, A = g^a, B = g^b \in \mathbb{F}_{p^2}$ can output $\left[g^{ab}\right]_1 \in \mathbb{F}_p$. Formally we consider the following assumption:

Assumption 2 (The Partial-CDH assumption over $\mathbb{F}_{p^2}$ [9]). *We say that the Partial-CDH problem is hard in $\mathbb{F}_{p^2}$ if for any PPT adversary $\mathcal{A}$, his Partial-CDH advantage for all $h \in I_2(p)$*

$$\mathbf{Adv}^{\mathrm{pcdh}}_{\mathcal{A},\mathbb{F}_{p^2}} = \Pr\left[\mathcal{A}(p,g,h,g^a,g^b) = \left[g^{ab}\right]_1 \middle| a,b \overset{\$}{\leftarrow} \left\{1,\cdots,p^2-1\right\}\right]$$

is negligible in l.

It is easy to see that the Partial-CDH problem is weaker than the regular CDH problem over $\mathbb{F}_{p^2}$. The following theorem shows that somewhat surprisingly, the regular CDH problem can be also reduced to the Partial-CDH problem in $\mathbb{F}_{p^2}$.

Theorem 1 *Suppose $\mathcal{A}$ is a Partial-CDH adversary that runs in time at most φ and achieves advantage $\mathbf{Adv}^{\mathrm{pcdh}}_{\mathcal{A},\mathbb{F}_{p^2}}$. Then there exists a CDH adversary $\mathcal{B}$, constructed from $\mathcal{A}$ in a blackbox manner, that runs in time at most 2φ plus the time to perform a small constant number of group operations and achieves advantage $\mathbf{Adv}^{\mathrm{cdh}}_{\mathcal{B},\mathbb{F}_{p^2}} \geq \mathbf{Adv}^{\mathrm{pcdh}}_{\mathcal{A},\mathbb{F}_{p^2}} - 1/p$.*

Proof: The case where $p = 2$ is trivial. In the following, we assume that $p \neq 2$. Our CDH adversary $\mathcal{B}$ works as follows, given input a random instance of the CDH problem $(g^a, g^b) \in (\mathbb{F}_{p^2})^2$ and given a Partial-CDH adversary $\mathcal{A}$.

To begin with, adversary $\mathcal{B}$ transforms the input random instance (g^a, g^b) into two *random* yet *correlated* instances of the Partial-CDH problem by applying the random self-reducibility (*i.e.,* RSR) property of the CDH problem. Namely, $\mathcal{B}$ chooses two integers $r, s \overset{\$}{\leftarrow} \mathbb{Z}_{p^2-1}$, and computes (g^{a+r}, g^{b+s}) and $(g^{2(a+r)}, g^{b+s})$. For brevity, let $A = a + r$ and $B = b + s$.

Adversary $\mathcal{B}$ then runs the Partial-CDH adversary $\mathcal{A}$ on the generated instances (g^A, g^B) and (g^{2A}, g^B) under the representation determined by $h(x) = x^2 + h_1 x + h_0 \in I_2(p)$ to obtain $\left[g^{AB}\right]_1$ and $\left[g^{2AB}\right]_1$.

In the finite field $\mathbb{F}_{p^2}$, we observe $(g^{AB})^2 \bmod h(x) = g^{2AB}$. Hence,

$$(\left[g^{AB}\right]_1 x + \left[g^{AB}\right]_0)^2 \bmod h(x) = \left[g^{2AB}\right]_1 x + \left[g^{2AB}\right]_0. \qquad (1)$$

Comparing the coefficient of the degree-1 term in the equation (1), we have that $\left[g^{AB}\right]_0$ is a root of the following equation with y being the unknown:

$$2\left[g^{AB}\right]_1 y = \left[g^{AB}\right]_1^2 h_1 + \left[g^{2AB}\right]_1. \qquad (2)$$

If $\left[g^{AB}\right]_1 \neq 0$, there is a unique solution to the equation (2). That is, adversary $\mathcal{B}$ can get $\left[g^{AB}\right]_0 = (\left[g^{AB}\right]_1^2 h_1 + \left[g^{2AB}\right]_1) \cdot (2\left[g^{AB}\right]_1)^{-1}$, and therefore get $g^{AB} = \left[g^{AB}\right]_1 x + \left[g^{AB}\right]_0$. Otherwise, $\mathcal{B}$ simply aborts.

In fact, we can prove that the probability that $\mathcal{B}$ aborts ($\left[g^{AB}\right]_1 = 0$) is negligible. Note that $\left[g^{AB}\right]_1 = 0$ if and only if $g^{AB} \in \mathbb{F}_p$. It is not difficult to see that for any integer $n \xleftarrow{\$} \mathbb{Z}_{p^2-1}$, $g^n \in \mathbb{F}_p$ if and only if $p+1|n$. This implies that the probability of $g^n \in \mathbb{F}_p$ is $1/p$. Therefore, given a random AB, the probability that $\left[g^{AB}\right]_1 = 0$ is only $1/p$.

Now by $A = a + r$ and $B = b + s$, we have the following equation:

$$g^{AB} = g^{ab+as+br+rs}. \tag{3}$$

Since adversary $\mathcal{B}$ knows r and r, by equation (3) it can easily obtain $g^{ab} = g^{AB}(g^a)^{-s}(g^b)^{-r}g^{-rs}$. The theorem now follows by a standard argument. ▮

Remark 1. Theorem 1 provides a firm foundation for the main result in FGPS. Informally we have the following corollary: assuming the hardness of the *conventional* CDH problem over $\mathbb{F}_{p^2}$, given g^a, g^b and a random representation of $\mathbb{F}_{p^2}$, every single bit of $K = [g^{ab}]_1$ is hard-core. We omit the formal statement because we will in a moment be presenting an essentially stronger and generalized result: every single bit of the CDH value over $\mathbb{F}_{p^2}$ for the regular CDH problem is unpredictable.

Remark 2. This also answers the second question that FGPS raised: It is less likely to reduce the Partial-CDH over $\mathbb{F}_{p^2}$ to the regular CDH problem over $\mathbb{F}_p$. Otherwise, the CDH problems over $\mathbb{F}_{p^2}$ and $\mathbb{F}_p$ would be equivalent.

Remark 3. We can define a *dual* variant of the Partial-CDH problem over $\mathbb{F}_{p^2}$: We say that the *Dual-Partial-CDH* problem is hard in $\mathbb{F}_{p^2}$ if for all $h \in I_2(p)$ no efficient algorithm given $g, A = g^a, B = g^b \in \mathbb{F}_{p^2}$ can output $\left[g^{ab}\right]_0 \in \mathbb{F}_p$. We can show that the Dual-Partial-CDH problem is also as hard as the conventional CDH problem. The proof can be found in Appendix A. Therefore, both the Partial-CDH and Dual-Partial CDH problems are as hard as the conventional CDH problem over $\mathbb{F}_{p^2}$.

Remark 4. In Section 5, we will further develop the idea and use more complex techniques to prove a much more general result over finite fields $\mathbb{F}_{p^t}$ for any polynomial $t > 1$.

4 All Bits Security of the CDH Problems over $\mathbb{F}_{p^2}$

In this section, we show the following two results: (1) assuming the hardness of the Partial-CDH problem over $\mathbb{F}_{p^2}$, we prove the unpredictability of every single bit of the *other* coordinate of the secret CDH value; (2) we go on to prove the unpredictability of every single bit of the secret CDH value for the regular CDH problem over $\mathbb{F}_{p^2}$.

Bits Security for the other coordinate. Let $B_k \colon \mathbb{F}_p \to \{\pm 1\}$ denote the k-th bit predicate (with a 0 bit being encoded as $+1$). Let β_k be the bias of B_k. For all $h, \widehat{h} \in I_2(p)$ there exists an easily computable isomorphism $\phi_{h,\widehat{h}} \colon \mathbb{F}_p[x]/(h) \to \mathbb{F}_p[x]/(\widehat{h})$. Informally we show that when given an oracle $\mathcal{O}$ that predicts the k-th bit of the degree 0 coefficient of the CDH value with non-negligible advantage, and the representation of the field, then we can break the Partial-CDH assumption with non-negligible advantage.

Theorem 2 *Under the Partial-CDH assumption over $\mathbb{F}_{p^2}$ (i.e., Assumption 2), for any PPT adversary $\mathcal{O}$, we have that for all $h \in I_2(p)$ the following quantity is negligible in l:*

$$\left| \Pr\left[\mathcal{O}(h, \widehat{h}, g, g^a, g^b) = B_k\big(\left[\phi_{h,\widehat{h}}(g^{ab})\right]_0\big) \middle| \widehat{h} \xleftarrow{\$} I_2(p); a, b \xleftarrow{\$} \{1, \cdots, p^2-1\}\right] - \beta_k \right|.$$

We first give an informal intuition of the proof of the theorem. We aim at constructing a code similar to those of FGPS and Duc and Jetchev [7]. For an element $\alpha \in \mathbb{F}_{p^2}$ and a monic irreducible polynomial $h \in I_2(p)$, we would define the following codeword:

$$C_\alpha(\widehat{h}) = B_k([\phi_{h,\widehat{h}}(\alpha)]_0). \tag{4}$$

Similar to the code defined in FGPS, the code in (4) is accessible using $\mathcal{O}$. However, the predicate B_k is evaluated on the *other* coordinate of $\phi_{h,\widehat{h}}(\alpha)$. In this case, it holds that $[\phi_{h,\widehat{h}}(\alpha)]_0 = \eta[\alpha]_1 + [\alpha]_0$ for some $\eta \in \mathbb{F}_p$ (and $\lambda \in \mathbb{F}_p^*$), according to FGPS [9, Lemma 5.3] (recalled in Lemma 3 below).

