SUBSET PRODUCT WITH ERRORS OVER UNIQUE FACTORIZATION DOMAINS AND IDEAL CLASS GROUPS OF DEDEKIND DOMAINS

TREY LI

ABSTRACT. It has been half a century since the first several NP-complete problems were discovered by Cook, Karp and Levin in the early 1970s. Till today, thousands of NP-complete problems have been found. Most of them are of combinatorial flavor. We discover new possibilities in purer mathematics and introduce more structures to the theory of computation. We propose a family of abstract problems related to the subset product problem. To describe hardness of abstract problems, we propose a new hardness notion called global-case hardness, which is stronger than worst-case hardness and incomparable with average-case hardness. It is about whether all prespecified subproblems of a problem are NP-hard. We prove that our problems are generally NP-hard in all/a wide range of unique factorization domains with efficient multiplication or all/a wide range of ideal class groups of Dedekind domains with efficient ideal multiplication.

1. INTRODUCTION

The study of NP-completeness was initiated around the early 1970s. In 1971, Cook discovered NP-completeness by reducing NP problems to the Boolean satisfiability problem [Coo71]; Karp further found another 21 NP-complete problems in 1972 [Kar72]; and Levin published his independent discovery of NP-completeness in 1973 [Lev73].

The subset product problem (SP) is a classical NP-complete problem [GJ79, p. 224]. It is given $n+1$ integers $a_1,\dots,a_n,X$, find a binary vector $(x_1,\dots,x_n)\in\{0,1\}^n$ such that $\prod_{i=1}a_i^{x_i}=X$. A natural variant is the modular subset product problem (MSP), which is given $n+2$ integers $a_1,\dots,a_n,X,N$, find a binary vector $(x_1,\dots,x_n)\in\{0,1\}^n$ such that $\prod_{i=1}a_i^{x_i}\equiv X \pmod N$. Recently in [GZ19] and [GL21] a special MSP with prime bases $a_i=p_i$ and safe prime modulus $N=2p+1$ was found to have applications in cryptography. Assuming certain conjectures, the authors showed that average-case the special problem is at least as hard as worst-case discrete logarithm problem in $\mathbb{Z}_q^\times$.

In this paper we propose a family of problems related to SP. The representatives are the subset product with errors problems over unique factorization domains (UFDs) and ideal class groups of Dedekind domains. We restrict our discussions on their NP-hardness and leave open whether they are in NP. We show that the non-modular problems are generally NP-hard for *all* UFDs with efficient multiplication and *all* Dedekind domains with efficient ideal multiplication; and the modular problems are generally NP-hard for efficient-Gaussian-predictable UFDs with efficient multiplication and Dedekind domains with efficient ideal multiplication and efficient-Gaussian-predictable ideal class group; where "efficient-Gaussian-predictability" is the key tool for our proofs, which allows us to embed a non-modular problem into a modular problem by a lattice argument.

This is the 1st paper of the series.

Date: October 1, 2022.

Email: treyquantum@gmail.com

2. SUBSET PRODUCT WITH ERRORS

Note that if a lot of random SP or MSP instances with respect to the same solution $(x_1,\ldots,x_n)$ are provided, the solution is believed to be easy to find. In fact, if any one of the instances is of low density (i.e., $n \ll \log_2 N$), then by a (quantum) discrete logarithm algorithm the problem is transformed into a special closest vector problem that can be solved by the algorithm in [LM09] via solving an LLL solvable unique shortest vector problem.

This motivates the invention of our subset product with errors problem, which is believed to be hard even provided polynomially many random instances with respect to the same solution.

The idea is to reserve some of the integers $a_{n-k+1},\ldots,a_n$ in SP, or equivalently, insert invisible errors $a_n,\ldots,a_{n+k}$ into SP, so that the reduction to the easy closest vector problem is interrupted due to the lattice basis cannot be created without knowing all a_i's.

In particular, we use primes as the errors and define the *subset product with errors problem* (SPE) as given $n+1$ integers $a_1,\ldots,a_n,X$ and a set of primes $L=\{\ell_1,\ldots\ell_k\}$ that are coprime to $a_1,\ldots,a_n$, find a binary vector $(x_1,\ldots,x_n)\in\{0,1\}^n$ and a square-free integer E factored over L such that

$$\prod_{i=1}^{n} a_i^{x_i}\cdot E = X.$$

SPE is at least as hard as SP because given an SP

$$(a_1,\ldots,a_n,X)$$

we can choose a set of primes $L=\{\ell_1,\ldots,\ell_k\}$ that are coprime to $a_1,\ldots,a_n$ and solve the SPE

$$(a_1,\ldots,a_n,L,X')$$

with

$$X'=XE,\ E=\prod_{i=1}^{k}\ell_i^{e_i},\ e_i\in\{0,1\}$$

for a vector $(x_1,\ldots,x_n)\in\{0,1\}^n$ and an integer E' such that

$$\prod_{i=1}^{n} a_i^{x_i}\cdot E' = X'.$$

By the unique factorization theorem,

$$E'=E$$

and thus

$$\prod_{i=1}^{n} a_i^{x_i} = X.$$

Therefore the NP-hardness of SPE is clear.

Also it makes more sense to consider a set L of size super-polynomial in n, since otherwise if $|L|$ is polynomial then SPE is simply SP with n being extended to $n+|L|$. Specifically, suppose $L=\{\ell_1,\ldots,\ell_k\}$ is small, then given an SPE instance

$$(a_1,\ldots,a_n,L,X),$$

we solve the SP

$$(a_1,\ldots,a_n,\ell_1,\ldots,\ell_k,X)$$

for a binary vector $(x_1,\ldots,x_n,e_1,\ldots,e_k)\in\{0,1\}^{n+k}$ such that

$$\prod_{i=1}^{n} a_i^{x_i}\cdot\prod_{j=1}^{k}\ell_j^{e_j}=X.$$

Then $(x:=(x_1,\ldots,x_n),E:=\prod_{j=1}^{k}\ell_j^{e_j})$ is a solution to the SPE.

The modular version of SPE, called the *modular subset product with errors problem* (MSPE) is defined in the natural way, namely given $n+2$ integers $a_1,\ldots,a_n,X,N$ and a set of primes $L=\{\ell_1,\ldots,\ell_k\}$ that are coprime to $a_1,\ldots,a_n$, find a binary vector $(x_1,\ldots,x_n)\in\{0,1\}^n$ and a square-free integer E factored over L such that

$$\prod_{i=1}^{n} a_i^{x_i}\cdot E\equiv X \pmod{N}.$$

We prove its NP-hardness by proving the general NP-hardness of the following more general problem over UFDs.

3. UFD SUBSET PRODUCT WITH ERRORS

Now we generalize the ground set of SP from $\mathbb{Z}$ to UFDs[1] and define the *unique factorization domain subset product problem* (USP) to be given $n+1$ elements $a_1,\ldots,a_n,X$ of a UFD R, find a binary vector $(x_1,\ldots,x_n)\in\{0,1\}^n$ such that

$$\prod_{i=1}^{n} a_i^{x_i}=X.$$

Note that SP is a special USP with $R=\mathbb{Z}$.

The modular version of USP, called the *modular unique factorization domain subset product problem* (MUSP), is given $n+1$ elements $a_1,\ldots,a_n,X$ of a UFD R, an ideal I of R, find a binary vector $(x_1,\ldots,x_n)\in\{0,1\}^n$ such that

$$\prod_{i=1}^{n} a_i^{x_i}\equiv X \pmod{I},$$

where we define $a=b \pmod{I}$ to be $a+I=b+I$ for $a,b\in R$.

The noisy version of USP, called the *unique factorization domain subset product with errors problem* (USPE), is given $n+1$ elements $a_1,\ldots,a_n,X$ of a UFD R as well as a set $L\subset R$ of prime elements of R that are coprime to all a_i, find a binary vector $(x_1,\ldots,x_n)\in\{0,1\}^n$ and a square-free ring element E such that

$$\prod_{i=1}^{n} a_i^{x_i}\cdot E=X.$$

The modular version of USPE, called the *modular unique factorization domain subset product with errors problem* (MUSPE), is given $n+1$ elements $a_1,\ldots,a_n,X$ of a UFD R, an ideal I of R, as well as a set $L\subset R$ of prime elements of R that are coprime prime to all

[1]In this paper, whenever we talk about a computational problem over a UFD or an ideal class group, we assumably mean a UFD with efficient multiplication or an ideal class group with efficient ideal multiplication respectively, where "efficient " conventionally means "polynomial time computable".

a_i, find a binary vector $(x_1,\ldots,x_n)\in\{0,1\}^n$ and a square-free ring element E factored over L such that

$$\prod_{i=1}^{n} a_i^{x_i}\cdot E \equiv X \pmod{I}.$$

4. GLOBAL-CASE HARDNESS

Note that our problems are defined in an abstract way. For an abstract problem, the traditional notion of NP-hardness is not very interesting. We want not only to know whether the problem is hard in the worst-case, but also whether it is hard in *all* specified cases.

For example, to prove NP-hardness of USP in the traditional sense, we only need to prove the existence of *one* UFD such that the subproblem over that specific UFD is NP-hard. However, we are more interested in whether the problem is generally NP-hard for *all* UFDs. We therefore introduce a stronger notion beyond NP-hardness.