Lemma 3 ([9, Lemma 5.3]). *For any $h \in I_2(p)$, there exists a unique function $L_h \colon \mathbb{F}_p \times \mathbb{F}_p^* \to I_2(p)$ which takes a pair (η, λ) to the polynomial $\widehat{h} = L_h(\eta, \lambda)$ such that the matrix $\left(\begin{smallmatrix} 1 & \eta \\ 0 & \lambda \end{smallmatrix}\right)$ defines an isomorphism from $\mathbb{F}_p[x]/(h)$ to $\mathbb{F}_p[x]/(\widehat{h})$ that sends $[\alpha]_1 x + [\alpha]_0 \mapsto \lambda[\alpha]_1 x + \eta[\alpha]_1 + [\alpha]_0$.*

Intuitively, one would consider the following code: for $\alpha \in \mathbb{F}_{p^2}$ and for $\eta \in \mathbb{F}_p$ (and $\lambda \in \mathbb{F}_p^*$), set

$$C_\alpha(\eta) = B_k(\eta[\alpha]_1 + [\alpha]_0). \tag{5}$$

Unfortunately, the above code in (5) is not *multiplicative*. In particular, this makes it hard to prove concentration and recoverability. This is why FGPS [9] considered defining the Partial-CDH problem over $\mathbb{F}_{p^2}$ as outputting the coefficient of the degree 1 term of g^{ab}, instead of the coefficient of the degree 0 term. More generally, the list decoding approach has only been proven successful for multiplicative codes so far [1, 7, 15]. One natural question is if it is (even) possible to apply list decoding approach to the case of non-multiplicative codes.

With a careful analysis, we are still able to show that the code in (5) is concentrated and recoverable. Concentration will follow from the key observation that the Fourier transform of the code in (5) is equal to that of a multiplication code (to be defined shortly) up to a factor of a character (as will be proved in Lemma 4). Hence, the l_2-norm of the Fourier transform of the code is equal to that of the multiplication code. That is, the code in (5) is concentrated if and only if the multiplication code is. Note that it is easy to argue that the multiplication code is concentrated.

The goal of recoverability is to recover the secret value from the heavy characters of the code C_α. We find that a character χ_β is heavy for C_α if and only if χ_β is heavy for a multiplicative code C'_α. The associated constant of a heavy character χ_β for the multiplicative code C'_α equals the product of the secret value and an (easily determined) factor. Therefore, one can recover the secret value with a heavy character. We begin by proving Lemma 4.

Lemma 4 *Let F_1, F_2 be functions mapping $\mathbb{Z}_n$ to $\mathbb{C}$. If for any y, $F_2(y) = F_1(y - \sigma)$, where σ is a constant in $\mathbb{Z}_n$, then we have for $\alpha \in \mathbb{Z}_n$, $\widehat{F_2}(\alpha) = \chi_\alpha(\sigma)\widehat{F_1}(\alpha)$.*

Proof of Lemma 4: By the definition of Fourier transform and $F_2(y) = F_1(y - \sigma)$, we have

$$\widehat{F_2}(\alpha) = 1/n \sum_{y \in \mathbb{Z}_n} \overline{F_1(y - \sigma)} \chi_\alpha(y). \tag{6}$$

It is easily seen that if y traverses the complete residue system modulo n then so does $y - \sigma$. Hence, we have $\{y - \sigma\}_{y \in \mathbb{Z}_n} = \mathbb{Z}_n$. Equation (6) can be re-written as

$$\widehat{F_2}(\alpha) = 1/n \sum_{y \in \mathbb{Z}_n} \overline{F_1(y)} \chi_\alpha(y + \sigma). \tag{7}$$

Since $\chi_\alpha(y+\sigma) = \chi_\alpha(\sigma)\chi_\alpha(y)$, equation (7) becomes

$$\widehat{F_2}(\alpha) = 1/n \sum_{y \in \mathbb{Z}_n} \overline{F_1(y)}\chi_\alpha(y)\chi_\alpha(\sigma).$$

It thus follows that $\widehat{F_2}(\alpha) = \chi_\alpha(\sigma)\widehat{F_1}(\alpha)$. This completes the proof the lemma. □

We are now ready to prove Theorem 2.

Proof of Theorem 2: Suppose that there exists an oracle $\mathcal{O}$ such that

$$\big|\Pr_{\eta,a,b}\big[\mathcal{O}(h,\widehat{h},g,g^a,g^b) = B_k\big(\big[\phi_{h,\widehat{h}}(g^{ab})\big]_0\big)\big] - \beta_k\big|$$

is larger than a non-negligible quantity ϵ. We construct another oracle $\mathcal{O}'$ that takes as input a base representation $h \in I_2(p)$, a Diffie-Hellman triple $g, g^a, g^b \in \mathbb{F}_{p^2}$, and an element of $\eta \in \mathbb{F}_p$ (instead of $\widehat{h} \in I_2(p)$). The new oracle selects $\lambda \xleftarrow{\$} \mathbb{F}_p^*$, constructs an isomorphism $\widehat{h}$ from the matrix $\left(\begin{smallmatrix} 1 & \eta \\ 0 & \lambda \end{smallmatrix}\right)$ as described in Lemma 3, and returns $\mathcal{O}(h,\widehat{h},g,g^a,g^b)$. One can then show that

$$\big|\Pr_{\eta,a,b}\big[\mathcal{O}'(h,\eta,g,g^a,g^b) = B_k\big(\eta\big[g^{ab}\big]_1 + \big[g^{ab}\big]_0\big)\big] - \beta_k\big|$$

is also larger than a non-negligible quantity.

For any element $\alpha \in \mathbb{F}_{p^2}$, we construct the following encoding of $\eta[\alpha]_1 + [\alpha]_0$ in its polynomial representation for $\mathbb{F}_p[x]/(h)$:

$$C_\alpha \colon \mathbb{F}_p \to \{\pm 1\} \text{ such that } C_\alpha(\eta) = B_k(\eta[\alpha]_1 + [\alpha]_0),$$

where, above, $[\alpha]_1$ and $[\alpha]_0$ are under the representation determined by h.

Accessibility. Accessibility proof is the same as that of FGPS. In particular, the oracle $\mathcal{O}'$ allows us to have access to a corrupted codeword $\widetilde{C}_\alpha$ of the above codeword defined as $\widetilde{C}_\alpha = \mathcal{O}'(h,\eta,g,g^a,g^b)$. The code $C_\alpha(\eta)$ is conceptually the same as the code $C_\alpha(\widehat{h})$. Therefore, if the oracle $\mathcal{O}$ has advantage ϵ then we have $|\Pr_\eta[C_\alpha(\eta) = \widetilde{C}_\alpha(\eta)]| \geq \beta_k + \epsilon$. Accessibility of the code C_α follows.

Concentration. We now prove that the codeword C_α is a Fourier concentrated code. To prove that $C_\alpha(\eta)$ is a concentrated code, we define the following related code:

$$C'_\alpha(\eta) = B_k(\eta[\alpha]_1).$$

It is easy to see that $C'_\alpha(\eta) = C_\alpha(\eta - [\alpha]_1^{-1}[\alpha]_0)$. According to Lemma 4, we can obtain

$$\chi_\beta([\alpha]_1^{-1}[\alpha]_0)\widehat{C_\alpha}(\beta) = \widehat{C'_\alpha}(\beta).$$

This immediately implies $|\widehat{C_\alpha}(\beta)| = |\widehat{C'_\alpha}(\beta)|$. Therefore, the code $C_\alpha(\eta)$ is concentrated if and only if the code $C'_\alpha(\eta)$ is. Note that it is easy to argue that $C'_\alpha(\eta)$ is a multiplication code. The proof for concentration of the code $C'_\alpha(\eta)$ is similar to those of [9, 15], and it can be found in Appendix B.

Character χ_β is τ-heavy for C_α if and only if χ_β is τ-heavy for C'_α. Therefore, according to the discussion in FGPS, for a threshold $\tau > 0$, the τ-heavy characters of C_α belong to the set

$$\Gamma_{\alpha,\tau} = \{\chi_\beta \colon \beta = \eta[\alpha]_1 \text{ for } \eta \in \Gamma_\tau\},$$

where Γ_τ is a set containing the τ-heavy coefficients of the function B_k. For each $\eta \in \Gamma_\tau$, there exists a unique integer pair $(\xi_\eta, \varsigma_\eta) \in [0, 1/\tau] \times [0, 1/\tau]$. Note that by [15, Lemma 9], the size of Γ_τ is at most $4\tau^{-2}$.

Recoverability. The proof for recoverability is similar to those of [9, 15]. According to Lemma 1, we know that there exists a threshold τ which is polynomial in the non-negligible quantity ϵ and at least one τ-heavy Fourier character $\chi \neq 0$ for C_α and $\widetilde{C}_\alpha$ such that $\chi \in \mathsf{Heavy}_\tau(C_\alpha) \cap \mathsf{Heavy}_\tau(\widetilde{C}_\alpha)$.

Fixing a polynomial $h(x) \in I_2(p)$, on input $g, g^a, g^b \in \mathbb{F}_{p^2}$, the following algorithm that has access to $\mathcal{O}$ produces a polynomial size list of elements in $\mathbb{F}_{p^2}$ which contains g^{ab} with probability $1 - \delta$.

Let τ be the threshold determined by Lemma 1. We write $\alpha = [\alpha]_1 x + [\alpha]_0$ to denote $g^{ab} \in \mathbb{F}_{p^2}$. Using the learning algorithm of AGS [1] (*i.e.,* the algorithm in Lemma 2), we obtain a polynomial size list L_α of all the τ-heavy Fourier characters for $\widetilde{C}_\alpha$. If χ_β is a non-trivial τ-heavy character for C_α, we have $[\alpha]_1 = \eta^{-1}\beta$. Given $\chi_\beta \in L_\alpha$, we define $L_\beta = \{[\alpha]_1 \colon [\alpha]_1 = \eta^{-1}\beta \text{ for } \eta \in \Gamma_\tau\}$.

Let $L = \bigcup_{\chi_\beta \in L_\alpha} L_\beta$. Note that L is of polynomial size and $\alpha \in L$ with probability $1 - \delta$. Since this is a polynomial size set, we can guess a result for $[\alpha]_1$ and hence get $[g^{ab}]_1$. The theorem now follows. ▮

Hard-core predicates for the CDH problem over $\mathbb{F}_{p^2}$. Let l be the binary length of a prime p. Note that for a fixed $h \in I_2(p)$, any element $\alpha \in \mathbb{F}_{p^2}$ of length $2l$ can be written as $[\alpha]_1 x + [\alpha]_0$. We assume without loss of generality that $[\alpha]_1$ and $[\alpha]_0$ are the leftmost and rightmost l bits value of α, respectively. Let $\widetilde{B}_k \colon \mathbb{F}_{p^2} \to \{\pm 1\}$ denote the k-th bit predicate (where $1 \leq k \leq 2l$) and let β_k be the bias of $\widetilde{B}_k$. In the following, we prove that given an oracle $\mathcal{O}$ that predicts the k-th bit of the CDH value over a random representation of the field $\mathbb{F}_{p^2}$ with non-negligible advantage, we can solve the *regular* CDH problem over $\mathbb{F}_{p^2}$ with non-negligible probability.

Theorem 3 *Under the CDH assumption over $\mathbb{F}_{p^2}$ (i.e., Assumption 1), for any PPT adversary $\mathcal{O}$, we have that for all $h \in I_2(p)$ the following quantity is negligible in l:*

$$\left| \Pr\left[\mathcal{O}(h, \widehat{h}, g, g^a, g^b) = \widetilde{B}_k\big(\phi_{h,\widehat{h}}(g^{ab})\big) \,\middle|\, \widehat{h} \xleftarrow{\$} I_2(p); a, b \xleftarrow{\$} \{1, \cdots, p^2 - 1\}\right] - \beta_k \right|.$$

Proof Sketch: For an element $\alpha \in \mathbb{F}_{p^2}$ and a fixed $h \in I_2(p)$, we define a codeword as follows:

$$C_\alpha(\widehat{h}) = \widetilde{B}_k(\phi_{h,\widehat{h}}(\alpha)).$$

If $k \leq l$, we have

$$\widetilde{B}_k(\phi_{h,\widehat{h}}(\alpha)) = B_k([\phi_{h,\widehat{h}}(\alpha)]_0).$$

Otherwise if $k > l$, we have

$$\widetilde{B}_k(\phi_{h,\widehat{h}}(\alpha)) = B_{k-l}([\phi_{h,\widehat{h}}(\alpha)]_1).$$

Along the same lines as the proofs of [9, Theorem 5.2] and Theorem 3, predicting any individual bit of the secret CDH value defined above can break the Partial-CDH assumption over $\mathbb{F}_{p^2}$, and hence break the CDH assumption over $\mathbb{F}_{p^2}$, as shown in Theorem 1. ▮

The above result strengthens and generalizes the FGPS result: first, the underlying assumption is the conventional CDH problem instead of the Partial-CDH problem; second, it proves that every individual bit of the secret CDH value, not just every bit of one of the coordinates, is hard-core.

5 The d-th CDH Assumption over $\mathbb{F}_{p^t}$ and its Equivalence to the CDH Assumption

As mentioned in FGPS, the Partial-CDH problems can be defined for general finite fields $\mathbb{F}_{p^t}$ with any $t > 1$ as outputting the coefficient of the term of degree $t-1$ (*i.e.,* the maximum degree). In this section, we define a much more generalized set of problems over $\mathbb{F}_{p^t}$—the d-th CDH problems, as outputting the coefficient of the degree d term for every $0 \leq d \leq t-1$.

For a given prime p, there are many different fields $\mathbb{F}_{p^t}$, but they are all isomorphic to each other. Let $h(x) = x^t + h_{t-1}x^{t-1} + \cdots + h_1 x + h_0$ be a monic irreducible polynomial of degree t in $\mathbb{F}_p$. It is well known that $\mathbb{F}_{p^t}$ is isomorphic to the field $\mathbb{F}_p[x]/(h)$, where $(h(x))$ is a principal ideal in the polynomial ring $\mathbb{F}_p[x]$ and therefore elements of $\mathbb{F}_{p^t}$ can be written as polynomials of degree $t-1$, *i.e.,* if $g \in \mathbb{F}_{p^t}$ then $g = g_{t-1}x^{t-1} + g_{t-2}x^{t-2} + \cdots + g_1 x + g_0$ and addition and multiplication are performed as polynomial operations modulo h. In the following, given $g \in \mathbb{F}_{p^t}$ we denote by $[g]_i$ the coefficient of the degree-i term, *i.e.,* $g_i = [g]_i$. Let $I_t(p)$ be the set of monic irreducible polynomials of degree t in $\mathbb{F}_p$, and let g be a generator of the multiplicative group of $\mathbb{F}_{p^t}$.

First, the CDH problem can be easily extended to the case of finite fields $\mathbb{F}_{p^t}$ for any $t > 1$.

Assumption 3 (The CDH assumption over $\mathbb{F}_{p^t}$). *We say that the CDH problem is hard in $\mathbb{F}_{p^t}$ for $t > 1$ if for any PPT adversary $\mathcal{A}$, his CDH advantage*

$$\mathbf{Adv}^{\mathrm{cdh}}_{\mathcal{A},\mathbb{F}_{p^t}} = \Pr\left[\mathcal{A}(p, g, h, g^a, g^b) = g^{ab} \middle| a, b \xleftarrow{\$} \{1, \cdots, p^t - 1\}\right]$$

is negligible in l.

We consider the following relaxed variations of the CDH problems over $\mathbb{F}_{p^t}$ for $t > 1$: We say that the *d-th CDH* problem (where $0 \leq d \leq t-1$) is hard in $\mathbb{F}_{p^t}$ if for all $h \in I_t(p)$ no efficient algorithm given $g, A = g^a, B = g^b \in \mathbb{F}_{p^t}$ can output $\left[g^{ab}\right]_d \in \mathbb{F}_p$. Formally we consider the following assumption:

Assumption 4 (The d-th CDH assumption over $\mathbb{F}_{p^t}$). *We say that the d-th CDH problem (where $0 \leq d \leq t-1$) is hard in $\mathbb{F}_{p^t}$ (for $t > 1$) if for any PPT adversary $\mathcal{A}$, his d-th CDH advantage for all $h \in I_t(p)$*

$$\mathbf{Adv}^{\mathrm{dcdh}}_{\mathcal{A},\mathbb{F}_{p^t}} = \Pr\left[\mathcal{A}(p, g, h, g^a, g^b) = \left[g^{ab}\right]_d \middle| a, b \xleftarrow{\$} \{1, \cdots, p^t - 1\}\right]$$

is negligible in l.

We prove that the regular CDH problem over $\mathbb{F}_{p^t}$ where $t > 1$ and $t \in poly(l)$ can be reduced to *any* d-th CDH problem ($0 \leq d \leq t-1$) over $\mathbb{F}_{p^t}$. Therefore, all the d-th CDH problems over finite fields $\mathbb{F}_{p^t}$ for any polynomial $t > 1$ are as hard as the regular CDH problem over the same fields.