Intuitively, worst-case hardness means *some instances are hard*; and average-case hardness means *most instances are hard*. Now we define *global-case hardness* to mean *every specified subproblem has some hard instances*. If all the specified subproblems have hard instances, we call the problem *generally NP-hard* (with respect to the specification).

Global case hardness is stronger than worst-case hardness but incomparable with average-case hardness. The first relation is obvious since globally distributed hard instances definitely imply the existence of some hard instances. The second relation is not hard to see either. For one direction, suppose the problem is globally hard for a family of specified subproblems, i.e., all specified subproblems has hard instances. However these hard instances might just cover a small portion of the subproblems respectively and hence only cover a small portion of the whole problem and thus the whole problem is not average-case hard. For the other direction, even if the problem is average-case hard, i.e., most instances are hard, it is still possible that there is a specified subproblem whose instances are all easy and thus the whole problem is not global-case hard with respect to the specification.

The main motivation of our study of global-case hardness is to discover more possibilities in pure mathematics for the theory of computation, and to introduce more structures to classical NP-hard problems.

The other motivation is from cryptography. In cryptography, we have always been struggling about choosing a good ground structure (e.g., a group, a ring, etc.) for a computational problem to be hard. For example, the discrete logarithm problem (DLP) is easy in the additive group modulo N; and thus for application in cryptography, we have to choose a proper group (e.g., a multiplicative group modulo a prime or an elliptic curve group) so that the problem is not that trivial.

To think outside this research convention, we want to investigate whether there is a problem that is generally hard for an extremely wide range of mathematical structures. For instance, suppose a problem is generally NP-hard for all groups/rings, then we can choose any group/ring to use[2].

We will show that USPE is generally NP-hard for all UFDs; and MUSPE is generally NP-hard for a very wide range of UFDs.

[2]Of cause, a basic requirement is that the group/ring operation should be polynomial time computable.

5. PRODUCT EMBEDDINGS

Let A and B be two search problems, which are essentially two sets of binary strings, where each string represents a problem instance. A reduction from A to B is a polynomial time algorithm that takes as input any instance $a \in A$ and transforms it into an instance $b \in B$ such that finding any solution to b guarantees a solution to a. To prove the NP-hardness of a problem B, it is sufficient to find a reduction from an NP-hard problem A to B.

The exact cover problem (XC) is a well-known NP-hard problem, which is given a set S and a family C of subsets of S, find a subfamily D of C with the sets in D pairwise disjoint and union to S. To prove the NP-hardness of MUSPE, we prove the reduction chain

$$\text{XC} \leq \text{USP} \leq \text{USPE} \leq \text{MUSPE}$$

by proving a polynomial time computable embedding chain

$$\text{XC} \hookrightarrow \text{USP} \hookrightarrow \text{USPE} \hookrightarrow \text{MUSPE}$$

that preserves the solution set, where embedding means bijective mapping.

The embedding from XC to USP can be achieved in the traditional sense, i.e. by interpreting subsets as integers and elements as prime factors. The embedding from USP to USPE is analogous to the reduction from SP to SPE discussed earlier. In the following we provide a general tool to deal with the embedding from USPE to MUSPE, namely from a non-modular problem to a modular problem.

We consider product embeddings for general groups/rings. Let M be a group/ring; and M/N be a quotient group/ring. We consider how to embed a set of products

$$S = \left\{ \prod_{i=1}^{n} a_i^{x_i} \;\middle|\; a_i \in M, x_i \in \{0,1\} \right\}$$

into M/N. It is sufficient to embed the set of products of irreducible factors

$$T = \left\{ \prod_{i=1}^{m} p_j^{e_j} \;\middle|\; p_j \middle| \prod_{i=1}^{n} a_i^{x_i}, e_j \in \{0,s\} \right\}$$

into M/N, where $s = nd$ with d the greatest exponent in all the factorizations of $a_1, \ldots, a_n$.

Denote the equivalence class a modulo N as $[a]$. In order to show that the morphism

$$\varphi : T \to M/N;\ \prod_{i=1}^{m} p_j^{e_j} \mapsto \left[\prod_{i=1}^{m} p_j^{e_j} \right]$$

is injective, i.e., there do not exist two different vectors $(e_1, \ldots, e_m) \neq (f_1, \ldots, f_m) \in \{0, \ldots, s\}^n$ such that

$$\prod_{i=1}^{m} p_j^{e_j} \neq \prod_{i=1}^{m} p_j^{f_j}$$

but

$$\left[\prod_{i=1}^{m} p_j^{e_j} \right] = \left[\prod_{i=1}^{m} p^{f_j} \right],$$

it is sufficient to avoid nonzero short vectors $z = (e_1 - f_1, \ldots, e_m - f_m) \in \{-s, \ldots, s\}^m$ of length (i.e. Euclidean norm) $\leq s\sqrt{m}$. For this, it is sufficient to require that the first minimum (i.e.

the Euclidean norm of the shortest vector) of the relation lattice[3]

$$\Lambda = \left\{ (z_1, \ldots, z_m) \in \mathbb{Z}^m \;\middle|\; \left[\prod_{i=1}^{m} p_j^{z_j} \right] = 1_{M/N} \right\}$$

defined by the kernel $\phi^{-1}(1_{M/N})$ of the group/ring homomorphism

$$\phi : \mathbb{Z}^m \to M/N;\ (z_1, \ldots, z_m) \mapsto \left[\prod_{i=1}^{m} p_j^{z_j} \right]$$

to be greater than $s\sqrt{m}$.

For this purpose, we define (c, s, m)-Gaussian-predictable quotient groups/rings, which are quotient groups/rings that the first minimum of the relation lattice Λ with respect to some primes $p_1, \ldots, p_m$ is predicted by the Gaussian heuristic with a one-sided bias upper bounded by a factor of c.

Definition 1. Let $c \in \mathbb{R}_{>1}$, $m, s \in \mathbb{N}$. Let M be a (multiplicative) group/ring. Let $P = \{p_1, \ldots, p_m\}$ be a set of irreducible elements of M; and $S = \left\{ \prod_{i=1}^{m} p_i^{e_i} \;\middle|\; p_i \in P, e_i \in \{0, \ldots, s\} \right\}$ be the set of products of elements in P with exponents upper bounded by s. Let $\lambda_1(\Lambda)$ be the first minimum of the lattice Λ and $gh(\Lambda)$ be the Gaussian heuristic for $\lambda_1(\Lambda)$. Then the pair $(M/N, P)$ is said to be (c, m, s)-*Gaussian-predictable* if

(1) $M/N = (p_1, \ldots, p_m)$;
(2) $|M/N| > (cs\sqrt{2\pi e})^m$;
(3) $gh(\Lambda) < c \cdot \lambda_1(\Lambda)$.

We sometimes say the quotient group/ring M/N itself (c, m, s)-*Gaussian-predictable*, with P implied. We say the group/ring M *Gaussian-predictable* if it has a (c, m, s)-Gaussian-predictable quotient group/ring M/N. We say M *efficient-Gaussian-predictable* if there is a polynomial time algorithm to find a (c, m, s)-Gaussian-predictable pair $(M/N, P)$.

Note that Condition (1) and (2) are for the use of the Gaussian heuristic

$$gh(\Lambda) = \sqrt{\frac{m}{2\pi e}} \operatorname{vol}(\Lambda)^{\frac{1}{m}}$$

of the first minimum $\lambda_1(\Lambda)$. (See [Duc18] for an introduction to the Gaussian heuristic.)

Also Condition (3) is not a serious restriction because if we fix $\epsilon > 0$, then for all sufficiently large m, we have $(1-\epsilon) \cdot gh(\Lambda) < \lambda_1(\Lambda) < (1+\epsilon) \cdot gh(\Lambda)$ for a randomly chosen lattice Λ of dimension m [Sie45; HPS08, p. 402]. Now Condition (3) only requires a one-sided bound with respect to a flexible constant c, which is easy to satisfy.

[3] To see why Λ is a lattice, we first notice that any (additive) subgroup of $\mathbb{R}^m$ is a lattice. We then consider the following two situations. (1) When M is a group, we view it as a multiplicative group so that products of the form $p_1^{z_1} \cdots p_m^{z_m}$ make sense. Then the group homomorphism ϕ maps from the additive group $\mathbb{Z}^m$ to the multiplicative quotient group M/N. The kernel $\ker(\phi) := \Lambda$ is an (additive) subgroup of $\mathbb{Z}^m$, which is an additive subgroup of $\mathbb{R}^m$. Hence Λ is a lattice. (2) When M is a ring, then the ring homomorphism ϕ maps from the ring $\mathbb{Z}^m$ to the quotient ring M/N. The kernel $\ker(\phi) := \Lambda$ is an ideal of $\mathbb{Z}^m$ hence is an additive subgroup of $\mathbb{Z}^m$ and thus an additive subgroup of $\mathbb{R}^m$. Hence Λ is a lattice. This is why we want M to be at least a group/ring; otherwise we may not have a lattice Λ and that we cannot argue about the first minimum of Λ or make use of the Gaussian heuristic.