Theorem 4 *Fix an l-bit prime p and an integer $t > 1$ such that $t \in poly(l)$. Suppose $\mathcal{A}$ is a d-th CDH adversary that runs in time at most φ and achieves advantage $\mathbf{Adv}^{\mathrm{dcdh}}_{\mathcal{A},\mathbb{F}_{p^t}}$. Then there exists a CDH adversary $\mathcal{B}$, constructed from $\mathcal{A}$ in a blackbox manner, that runs in time at most $t\varphi$ plus the time to perform $poly(l)$ group operations and achieves advantage $\mathbf{Adv}^{\mathrm{cdh}}_{\mathcal{B},\mathbb{F}_{p^t}} \geq 1/(2t) \cdot (1 - 1/p)^2 \cdot e^{-\frac{2}{p-1}} \cdot (\mathbf{Adv}^{\mathrm{dcdh}}_{\mathcal{A},\mathbb{F}_{p^t}})^t$.*

Proof intuition: The goal is to construct a CDH adversary which on input $g^a, g^b \in \mathbb{F}_{p^t}$ returns g^{ab}. Slightly generalizing the prior method in proving Theorem 1, adversary chooses integers $r, s \xleftarrow{\$} \mathbb{Z}_{p^t - 1}$

and computes (g^{a+r}, g^{b+s}). Let $A = a + r$ and $B = b + s$. Adversary then runs the d-th CDH problem oracle on the instances (g^A, g^B) to get the d-th coordinate of the CDH value $[g^{AB}]_d$. It is easy to see that $g^{AB} = g^{ab}g^{as+br+rs}$, *i.e.,*

$$\sum_{k=0}^{t-1}[g^{AB}]_k x^{\equiv}\left(\sum_{i=0}^{t-1}[g^{ab}]_i x^i\right)\left(\sum_{j=0}^{t-1}[g^{as+br+rs}]_j x^j\right) \bmod h(x).$$

The adversary can easily compute g^{as}, g^{br}, and g^{rs}, and therefore $[g^{AB}]_d$ can be written as a linear expression in the coordinates of g^{ab} with known coefficients $e_i \in \mathbb{F}_p$ $(0 \le i \le t-1)$ such that $[g^{AB}]_d = \sum_{i=0}^{t-1} e_i [g^{ab}]_i$.

We can repeat the above process to gather "enough" equations and use Gaussian elimination to compute the unknowns (*i.e.,* $[g^{ab}]_i$ for $0 \le i \le t-1$) and thus obtain g^{ab}. Typically, one would just need to repeat the process for t times to collect t equations, if one could prove that the corresponding coefficient matrix for the equation set has full rank with non-negligible probability. To achieve this goal, we will run the d-th CDH oracle under a "random" field representation by choosing a random irreducible polynomial. However, as we will see, making the above intuition into a rigorous argument takes some work.

Proof of Theorem 4: On input a random instance of the CDH problem $(g^a, g^b) \in (\mathbb{F}_{p^t})^2$, the goal is to construct a CDH adversary $\mathcal{B}$ that runs a d-th CDH adversary $\mathcal{A}$ and returns g^{ab}.

Preparation. Before proceeding to the proof, we introduce two useful lemmas. The first lemma (see [14, Ex. 3.26 and 3.27, p. 142]) ensures that the probability of a random polynomial of degree t being irreducible is high.

Lemma 5 *Let p be a prime and $t \in \mathbb{Z}_+$. Then the probability that a randomly selected polynomial of degree t over $\mathbb{F}_p$ is at least $1/(2t)$.*

The second lemma claims if all the entries in a square matrix are independently and uniformly chosen at random over a large finite field $\mathbb{F}_p$ then there is a good chance that the matrix is nonsingular. Note that we require that the probability depend only on the size of the finite field p, but not on the size of the matrix m. Similar results have been studied in, *e.g.,* [17].[1] For self-containedness, we include a simpler proof for Lemma 6 (with a better bound) in Appendix C.

Lemma 6 *Let M be an $m \times m$ square matrix over the finite field $\mathbb{F}_p$. If every element of the matrix is chosen independently and uniformly at random, then the probability that M is nonsingular is at least $e^{-\frac{2}{p-1}}$.*

Choosing a random base representation. To begin with, adversary $\mathcal{B}$ needs to run the d-th CDH problem adversary $\mathcal{A}$ over a random base representation of the field by choosing a random irreducible polynomial $h(x)$. This will allow us to use the distribution property of the coefficients of $h(x)$. To select such a random irreducible polynomial, adversary $\mathcal{B}$ would just need to choose a random monic polynomial with all its coefficients being uniformly and independently selected at random from the elements of $\mathbb{F}_p$. According to Lemma 5, the polynomial will be irreducible with $1/(2t)$ probability.[2]

Core algorithm. Adversary $\mathcal{B}$ chooses integers $r, s \xleftarrow{\$} \mathbb{Z}_{p^t-1}$ and computes (g^{a+r}, g^{b+s}). For clarity, we let $A = a + r$ and $B = b + s$. Adversary $\mathcal{B}$ then runs the d-th CDH adversary $\mathcal{A}$ under the

[1] There are even more references which study the probability problems with respect to both p and m. The problems, although important, are not our concern here.

[2] In order to check whether a given polynomial is irreducible, one can use, *e.g.*, Ben-Or's irreducibility test [3].

representation determined by $h(x)$ on the instances (g^A, g^B) to get the d-th coordinate of the CDH value $[g^{AB}]_d$ under the representation $h(x)$.

Expanding the equation $g^{AB} = g^{ab+as+br+rs}$, we obtain

$$\sum_{v=0}^{t-1} [g^{AB}]_v x^v \equiv \left(\sum_{i=0}^{t-1} [g^{ab}]_i x^i \right) \left(\sum_{j=0}^{t-1} [g^{as+br+rs}]_j x^j \right) \bmod h(x). \tag{8}$$

As adversary $\mathcal{B}$ can easily compute g^{as}, g^{br}, and g^{rs}, it can then compute each coordinate $[g^{as}g^{br}g^{rs}]_i$ $(0 \leq i \leq t-1)$. Therefore, $[g^{AB}]_d$ can be represented as a linear equation with known coefficients $e_i \in \mathbb{F}_p$, $0 \leq i \leq t-1$, such that

$$[g^{AB}]_d = \sum_{i=0}^{t-1} e_i [g^{ab}]_i. \tag{9}$$

Repeating the process for t times, $\mathcal{B}$ obtain t equations with t unknowns (*i.e.,* $[g^{ab}]_i$ for $0 \leq i \leq t-1$). Note that the probability that all t invocations of adversary $\mathcal{A}$ return correct answers is at least $(\mathbf{Adv}^{\mathrm{dcdh}}_{\mathcal{A},\mathbb{F}_{p^t}})^t$. If the coefficient matrix of the equation set has full rank then $\mathcal{B}$ can easily compute g^{ab} in polynomial time of l and t (note that $t \in poly(l)$).

Next we will prove that the coefficient matrix for the above equation set *indeed* has full rank with non-negligible probability, and the theorem will then follow.

Nonsingularity of the coefficient matrix. We first need to provide an explicit expression of e_i for equation (9). Let $h(x) = x^t + h_{t-1}x^{t-1} + \cdots + h_1 x + h_0$. For brevity, let $x_i = [g^{ab}]_i$ and $y_i = [g^{as+br+rs}]_i$ for $0 \leq i \leq t-1$. Then

$$g^{ab} = \sum_{i=0}^{t-1} x_i x^i \quad \text{and} \quad g^{as+br+rs} = \sum_{j=0}^{t-1} y_j x^j.$$

We observe two simple facts: first, if one can compute all variables x_i $(0 \leq i \leq t-1)$ then one can recover g^{ab}; second, each variable y_i $(0 \leq i \leq t-1)$ is uniformly and independently chosen at random. Suppose that

$$(\sum_{i=0}^{t-1} x_i x^i)(\sum_{j=0}^{t-1} y_j x^j) = \sum_{k=0}^{2t-2} \alpha_k x^k.$$

It is easy to see that

$$\alpha_k = \sum_{\substack{i+j=k \\ 0 \leq i,j \leq t-1}} x_i y_j$$

Equation (8) can be written as

$$\sum_{k=0}^{2t-2} \alpha_k x^k \equiv \sum_{v=0}^{t-1} [g^{AB}]_v x^v \bmod h(x). \tag{10}$$

By a rather complex calculation, we obtain from equation (10) for $0 \leq v \leq t-1$,

$$[g^{AB}]_v = \alpha_v + \sum_{i=0}^{v} h_i \beta_{t-1-v+i},$$

where we have

$$\beta_1 = \alpha_{2t-2},$$

and for $2 \leq k \leq t-1$

$$\beta_k = \alpha_{2t-k-1} - \sum_{i=1}^{k-1} h_{t-k+i}\beta_i.$$

In particular, we are interested in the equation for d-th coordinate

$$[g^{AB}]_d = \alpha_d + \sum_{i=0}^{d} h_i \beta_{t-1-d+i}.$$

From the definition of β_i for $1 \leq i \leq t-1$, we have

$$[g^{AB}]_d = \alpha_d + \xi_t\alpha_t + \xi_{t+1}\alpha_{t+1} + \xi_{t+2}\alpha_{t+2} + \cdots \xi_{2t-2}\alpha_{2t-2}, \tag{11}$$

where each ξ_z ($t \leq z \leq 2t-2$) is a *non-trivial* polynomial[3] of $h_0, h_1, \cdots$, and h_{t-1} (and has the form of $\sum h_0^{i_0} h_1^{i_1} \cdots h_{t-2}^{i_{t-2}} h_{t-1}^{i_{t-1}}$ satisfying $i_0 + i_1 + \cdots + i_{t-1} \leq t-1$), and $\deg(\xi_z) = z - t + 1$. It is important to note that $\xi_t = h_d$.