LEMMA 1. Let $(M/N, P)$ be a (c, m, s)-Gaussian-predictable pair, where $P := \{p_1, \ldots, p_m\}$. Let $T = \left\{\prod_{i=1}^{m} p_i^{e_i} \;\middle|\; e_i \in \{0, \ldots, s\}\right\}$. Then the morphism

$$\varphi : T \to M/N;\ \prod_{i=1}^{m} p_j^{e_j} \mapsto \left[\prod_{i=1}^{m} p_j^{e_j}\right]$$

is injective.

Proof. Consider the lattice Λ. Since $(M/N, P)$ is (c, m, s)-Gaussian-predictable, we have that

$$\lambda_1(\Lambda) > (1/c) \cdot gh(\Lambda) = (1/c) \cdot \sqrt{\frac{m}{2\pi e}}\, \mathrm{vol}(\Lambda)^{\frac{1}{m}},$$

where the volume $\mathrm{vol}(\Lambda)$ of the lattice Λ is given by the size of the image $\mathrm{im}(\phi)$ of ϕ by the first group/ring isomorphism theorems.

Now, since $(M/N, P)$ is (c, m, s)-Gaussian-predictable, P generates M/N. Hence ϕ is surjective and thus $\mathrm{vol}(\Lambda) = \mathrm{im}(\phi) = |M/N|$. Again by the Gaussian predictability of $(M/N, P)$, we have that $|M/N| > (cs\sqrt{2\pi e})^m$. We therefore have a lower bound for the first minimum as

$$\lambda_1(\Lambda) > (1/c) \cdot \sqrt{\frac{m}{2\pi e}} |M/N|^{\frac{1}{m}} = s\sqrt{m}.$$

Now, suppose for contradiction that there exist two different elements in T map to the same element in M/N. I.e., there exist two different vectors $(e_1, \ldots, e_m) \neq (f_1, \ldots, f_m) \in \{0, \ldots, s\}^m$ such that

$$\prod_{j=1}^{m} p_j^{e_j} \neq \prod_{i=1}^{m} p_j^{f_j}$$

but

$$\left[\prod_{j=1}^{m} p_j^{e_j}\right] = \left[\prod_{j=1}^{m} p_j^{f_j}\right].$$

Then

$$\left[\prod_{j=1}^{m} p_j^{e_j - f_j}\right] = 1_{M/N}$$

with $z := (e_1 - f_1, \ldots, e_m - f_m) \in \{-s, \ldots, s\}^m$ nonzero. I.e., there is a nonzero vector $z \in \Lambda$ of length $\leq s\sqrt{m}$, contradicting $\lambda_1(\Lambda) > s\sqrt{m}$. □

6. GENERAL NP-HARDNESS OF MUSPE

Having defined the concept of efficient-Gaussian-predictability, we are now ready to prove the general NP-hardness of the modular problem MUSPE.

Let $(S = \{s_1, \ldots, s_m\}, C = \{C_1, \ldots, C_n\})$ be an XC. For convenience we assign an order to the elements s_i as well as the subsets C_i and denote the problem as $(S = (s_1, \ldots, s_m), C = (C_1, \ldots, C_n))$. We define the *characteristic matrix* of (S, C) to be the binary matrix $A \in \{0, 1\}^{m \times n}$ with each column $A_{*,i}$ the characteristic vector of the subset C_i.

THEOREM 1. MUSPE is generally NP-hard for all UFDs R where there is a polynomial time algorithm to compute its multiplication and there is a polynomial time algorithm to find a Gaussian-predictable pair $(R/I, P)$ for some ideal I of R.

Proof. Let (S,C) be an XC with characteristic matrix $A \in \{0,1\}^{m\times n}$. Note that the characteristic matrix of any XC instance is easy to find hence we treat it as a part of the XC instance. Let d be the largest entry of A. We have that $d = 1$. We show the reduction chain

$$\text{XC} \leq \text{USP} \leq \text{USPE} \leq \text{MUSPE}.$$

(1) Find a $(c, m+k, s)$-Gaussian-predictable pair $(R/I, P)$ with R a UFD, I an ideal of R, $s := nd$, and k a natural number (typically super-polynomially large). Denote $P := \{p_1,\ldots,p_m,\ell_1,\ldots,\ell_k\}$ and $L := \{\ell_1,\ldots,\ell_k\}$.

(2) From XC to USP. Compute $a_1,\ldots,a_n$ with

$$a_i = \prod_{j=1}^{m} p_j^{A_{j,i}}$$

for all $i \in [n]$; and compute

$$X = \prod_{i=1}^{m} p_i.$$

By unique factorization in UFDs, a vector x is a solution to the USP $(a_1,\ldots,a_n,X)$ if and only if it is a solution to the XC (S,C).