Expanding $\alpha_k = \sum_{\substack{i+j=k \\ 0 \leq i,j \leq t-1}} x_i y_j$, equation (11) can be written as

$$[g^{AB}]_d = y_d x_0 + \sum_{i=1}^{d} \big(y_{d-i} + \sum_{j=1}^{i} \xi_{t+j-1} y_{t-i+j-1}\big) x_i + \sum_{i=d+1}^{t-1} \big(\sum_{j=1}^{i} \xi_{t+j-1} y_{t-i+j-1}\big) x_i. \tag{12}$$

As discussed, adversary $\mathcal{B}$ can repeat the above process for t times by randomly and independently choosing t pairs of integers $(r^{(u)}, s^{(u)}) \overset{\$}{\leftarrow} (\mathbb{Z}_{p^t-1})^2$ such that $1 \leq u \leq t$. Let $A^{(u)} = a + r^{(u)}$ and $B^{(u)} = b + s^{(u)}$. Adversary $\mathcal{B}$ then runs the d-th CDH adversary $\mathcal{A}$ on each instance $(g^{A^{(u)}}, g^{B^{(u)}})$ under the representation $h(x)$ to get the d-th coordinate of the CDH value $[g^{A^{(u)}B^{(u)}}]_d$. From equation (12) we have

$$[g^{A^{(u)}B^{(u)}}]_d = y_d^{(u)} x_0 + \sum_{i=1}^{d} \big(y_{d-i}^{(u)} + \sum_{j=1}^{i} \xi_{t+j-1} y_{t-i+j-1}^{(u)}\big) x_i + \sum_{i=d+1}^{t-1} \big(\sum_{j=0}^{i} \xi_{t+j} y_{t-i+j}^{(u)}\big) x_i. \tag{13}$$

Adversary $\mathcal{B}$ now get a linear equation set with $x_0, x_1, \cdots$, and x_{t-1} being the unknowns. If the coefficient matrix for the equation set has full rank then $\mathcal{B}$ can easily compute its unique solution set in polynomial time. In what follows, we will show that the coefficient matrix is indeed nonsingular with high probability. Specifically, we will show that each entry of the above matrix is independently and uniformly chosen at random and then conclude according to Lemma 6.

First, we prove that with high probability all the entries of the coefficient matrix are uniformly distributed at random. As each row has the same representation according to equation (13), we can consider without loss of generality the elements of the first row. The first element $y_d^{(1)}$ is clearly uniformly distributed at random. Each of the remaining $t-1$ elements at least contains one vector of the form $\xi_t y_i^{(1)}$ (such that $1 \leq i \leq t-1$). Recall that $\xi_t = h_d$ and therefore the probability of ξ_t being non-trivial is at least $1 - 1/p$. Hence, the probability that the remaining $t-1$ elements are also uniformly distributed at random is at least $1 - 1/p$.

Second, we observe that since independently random numbers are used for different rows, any two entries from two different rows are linearly independent. We go on to prove that with high probability

[3] As we will see, the non-triviality of the polynomial is essential to our proof.

all entries from the same row are *also* linearly independent. Again we just need to consider with loss of generality the first row of the coefficient matrix.

Note that each entry of the first row of the coefficient matrix is a linear combination of $y_0^{(1)}, \cdots, y_{t-1}^{(1)}$. Regarding each such entry as a column vector, we can obtain for the first row a $t \times t$ square matrix, whose antidiagonal (*i.e.,* secondary diagonal) is formed by all zero elements. We would just need to prove the nonsingularity of this new matrix.

$$\begin{pmatrix}
0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & \xi_t \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \xi_t & \xi_{t+1} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
0 & 0 & 1 & \cdots & 0 & 0 & \xi_t & \cdots & \xi_{t+d-5} & \xi_{t+d-4} & \xi_{t+d-3} \\
0 & 1 & 0 & \cdots & 0 & \xi_t & \xi_{t+1} & \cdots & \xi_{t+d-4} & \xi_{t+d-3} & \xi_{t+d-2} \\
1 & 0 & 0 & \cdots & \xi_t & \xi_{t+1} & \xi_{t+2} & \cdots & \xi_{t+d-3} & \xi_{t+d-2} & \xi_{t+d-1} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \xi_{t+d-5} & \xi_{t+d-4} & \xi_{t+d-3} & \cdots & \xi_{2t-6} & \xi_{2t-5} & \xi_{2t-4} \\
0 & 0 & \xi_t & \cdots & \xi_{t+d-4} & \xi_{t+d-3} & \xi_{t+d-2} & \cdots & \xi_{2t-5} & \xi_{2t-4} & \xi_{2t-3} \\
0 & \xi_t & \xi_{t+1} & \cdots & \xi_{t+d-3} & \xi_{t+d-2} & \xi_{t+d-1} & \cdots & \xi_{2t-4} & \xi_{2t-3} & \xi_{2t-2}
\end{pmatrix}$$

Denote by γ_i the i-th row vector for $1 \leq i \leq t$. Suppose there are elements $a_1, a_2, \cdots, a_t \in \mathbb{F}_p$ such that

$$a_1\gamma_1 + a_2\gamma_2 + \cdots + a_t\gamma_t = 0, \tag{14}$$

where $a_1, a_2, \cdots, a_t$ are not all zero. The equation (14) implies the following equation set

$$\begin{aligned}
a_d &= 0, \\
\sum_{i=0}^{k} a_{t-i}\xi_{t+k-i} + a_{d-k-1} &= 0, \qquad 1 \leq k \leq d-1, \\
\sum_{i=d}^{k} a_{t-i}\xi_{t+k-i-1} &= 0, \qquad d \leq k \leq t-1.
\end{aligned}$$

Observe that $a_2, \cdots, a_t$ are not all zero, as otherwise a_1 must also be zero by the d-th equation above, contradicting that they are not all zero. Assume that a_κ ($2 \leq \kappa \leq t$) is the last non-zero one. Let's consider only the κ-th equation of the equation set: if $2 \leq \kappa \leq d$, the κ-th equation is $a_\kappa\xi_t + a_{d-\kappa+1} = 0$, otherwise $a_\kappa\xi_t = 0$. Recall that $\xi_t = h_d$ and therefore the κ-th equation has at most one solution. As $h_0, h_1, \cdots, h_{t-1}$ are chosen independently and uniformly at random, the probability that the κ-th equation has solution is at most $1/p$. Hence, the probability that $\gamma_1, \cdots, \gamma_t$ are linearly independent over $\mathbb{F}_p$ is at least $1 - 1/p$.

We can now conclude that with probability at least $(1 - 1/p)^2$, the coefficient matrix from equation (13) is indeed a random $t \times t$ matrix over $\mathbb{F}_p$, where each entry is chosen independently and uniformly at random from the elements of $\mathbb{F}_p$. According to Lemma 6, if a matrix satisfies the above property then the probability of the matrix being nonsingular is at least $e^{-\frac{2}{p-1}}$.

Putting the pieces together and end of proof. Our reduction works on a random polynomial of degree t over finite field $\mathbb{F}_p$. The probability of a random polynomial being irreducible is at least

$1/(2t)$. With a random irreducible polynomial, we have shown that if the coefficient matrix for the equation set (13) has full rank then we are able to solve the equation set in polynomial time (*e.g.,* using Gaussian elimination) and therefore recover the secret CDH value over $\mathbb{F}_{p^t}$. We have also proven that the probability of the matrix having full rank is at least $(1-1/p)^2 \cdot e^{-\frac{2}{p-1}}$.

Therefore, selecting a random polynomial of degree t and invoking the core algorithm (*i.e.,* running adversary $\mathcal{A}$ for t times and solving the equation set obtained), adversary $\mathcal{B}$ can compute the desired CDH value, that runs in time at most $t\varphi$ plus the time to perform *poly*(l) group operations with a non-negligible advantage $1/(2t) \cdot (1-1/p)^2 \cdot e^{-\frac{2}{p-1}} \cdot (\mathbf{Adv}^{\mathrm{dcdh}}_{\mathcal{A},\mathbb{F}_{p^t}})^t$. The theorem now follows. ∎

6 Almost All Bits Security of the CDH Problems over $\mathbb{F}_{p^t}$ for $t > 1$

In this section, we show the following result: assuming the hardness of the d-th CDH problem over $\mathbb{F}_{p^t}$ with polynomial $t > 1$, if $d \neq 0$, we prove the unpredictability of every single bit of the degree-d coordinate of the secret CDH value. Together with the equivalence result in Section 5, this implies that for the conventional CDH problems over $\mathbb{F}_{p^t}$ for an l-bit prime p and an integer $t > 1$, $(t-1)l$ out of tl secret CDH bits—including every individual bit except that of the degree 0 coordinate—are hard-core.

We begin with the definition of d-th residue modulo p. Let p be a prime and d be an integer. We say that an element $\alpha \in \mathbb{F}_p^*$ is a d-th residue modulo p, if the congruence equation $x^d \equiv \alpha \bmod p$ has solution in $\mathbb{F}_p$. We let $\mathbb{F}_p^d$ denote the set of the d-th residue modulo p. The following lemma provides a well-known result on d-th residue modulo p:

Lemma 7 *Let p be a prime and $d \in \mathbb{Z}_+$. There exists d-th residue modulo p and the number of d-th residue modulo p is $(p-1)/(d, p-1)$.*

We present a lemma that gives a characterization of the isomorphisms between two representations of the fields $\mathbb{F}_{p^t}$. The isomorphisms generalize that of finite fields $\mathbb{F}_{p^2}$ in FGPS to the case of general finite fields $\mathbb{F}_{p^t}$ for any $t > 1$. More importantly, they simplify that of FGPS in the sense we identify a more restrictive class of isomorphisms. This simplicity turns out to be essential to establishing the bits security for general finite fields.

Lemma 8 *For any $h(x) \in I_t(p)$, there exists a unique function $L_h \colon \mathbb{F}_p^* \to I_t(p)$ which takes λ to the polynomial $\widehat{h}_\lambda = L_h(\lambda) = \frac{h(\lambda x)}{\lambda^t}$ such that λ defines an isomorphism from $\mathbb{F}_p[x]/(h)$ to $\mathbb{F}_p[x]/(\widehat{h})$ that sends*

$$\sum_{i=0}^{t} [\alpha]_i x^i \mapsto \sum_{i=0}^{t} \lambda^i [\alpha]_i x^i.$$

Proof. For any $\lambda \in \mathbb{F}_p^*$, let $\widehat{h}_\lambda(x) = \frac{h(\lambda x)}{\lambda^t}$. It is easy to see that $\widehat{h}_\lambda(x)$ is a monic irreducible polynomial over $\mathbb{F}_p$, *i.e.,* $\widehat{h}_\lambda(x) \in I_t(p)$. Hence, there is an isomorphism from $\mathbb{F}_p[x]/(h)$ to $\mathbb{F}_p[x]/(\widehat{h}_\lambda)$. In order to specify a homomorphism ψ from $\mathbb{F}_p[x]/(h)$ to another field J of characteristic p, it is both necessary and sufficient to choose $\psi(x) = y \in J$ such that $h(y) = 0$ in J. The definition of $\widehat{h}_\lambda$ implies that x sends to λx. The lemma now follows.

Theorem 5 *Under the d-th CDH assumption over $\mathbb{F}_{p^t}$ for any polynomial $t > 1$ (i.e., Assumption 4), for any PPT adversary $\mathcal{O}$, if $d \neq 0$, we have that for all $h \in I_t(p)$ the following quantity is negligible* *in l:*

$$\big|\Pr\big[\mathcal{O}(h,\lambda,g,g^a,g^b) = B_k\big(\big[\phi_{h,\widehat{h}_\lambda}(g^{ab})\big]_d\big)\big|\lambda \xleftarrow{\$} \mathbb{F}_p^*; a,b \xleftarrow{\$} \{1,\cdots,p^2-1\}\big] - \beta_k\big|.$$

Proof: For an element $\alpha \in \mathbb{F}_{p^t}$ and a monic irreducible polynomial $h \in I_t(p)$, $\lambda \xleftarrow{\$} \mathbb{F}_p^*$, the prediction oracle $\mathcal{O}$ gives noisy access to the codeword $B_k(\lambda^d[\alpha]_d)$. Note that when $d \neq 1$ the above code is not *multiplicative*. Again, this would make it hard to prove concentration and recoverability. In order to apply the techniques of [1], we would need noisy access to the multiplication code

$$C_\alpha : \mathbb{F}_p \mapsto \{\pm 1\}, \ \text{ defined as } \ C_\alpha(\lambda) = B_k(\lambda[\alpha]_d) \ \text{ (extended by } \ C_\alpha(0) = -1).$$

We construct another oracle $\mathcal{O}'$ that takes as input a base representation $h \in I_t(p)$, a Diffie-Hellman triple $g, g^a, g^b \in \mathbb{F}_{p^t}$, and $\lambda \xleftarrow{\$} \mathbb{F}_p^*$, and returns $\mathcal{O}(h, r_\lambda, g, g^a, g^b)$ if λ is d-th residue modulo p, where $r_\lambda^d \equiv \lambda (\text{mod } p)$, otherwise tosses a β_k-biased coin.

Suppose that there exists an oracle $\mathcal{O}$ such that

$$\big|\Pr_{\lambda,a,b}\big[\mathcal{O}(h,\lambda,g,g^a,g^b) = B_k\big(\big[\phi_{h,\widehat{h}_\lambda}(g^{ab})\big]_d\big)\big] - \beta_k\big| \geq \epsilon \tag{15}$$

where ϵ is a non-negligible quantity. Following the technique in Boneh and Shparlinski [5], we now show that

$$\big|\Pr_{\lambda,a,b}\big[\mathcal{O}'(h,\lambda,g,g^a,g^b) = B_k\big(\lambda\big[g^{ab}\big]_d\big)\big] - \beta_k\big| \geq \epsilon/d.$$

Let $E_{g^{ab}}$ be the event that $\mathcal{O}'(h,\lambda,g,g^a,g^b) = B_k\big(\lambda\big[g^{ab}\big]_d\big)$. Note that if λ is uniform in $\mathbb{F}_p^d \setminus \{0\}$ then r_λ is uniform in $\mathbb{F}_p^*$. Therefore, we have

$$\begin{aligned} \Pr[E_{g^{ab}}] \ &= \tfrac{1}{(d,p-1)} \Pr[E_{g^{ab}} | \lambda \in \mathbb{F}_p^d] + (1 - \tfrac{1}{(d,p-1)}) \Pr[E_{g^{ab}} | \lambda \notin \mathbb{F}_p^d] \quad \text{(according to Lemma 7)} \\ &\geq \tfrac{1}{(d,p-1)}(\beta_k + \epsilon) + (1 - \tfrac{1}{(d,p-1)})\beta_k \quad \text{(according to condition (15))} \\ &= \beta_k + \tfrac{\epsilon}{(d,p-1)} \geq \beta_k + \tfrac{\epsilon}{d}. \end{aligned}$$

Note that $t > d$ and $t \in poly(l)$ and therefore the above quantity is non-negligible.[4]

Accessibility. The oracle $\mathcal{O}'$ allows us to have access to a corrupted codeword $\widetilde{C}_\alpha$ of the above codeword defined as $\widetilde{C}_\alpha = \mathcal{O}'(h,\lambda,g,g^a,g^b)$. Therefore, if the oracle $\mathcal{O}$ has advantage ϵ then we have $|\Pr[C_\alpha(\lambda) = \widetilde{C}_\alpha(\lambda)]| \geq \beta_k + \epsilon/d$. Accessibility of the code C_α follows.

Concentration. The proof slightly generalizes that of FGPS. For a threshold $\tau > 0$, the τ-heavy characters of C_α belong to the set

$$\Gamma_{\alpha,\tau} = \{\chi_\beta \colon \beta = \lambda[\alpha]_d \text{ for } \lambda \in \Gamma_\tau\},$$

where Γ_τ is a set containing the τ-heavy coefficients of the function B_k. For each $\lambda \in \Gamma_\tau$, there exists a unique integer pair $(\xi_\lambda, \varsigma_\lambda) \in [0, 1/\tau] \times [0, 1/\tau]$. As in Theorem 2, the proof for concentration of the code $C_\alpha(\lambda)$ is now similar to those of [9, 15].

Recoverability. The proof for recoverability is also a generalization of the one in FGPS. First, by Lemma 1 we know that there exists a threshold τ which is polynomial in the non-negligible quantity ϵ and at least one τ-heavy Fourier character $\chi \neq 0$ for C_α and $\widetilde{C}_\alpha$ such that $\chi \in \mathsf{Heavy}_\tau(C_\alpha) \cap \mathsf{Heavy}_\tau(\widetilde{C}_\alpha)$.

[4] Note that it is possible that even if t is exponentially large, the value for $\epsilon/(d, p-1)$ may still be non-negligible—*e.g.,* when $(d, p-1)$ is small.

Given a fixed polynomial $h(x) \in I_t(p)$, on input $g, g^a, g^b \in \mathbb{F}_{p^t}$, the following algorithm that has access to $\mathcal{O}$ produces a polynomial size list of elements in $\mathbb{F}_{p^t}$ which contains g^{ab} with probability $1-\delta$.

Let τ be the threshold determined by Lemma 1. We write $\alpha = \sum_{i=0}^{t-1}[\alpha]_i x^i$ to denote $g^{ab} \in \mathbb{F}_{p^t}$. Again using the learning algorithm of AGS [1], we obtain a polynomial size list L_α of all the τ-heavy Fourier characters for $\widetilde{C}_\alpha$. If χ_β is a non-trivial τ-heavy character for C_α, we have $[\alpha]_d = \lambda^{-1}\beta$. Given $\chi_\beta \in L_\alpha$, we define $L_\beta = \{[\alpha]_d \colon [\alpha]_d = \lambda^{-1}\beta \text{ for } \eta \in \Gamma_\tau\}$.

Let $L = \bigcup_{\chi_\beta \in L_\alpha} L_\beta$, which is a set of polynomial size. Also we have $\alpha \in L$ with probability $1-\delta$. We can guess a result for $[\alpha]_d$ and hence get $[g^{ab}]_d$. The theorem now follows. ▮

Discussion.[5] It is worth mentioning that Theorem 5 proves what is slightly different in concept from that of the FGPS paper. In FGPS, it is shown that any bit prediction oracle must have negligible success probability ranging over *all representations*, whereas Theorem 5 shows that the success probability must be negligible ranging over a restricted class. On the surface, this leaves the possibility that there exists an oracle which predicts B_k for a large (non-negligible) class of representations that are not of the form given in Lemma 8, making it seem like a slightly weaker concept. This is also the reason that FGPS did not choose to use it. However, in any application, participants would agree upon some representation that they would like to use, and therefore our result does not limit its applicability and it is in fact simpler.