(3) From USP to USPE. Choose a binary vector $(u_1,\ldots,u_k) \in \{0,1\}^k$ of Hamming weight polynomial in n (e.g. 1) and compute

$$E = \prod_{i=1}^{k} \ell_i^{u_i}.$$

Then $(a_1,\ldots,a_n,L,XE)$ is an USPE. By unique factorization, the pair (x,E) is a solution to this USPE if and only if x is a solution to the USP $(a_1,\ldots,a_n,X)$.

(3) From USPE to MUSPE. For one direction, it is obvious that all solutions to the USPE are solutions to the MUSPE. For the other direction, suppose for contradiction that there exists a solution to the MUSPE that is not a solution to the USPE. I.e., there exists a pair of binary vectors $(x_1,\ldots,x_n,u_1,\ldots,u_k) \neq (y_1,\ldots,y_n,v_1,\ldots,v_k) \in \{0,1\}^{m+k}$ such that

$$\left[\prod_{i=1}^{n} a_i^{x_i} \prod_{j=1}^{k} \ell_j^{u_j}\right] = \left[\prod_{i=1}^{n} a_i^{y_i} \prod_{j=1}^{k} \ell_j^{v_j}\right]$$

but

$$\prod_{i=1}^{n} a_i^{x_i} \prod_{j=1}^{k} \ell_j^{u_j} \neq \prod_{i=1}^{n} a_i^{y_i} \prod_{j=1}^{k} \ell_j^{v_j}.$$

This means that there exist two vectors $(e_1,\ldots,e_{m+k}) \neq (f_1,\ldots,f_{m+k}) \in \{0,\ldots,s\}^{m+k}$ such that

$$\left[\prod_{i=1}^{m+k} p_i^{e_i}\right] = \left[\prod_{i=1}^{m+k} p_i^{f_i}\right]$$

but

$$\prod_{i=1}^{m+k} p_i^{e_i} \neq \prod_{i=1}^{m+k} p_i^{f_i}$$

where $(p_{m+1},\ldots,p_{m+k}) := (\ell_1,\ldots,\ell_k)$. This contradicts Lemma 1. Hence a pair (x,E) is a solution to the MUSPE if and only if it is a solution to the USPE. □

Remark. We take $R := \mathbb{Z}$ as an example to see why a Gaussian-predictable pair finding algorithm is not a strong assumption. We find a (c,m,s)-Gaussian-predictable pair as follows. We first set $c = 2^m$ (very large) and $s := nd$; then set q to be the smallest prime such that $q-1 > (cs\sqrt{2\pi e})^m$; then sample m primes $p_1,\ldots,p_m$ from $[q^{1/m},q]$. Note that $\prod_{i=1}^m p_i > q-1$. Hence it is not the trivial case where all products $\prod_{i=1}^m p_i^{x_i}$ with $x_i \in \{0,1\}$ are initially in $\mathbb{Z}_q^\times$. Now note that there are $(q-1)^m \gg q-1$ different integers of the form $\prod_{i=1}^m p_i^{z_i}$, where $z_i \in \{0,\ldots,q-2\}$; and that the primes are random thus we expect that the mod q reduced numbers $\prod_{i=1}^m p_i^{z_i} \pmod q$ distribute quite uniformly over $\mathbb{Z}_q^\times$. Hence we expect that $\mathbb{Z}_q^\times = (p_1,\ldots,p_m)$ with high (i.e. at least noticeable if not overwhelming) probability. Namely Condition (1) of Gaussian predictability is satisfied with high probability. Also, Condition (2), namely $|\mathbb{Z}_q^\times| > (cs\sqrt{2\pi e})^m$ follows immediately from the choice of q. As to Condition (3), by the randomness of the primes $p_1,\ldots,p_m$, it is very reasonable to expect that the Gaussian heuristic $gh(\Lambda) < 2^m \cdot \lambda_1(\Lambda)$ with extremely high (at least noticeable if not overwhelming) probability otherwise the Gaussian heuristic makes no sense — recall that if we fix $\epsilon > 0$, then for all sufficiently large m, we have $(1-\epsilon)\cdot gh(\Lambda) < \lambda_1(\Lambda) < (1+\epsilon)\cdot gh(\Lambda)$ for a randomly chosen lattice Λ of dimension m; and now we replace the tiny coefficient $1/(1-\epsilon)$ by the super huge number 2^m.

7. IDEAL SUBSET PRODUCT WITH ERRORS

Now we move one step more abstract and consider subset products over ideals of a Dedekind domain. The relation between Dedekind domain and UFD is that a Dedekind domain is a UFD if and only if it is a principal ideal domain (PID).

Let R be a Dedekind domain. Let K be a field of fractions of R. Let $\mathcal{F}$ be the group of fractional ideals of R; and $\mathcal{P}$ be the group of principal ideals of R. Let $Cl = \mathcal{F}/\mathcal{P}$ be the ideal class group of K. For a concrete picture, we can think about $R := \mathcal{O}_K$ as the ring of integers of some number field $K/\mathbb{Q}$ or some function field $K/\mathbb{F}_q(X)$.

We define the *ideal subset product problem* (ISP) to be given n ideals $\mathfrak{a}_1,\ldots,\mathfrak{a}_n,\mathfrak{X}$ of R, find a binary vector $(x_1,\ldots,x_n) \in \{0,1\}^n$ such that

$$\prod_{i=1}^n \mathfrak{a}_i^{x_i} = \mathfrak{X}.$$

The *modular ideal subset product problem* (MISP) is given n ideals $\mathfrak{a}_1,\ldots,\mathfrak{a}_n,\mathfrak{X}$ of R, find a binary vector $(x_1,\ldots,x_n) \in \{0,1\}^n$ such that

$$\left(\prod_{i=1}^n \mathfrak{a}_i^{x_i}\right)\cdot\mathcal{P} = \mathfrak{X}\cdot\mathcal{P}.$$

The *ideal subset product with errors problem* (ISPE) is given $n+1$ ideals $\mathfrak{a}_1,\ldots,\mathfrak{a}_n,\mathfrak{X}$ of R and a set $L \subset Spec(R)$ of prime ideals of R such that no ideal in L divides any $\mathfrak{a}_i$, find a binary vector $x = (x_1,\ldots,x_n) \in \{0,1\}^n$ and a square-free ideal $\mathfrak{E}$ factored over L such that

$$\prod_{i=1}^n \mathfrak{a}_i^{x_i}\cdot\mathfrak{E} = \mathfrak{X}.$$

The *modular ideal subset product with errors problem* (MISPE) is given $n+1$ ideals $\mathfrak{a}_1,\ldots,\mathfrak{a}_n,\mathfrak{X}$ of R as well as a set $L \subset Spec(R)$ of prime ideals of R such that no ideal in

L divides any $\mathfrak{a}_i$, find a binary vector $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ and a square-free ideal $\mathfrak{E}$ factored over L such that

$$\left(\prod_{i=1}^{n} \mathfrak{a}_i^{x_i} \cdot \mathfrak{E}\right) \cdot \mathcal{P} = \mathfrak{X} \cdot \mathcal{P}.$$

If the ideals $\mathfrak{a}_i$ are loosen to be fractional ideals in Cl, and the prime ideals $\mathfrak{l}_j$ are loosen to be fractional ideals of the form $\mathfrak{l}_i\mathfrak{t}_i^{-1}$ with $\mathfrak{l}_i$ and $\mathfrak{t}_i$ prime ideals, then we have the *fractional ideal subset product* variants. Their definitions are similar to the ideal problems with the above changes, about which we do not go into details any more. Instead, we give abbreviations to them as FISP, MFISP, FISPE and MFISPE respectively, where FI is short for "fractional ideal". Remind that fractional ideals are not ideals but submodules of R. Hence these variants are problems over a group of submodules.

THEOREM 2. MISPE and MFISPE are generally NP-hard for Dedekind domains R where there is a polynomial time algorithm to compute its ideal multiplication and there is a polynomial time algorithm to find a Gaussian-predictable pair $(Cl(R), P)$.

Proof. The proofs for the reduction chains

$$\text{XC} \leq \text{ISP} \leq \text{ISPE} \leq \text{MISPE}$$

and

$$\text{XC} \leq \text{FISP} \leq \text{FISPE} \leq \text{MFISPE}$$

are analogous to the case of UFD, i.e., Theorem 1. All we need to do is to modify the proof of Theorem 1 as the following: replace the elements a_i by ideals $\mathfrak{a}_i$ (resp. fractional ideals $\mathfrak{a}_i$); replace irreducible elements p_i by prime ideals $\mathfrak{p}_i$ (resp. fractional ideals of the form $\mathfrak{p}_i\mathfrak{q}_i^{-1}$ with $\mathfrak{p}_i$ and $\mathfrak{q}_i$ prime ideals); and replace the irreducible elements ℓ_i by prime ideals $\mathfrak{l}_i$ (resp. fractional ideals of the form $\mathfrak{l}_i\mathfrak{t}_i^{-1}$ with $\mathfrak{l}_i$ and $\mathfrak{t}_i$ prime ideals). □

8. SUBSET PRODUCT FAMILY

In fact, by similar arguments to the proof of Theorem 1, we have proved the following reduction tree, which summarizes the main conclusions of the paper.