Following from Theorem 4 and Theorem 5, we obtain another main result: almost all individual bits of the CDH value of the traditional CDH problem over finite fields $\mathbb{F}_{p^t}$ for polynomial $t > 1$ are hard-core. Formally we have the following theorem:

Theorem 6 *Under the CDH assumption over $\mathbb{F}_{p^t}$ for any polynomial $t > 1$ (i.e., Assumption 3), for any PPT adversary $\mathcal{O}$, if $d \neq 0$, we have that for all $h \in I_t(p)$ the following quantity is negligible in l:*

$$\left| \Pr\left[\mathcal{O}(h, \lambda, g, g^a, g^b) = B_k\left(\left[\phi_{h,\widehat{h}_\lambda}(g^{ab})\right]_d\right) \middle| \lambda \xleftarrow{\$} \mathbb{F}_p^*; a, b \xleftarrow{\$} \{1, \cdots, p^2-1\}\right] - \beta_k \right|.$$

7 Concluding Remarks

FGPS provided the first results known for hard-core predicates for the CDH problem over finite fields. They defined a weaker variant of the CDH problem over finite fields of the form $\mathbb{F}_{p^t}$, and proved that the problem over $\mathbb{F}_{p^2}$ has a class of hard-core predicates.

We proved that the Partial-CDH problem over finite fields $\mathbb{F}_{p^2}$ is as hard as the conventional CDH problem over $\mathbb{F}_{p^2}$. This result provides a firm foundation for the main result of FGPS. Namely, FGPS and our result together prove the existence of hard-core predicates for the CDH problems over finite fields based on the traditional CDH assumption. This also answers the second question that FGPS raised: It is unlikely to reduce the Partial-CDH over $\mathbb{F}_{p^2}$ to the regular CDH problem over $\mathbb{F}_p$, as otherwise the CDH problems over $\mathbb{F}_{p^2}$ and $\mathbb{F}_p$ would be equivalent.

We advanced the list decoding approach, and for the first time, we applied it to the case of a non-multiplicative code. We proved that the Partial-CDH problem also admits the hard-core predicates for every individual bit of the other coordinate of the secret CDH value over a random representation of the finite field $\mathbb{F}_{p^2}$. By combining all our results, we obtained one of our main theorems: given an oracle $\mathcal{O}$ that predicts any bit of the CDH value over a random representation of the field $\mathbb{F}_{p^2}$ with non-negligible advantage, we can solve the *regular* CDH problem over $\mathbb{F}_{p^2}$ with non-negligible probability.

[5] This discussion is due to a personal communication with W. E. Skeith III (Aug. 2014).

We extended the Partial-CDH problem to the case of general finite fields $\mathbb{F}_{p^t}$ for $t > 1$ and defined the d-th CDH problem for any $0 \le d \le t-1$. We showed an interesting result that the regular CDH problem over $\mathbb{F}_{p^t}$ can be reduced to any d-th CDH problem ($0 \le d \le t-1$), and therefore all the d-th CDH problems over finite fields $\mathbb{F}_{p^t}$ are as hard as the regular CDH problem over the same fields.

We continued to prove that over finite fields $\mathbb{F}_{p^t}$ for any $t > 1$, each d-th CDH problem except $d \neq 0$ admits a large class of hard-core predicates, including every individual bit of d-th coordinate. Hence we obtain another strong result: almost all bits of the CDH value of the traditional CDH problem over finite fields $\mathbb{F}_{p^t}$ for $t > 1$ are hard-core.

We conclude this paper with two open questions. First, it is natural to ask whether 0-th CDH problem $\mathbb{F}_{p^t}$ for $t > 1$ admits hard-core predicates (and whether *all* individual bits of the secret CDH value over $\mathbb{F}_{p^t}$ for $t > 1$ are unpredictable). Second, we hope that the techniques developed in FGPS and our paper could be used to prove the existence of hard-core predicates for the CDH problem over $\mathbb{F}_p$.

Acknowledgments

We are greatly indebted to past PC members for pointing out an error in proving Theorem 2 (now fixed) and an initial idea for proving Theorem 4 on which our algorithm is based. Many thanks to William E. Skeith III for kindly verifying the proofs and for providing insightful corrections, comments, and suggestions. We thank Kai-Min Chung, Matt Franklin, Phil Rogaway, Xiaoyun Wang, and Haiyang Xue for helpful comments and discussion. Part of the work was carried out when Mingqiang visited UC Davis.

References

[1] A. Akavia, S. Goldwasser, and S. Safra. Proving hard-core predicates using list decoding. *FOCS 2003*, pp. 146–157, IEEE Computer Society, 2003.

[2] W. Alexi, B. Chor, O. Goldreich, and C. Schnorr. RSA and rabin functions: Certain parts are as hard as the whole. *SIAM Journal on Computing*, 17(2): 194–209, 1988.

[3] M. Ben-Or. Probabilistic algorithms in finite fields. *FOCS 1981*, 11: 394–398, 1981.

[4] M. Blum and S. Micali. How to generate cryptographically strong sequences of pseudorandom bits. *SIAM Journal on Computing*, 13(4): 850–864, 1984.

[5] D. Boneh and I. E. Shparlinski. On the unpredictability of bits of the elliptic curve diffie-hellman scheme. *CRYPTO 2001*, LNCS vol. 2139, pp. 201–212, Springer, 2011.

[6] W. Diffie and M. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT–22(6): 644–654, 1976.

[7] A. Duc and D. Jetchev. Hardness of computing individual bits for one-way functions on elliptic curves. *CRYPTO 2012*, LNCS vol. 7417, pp. 832–849, 2012.

[8] T. ElGamal. A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, IT–31(4): 469–472, 1985.

[9] N. Fazio, R. Gennaro, I. M. Perera, and W. E. Skeith III. Hard-core predicates for a Diffie-Hellman problem over finite fields. *CRYPTO 2013*, LNCS vol. 8043, pp. 148–165, 2013.

[10] J. von zur Gathen and J. Gerhard. Modern Computer Algebra. *Cambridge University Press*, 1999.

[11] O. Goldreich and L. A. Levin. A hard-core predicate for all one-way functions. *STOC 1989*, pp. 25–32, ACM press, 1989.

[12] S. Goldwasser and S. Micali. Probabilistic encryption. *JCSS*, 28(2):270–299, 1984.

[13] J. Håstad and M. Näslund. The security of individual RSA bits. *FOCS*, pp. 510–521, 1998.

[14] R. Lidl and H. Niederreiter. Finite Fields. *Addison-Wesley*, 1983.

[15] P. Morillo and C. Ràfols. The security of all bits using list decoding. *PKC 2009*, LNCS vol. 5443, pp. 15–33, Springer, 2009.

[16] M. Näslund. All bits in $ax + b \mod p$ are hard. *CRYPTO '96*, pp. 114–128, 1996.

[17] A. Slinko. A generalization of Komlós's theorem on random matrices. *New Zealand J. Math.* 30 (1): 81–86, 2001.

A. The Dual-Partial-CDH Problem

We define a *dual* variant of the Partial-CDH problem over $\mathbb{F}_{p^2}$: We say that the *Dual-Partial-CDH* problem is hard in $\mathbb{F}_{p^2}$ if for all $h \in I_2(p)$ no efficient algorithm given $g, A = g^a, B = g^b \in \mathbb{F}_{p^2}$ can output $\left[g^{ab}\right]_0 \in \mathbb{F}_p$. Formally we consider the following assumption:

Assumption 5 (The Dual-Partial-CDH assumption over $\mathbb{F}_{p^2}$). *We say that the Dual-Partial-CDH problem is hard in $\mathbb{F}_{p^2}$ if for any PPT adversary $\mathcal{A}$, his Dual-Partial-CDH advantage for all $h \in I_2(p)$*

$$\mathbf{Adv}^{\text{dpcdh}}_{\mathcal{A},\mathbb{F}_{p^2}} = \Pr\left[\mathcal{A}(p, h, g, g^a, g^b) = \left[g^{ab}\right]_0 \middle| a, b \xleftarrow{\$} \{1, \cdots, p^2 - 1\}\right]$$

is negligible in l.

The following theorem asserts that the Dual-Partial-CDH problem is also as hard as the conventional CDH problem. Therefore, both the Partial-CDH and Dual-Partial CDH problems are as hard as the conventional CDH problem over $\mathbb{F}_{p^2}$.

Theorem 7 *Suppose $\mathcal{A}$ is a Dual-Partial-CDH adversary that runs in time at most φ and achieves advantage $\mathbf{Adv}^{\text{dpcdh}}_{\mathcal{A},\mathbb{F}_{p^2}}$. Then there exists a CDH adversary $\mathcal{B}$, constructed from $\mathcal{A}$ in a blackbox manner, that runs in time at most 2φ plus the time to perform a small constant number of group operations and achieves advantage $\mathbf{Adv}^{\text{cdh}}_{\mathcal{B},\mathbb{F}_{p^2}} \geq 1/2 \cdot \mathbf{Adv}^{\text{dpcdh}}_{\mathcal{A},\mathbb{F}_{p^2}}$.*

Proof: We assume that $p \neq 2$, as the case $p = 2$ is trivial. Our CDH adversary $\mathcal{B}$ works as follows, given a challenge instance of the CDH problem $(g^a, g^b) \xleftarrow{\$} (\mathbb{F}_{p^2})^2$ and given a Dual-Partial-CDH adversary $\mathcal{A}$.