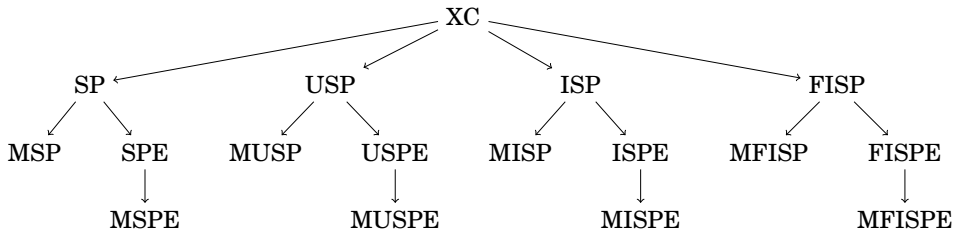

FIGURE 1. Subset Product Family.

A convenient way to memorize the abbreviations of the problems in the tree is to remember the following five rules: (1) the modular problems are those started with an "M"; (2) the errorized problems are those ended with an "E"; (3) "SP" always means "subset product"; (4) "I" always means "ideal"; and (5) "FI" always meand "fractional ideal". Put into words, the tree is the following theorem.

THEOREM 3 (Whole Theorem)**.** The subset product variants in Figure 1 are NP-hard. Specifically, the problems in the SP subtree are NP-hard; the non-modular problems in the USP subtree are generally NP-hard for all UFDs with efficient multiplication; the modular problems in the USP subtree are generally NP-hard for all efficient-Gaussian-predictable UFDs with efficient multiplication; the non-modular problems in the ISP and FISP subtrees are generally NP-hard for all Dedekind domains with efficient ideal multiplication; and the modular problems in the ISP and FISP subtrees are generally NP-hard for all Dedekind domains with efficient ideal multiplication and efficient-Gaussian-predictable ideal class group.

Proof. The four reduction chains from XC to the deepest leaves are Theorem 1 and 2 (note that the chain "XC → SP → SPE → MSPE" in the first subtree is implied by the chain "XC → USP → USPE → MUSPE" in the second subtree, i.e., Theorem 1); and the four remaining reductions SP ≤ MSP, USP ≤ MUSP, ISP ≤ MISP and FISP ≤ MFISP are similar to the reduction from USPE to MUSPE (i.e., Step (3) in the proof of Theorem 1) with the errors $\ell_1, \ldots, \ell_k$ being ignored. □

REFERENCES

[Coo71] Stephen A Cook. "The complexity of theorem-proving procedures". In: *Proceedings of the third annual ACM symposium on Theory of computing*. 1971, pp. 151–158.

[Duc18] Léo Ducas. "Shortest vector from lattice sieving: a few dimensions for free". In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2018, pp. 125–145.

[GJ79] Michael R Garey and David S Johnson. *Computers and intractability*. Vol. 174. freeman San Francisco, 1979.

[GL21] Steven D. Galbraith and Trey Li. "Small Superset and Big Subset Obfuscation". In: *Information Security and Privacy*. Ed. by Joonsang Baek and Sushmita Ruj. Cham: Springer International Publishing, 2021, pp. 68–87. ISBN: 978-3-030-90567-5.

[GZ19] Steven D. Galbraith and Lukas Zobernig. "Obfuscated Fuzzy Hamming Distance and Conjunctions from Subset Product Problems". In: *Theory of Cryptography*. Ed. by Dennis Hofheinz and Alon Rosen. Cham: Springer International Publishing, 2019, pp. 81–110. ISBN: 978-3-030-36030-6.

[HPS08] Jeffrey Hoffstein, Jill Pipher, and Joseph H Silverman. *An introduction to mathematical cryptography*. Vol. 1. Springer, 2008.

[Kar72] Richard M Karp. "Reducibility among combinatorial problems". In: *Complexity of computer computations*. Springer, 1972, pp. 85–103.

[Lev73] Leonid Anatolevich Levin. "Universal sequential search problems". In: *Problemy peredachi informatsii* 9.3 (1973), pp. 115–116.

[LM09] Vadim Lyubashevsky and Daniele Micciancio. "On Bounded Distance Decoding, Unique Shortest Vectors, and the Minimum Distance Problem". In: *Advances in Cryptology - CRYPTO 2009*. Ed. by Shai Halevi. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 577–594. ISBN: 978-3-642-03356-8.

[Sie45] Carl Ludwig Siegel. "A Mean Value Theorem in Geometry of Numbers". In: *Annals of Mathematics* 46.2 (1945), pp. 340–347. ISSN: 0003486X. URL: http://www.jstor.org/stable/1969027 (visited on 06/24/2022).