Adversary $\mathcal{B}$ begins by choosing integers $r, s \xleftarrow{\$} \mathbb{Z}_{p^2-1}$ and then computes two pairs (g^{a+r}, g^{b+s}) and $(g^{2(a+r)}, g^{b+s})$. Let $A = a + r$ and $B = b + s$. Adversary $\mathcal{B}$ runs the Dual-Partial-CDH adversary $\mathcal{A}$ on input the generated instances (g^A, g^B) and (g^{2A}, g^B) under the representation determined by $h(x) = x^2 + h_1 x + h_0 \in I_2(p)$, and obtains $\left[g^{AB}\right]_0$ and $\left[g^{2AB}\right]_0$.

In the finite field $\mathbb{F}_{p^2}$, we get $(g^{AB})^2 \bmod h(x) = g^{2AB}$. Hence,

$$(\left[g^{AB}\right]_1 x + \left[g^{AB}\right]_0)^2 \bmod h(x) = \left[g^{2AB}\right]_1 x + \left[g^{2AB}\right]_0.$$

Comparing the coefficient of the degree-0 term in the above equation, we have $\left[g^{AB}\right]_1$ is a root of the following equation with y being the unknown:

$$h_0 y^2 = \left[g^{AB}\right]_0^2 - \left[g^{2AB}\right]_0.$$

Note we have $h_0 \neq 0$, since otherwise $h(x)$ is reducible. Therefore, the equation has two roots. Namely, adversary $\mathcal{B}$ can get two possible values on $\left[g^{AB}\right]_1$. Along the same lines as in Theorem 1, adversary $\mathcal{B}$ can obtain two possible g^{ab} and guess the correct one. By a standard argument, the proof is easily completed. ▮

B. Fourier Concentration of $C'_\alpha(\eta)$

The proof of the Fourier concentration of the multiplication code $C'_\alpha(\eta) = B_k(\eta[\alpha]_1)$ is the same as that of FGPS which follows Morillo and Ràfols [15]. We now provide the proof in more detail.

For $\beta \in \mathbb{F}_p$, if $C'_\alpha(\eta)$ is ϵ-concentrated in $\Gamma_\alpha = \{\chi_\beta\}$ then $B_k(\eta[\alpha]_1)$ is ϵ-concentrated in the set $\{\chi_\eta \colon \eta = \beta[\alpha]_1^{-1}\}$. Thus, we just need to prove the Fourier concentration of $B_k(\eta[\alpha]_1)$. We would need to analyze the Fourier coefficients of $B_k \colon \mathbb{F}_p \to \{\pm 1\}$.

We define $g(x)$ as

$$g(x) = \frac{B_k(x) + B_k(x + 2^k)}{2}.$$

Morillo and Ràfols [15] notice that the Fourier transform of $B_k(x)$ and the Fourier transform of $g(x)$ can be related with the following equation:

$$\widehat{g}(\eta) = \frac{\omega_p^{2^k \eta} + 1}{2} \widehat{B_k}(\eta),$$

where $\eta \in \mathbb{F}_p$ and $\omega_p = e^{2\pi i/p}$.

In particular, assuming $\eta \in [-\frac{p-1}{2}, \frac{p-1}{2}]$, they consider the following two cases for η:

1. $\eta \geq 0$, consider $\delta_{\eta,k} := 2^k \eta - (p-1)/2 \bmod p$ and let $\lambda_{\eta,k} \in [0, 2^{k-1} - 1]$ be the unique integer for which $2^k \eta = (p-1)/2 + \delta_{\eta,k} + p\lambda_{\eta,k}$.

2. $\eta < 0$, consider $\delta_{\eta,k} := 2^k \eta + (p+1)/2 \bmod p$ and let $\lambda_{\eta,k} \in [0, 2^{k-1} - 1]$ be the unique integer for which $2^k \eta = -(p+1)/2 + \delta_{\eta,k} + p\lambda_{\eta,k}$.

For both cases, there are unique integers $\mu_{\eta,k} \in [0, r]$, where r is the largest integer less than $p/2^{k+1}$ and $r_{\eta,k} \in [0, 2^k - 1]$ such that $a_p(2^k \eta - (p-1)/2) = \mu_{\eta,k} 2^k + r_{\eta,k}$, where $a_p(x) = \min\{x \bmod p, p - x \bmod p\}$ for $x \bmod p$ being taken in $[0, p-1]$. The definition of Γ_τ in Section 4 is as follows

$$\Gamma_\tau = \{\eta \colon (\lambda_{\eta,k}, \mu_{\eta,k}) \in [0, 1/\tau] \times [0, 1/\tau]\}.$$

Here we select τ such that $1/\tau = \mathit{poly}(\log p)$. Morillo and Ràfols [15] obtain the following upper bound of $\widehat{B_k}(\eta)$:

$$|\widehat{B_k}(\eta)|^2 < O(\frac{1}{\lambda_{\eta,k}^2 \mu_{\eta,k}^2}).$$

Now one can conclude that $B_k(\eta[\alpha]_1)$ is Fourier concentrated.

C. Proof of Lemma 6

To help understand the proof, we slightly rephrase Lemma 6 in the language of vector space.

Lemma 9 *Let $\mathbb{V}$ be a vector space over the finite field $\mathbb{F}_p$ of dimension m. Let $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_m}$ be m independent, random vectors in the vector space $\mathbb{V}$. Let the matrix $M = (\boldsymbol{v'_1}, \boldsymbol{v'_2}, \cdots, \boldsymbol{v'_m})$, where $\boldsymbol{v'_i}$ (where $1 \le i \le m$) is the transpose of $\boldsymbol{v_i}$. Then the probability that the rank of matrix M (i.e.,*Rank(M)*) is m is at least $e^{-\frac{2}{p-1}}$.*

Proof: Let E_i $(1 \le i \le m)$ denote the event that $\boldsymbol{v_i} \notin \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$, where $\text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$ is the subspace generated by vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}$.

We first observe that assuming $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}$ are linearly independent, we have $\boldsymbol{v_i} \notin \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$ if and only if $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}$, and $\boldsymbol{v_i}$ are linearly independent. On the one hand, if $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}$, and $\boldsymbol{v_i}$ are linearly independent, we have $\boldsymbol{v_i} \notin \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$. Otherwise there exist $\lambda_z \in \mathbb{F}_p$, $z = 1, 2, \cdots, i-1$, such that $\sum_{z=1}^{i-1} \lambda_z \boldsymbol{v_z} = \boldsymbol{v_i}$. This clearly causes a contradiction. On the other, suppose $\boldsymbol{v_i} \notin \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$, and there exist $\lambda_z \in \mathbb{F}_p$, $z = 1, 2, \cdots, i$, such that $\sum_{z=1}^{i} \lambda_z \boldsymbol{v_z} = \boldsymbol{0}$. If $\lambda_i \neq 0$, we have $\boldsymbol{v_i} = \sum_{z=1}^{i-1} \lambda_i^{-1} \lambda_z \boldsymbol{v_z}$, contradicting $\boldsymbol{v_i} \notin \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$. Thus $\lambda_i = 0$. Since $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}$ are linearly independent, we have $\lambda_1 = \lambda_2 = \cdots = \lambda_i = 0$.

It is easy to see that if the events E_i $(1 \le i \le m)$ occur simultaneously then the rank of the matrix is m. Hence, we have $\Pr[\text{Rank}(M) = m] \ge \Pr[E_1 E_2 \cdots E_m]$.

Below, we give a lower bound on $\Pr[E_1 E_2 \cdots E_m]$. For random vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}, \boldsymbol{v_i}$ in vector space $\mathbb{V}$, we actually have that the probability of $\boldsymbol{v_i} \in \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$ is at most $\frac{1}{p^{m-i+1}}$. This is because the dimension of vector space $\mathbb{V}$ is m, while the dimension of vector subspace $\text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_{i-1}}\}$ is at most $i-1$. Thus, the probability that E_i occurs is at least $1 - \frac{1}{p^{m-i+1}}$. We now have

$$\Pr[E_1 E_2 \cdots E_m] \ge \prod_{i=1}^{m} (1 - \frac{1}{p^{m-i+1}}) = \prod_{i=1}^{m} (1 - \frac{1}{p^i}) > \prod_{i=1}^{\infty} (1 - \frac{1}{p^i}).$$

It remains to lower bound $\prod_{i=1}^{\infty} (1 - \frac{1}{p^i})$. As $\prod_{i=1}^{\infty} (1 - \frac{1}{p^i}) = e^{\sum_{i=1}^{\infty} \ln(1 - \frac{1}{p^i})}$, we only need to bound the sum $\sum_{i=1}^{\infty} \ln(1 - \frac{1}{p^i})$. Using Taylor series expansion, we obtain

$$\sum_{i=1}^{\infty} \ln(1 - \frac{1}{p^i}) = -\sum_{i=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{np^{in}} \ge -\sum_{i=1}^{\infty} \frac{2}{p^i} = -\frac{2}{p-1}.$$

Thus, $\Pr[\text{Rank}(M) = m] \ge \Pr[E_1 E_2 \cdots E_m] \ge e^{-\frac{2}{p-1}}$. This completes the proof of this lemma. □