Generalized Learning Problems and Applications to Non-Commutative Cryptography

Gilbert Baumslag[1], Nelly Fazio[1], Antonio Nicolosi[2], Vladimir Shpilrain[1], and William E. Skeith[1]

[1] The City College of CUNY
gilbert.baumslag@gmail.com, {fazio,wes}@cs.ccny.cuny.edu, shpil@groups.sci.ccny.cuny.edu
[2] Stevens Institute of Technology
nicolosi@cs.stevens.edu

Abstract. We propose a generalization of the *learning parity with noise* (LPN) and *learning with errors* (LWE) problems to an abstract class of group-theoretic learning problems that we term *learning homomorphisms with noise* (LHN). This class of problems contains LPN and LWE as special cases, but is much more general. It allows, for example, instantiations based on non-abelian groups, resulting in a new avenue for the application of combinatorial group theory to the development of cryptographic primitives. We then study a particular instantiation using relatively free groups and construct a symmetric cryptosystem based upon it.

Keywords. Learning with errors. Post-quantum cryptography. Non-commutative cryptography. Burnside groups.

1 Introduction

Motivation. One of the pillars of the modern reductionist approach to cryptography, as exemplified *e.g.,* in [17,18], has been the focus on explicit computational assumptions, precisely phrased in the language of probabilistic modeling. The resulting separation of cryptographic mechanisms from their underlying conjectured-hard problems has been instrumental to the development of a proper formalization of security for disparate cryptographic notions, and for the establishment of connections and elucidation of relations among crypto primitives.

Despite their fundamental role in the theory of cryptography, there is little variety in the family of intractability assumptions. Most of the cryptographic constructs which are used in practice today either rely on a small handful of computational assumptions related to factoring and discrete logs (*e.g.,* RSA, Diffie-Hellman), or lack a well-defined assumption altogether (*e.g.,* AES, and any of the SHA functions). A number of alternatives have surfaced, beginning with elliptic curve cryptosystems [31,27] and more recently with lattice-based constructions [1,2]. Both have turned out to provide revolutionary advances in the theory. Elliptic curves led to the development of identity based cryptosystems [37,11,12], and lattices have recently led to the development of the first fully homomorphic cryptosystems [14,39,15,16].

In this paper, we seek to tap into new sources of computational hardness. Inspired by the recent success of the *learning parity with noise* (LPN [26,10]) and *learning with errors* (LWE [36,30]) problems as a platform for a variety of cryptographic applications, we pursue a generalization of these problems into an abstract class of hard *group-theoretic learning* problems. Besides being of interest in its own right, this generalization opens the way to a new approach for basing cryptography on combinatorial group theory. The rich algebraic structure of non-abelian groups compares favorably with the rigid structure of cyclic groups. Moreover, no efficient quantum algorithms are known for most computational problems in combinatorial group theory, which provides substantial motivation for pursuing this direction of research.

Besides enriching the set of viable intractability assumptions and providing a plausible alternative for post-quantum cryptography, our approach brings into play tools and ideas that have traditionally not found much application in cryptography. For example, in Section 4, we develop an

instantiation of our abstract group-theoretic learning problem based on the theory of groups with exponent k, or *Burnside groups*. We hope that the computational properties of these mathematical objects will spark further work to develop new applications of group theory to cryptography.

A number of attempts to apply combinatorial group theory to cryptography exist in the literature (see below for a survey). Earlier efforts aimed at capitalizing on the algorithmic unsolvability of many of the standard computational problems in combinatorial group theory (*e.g.,* the *word problem*, the *conjugacy problem* and the *membership problem*). These attempts, however, overestimated the relevance of problems that are unsolvable in the *worst-case* for cryptographic purposes. Our approach instead suggests new group-theoretic problems and efficiently sampleable distributions on which it is reasonable to conjecture that these problems remain difficult *on average*.

Non-Commutative Cryptography. In 1984, Wagner and Magyarik [40] proposed the first construction of a group-theoretic asymmetric cryptosystem based on the hardness of the word problem for finitely-presented groups and semigroups. In a nutshell, their idea parallels that of Goldwasser and Micali [19]: rather than distinguishing between quadratic residues and non-residues, the underlying problem is to distinguish two words in a finitely presented group G.

The Wagner-Magyarik cryptosystem has been cryptanalyzed in a number of works, including [21,23,8]. The breakdown in the security was not caused by a weakness of the word problem for groups, but rather it stemmed from a general lack of precision when describing the system and the assumptions on which it was founded. This absence of proper formalization has been characteristic of a number of the early approaches to applying group theory to cryptography [20]. For example, as noted in [8], the description of the protocol in [40] is quite ambiguous, and many design choices were left unspecified. More precisely, the authors failed to provide polynomial time algorithms to generate system parameters (*e.g.,* the group G), as well as the public and private keys, and also failed to provide a complete description of the decryption algorithm. Formal definitions of security were also lacking. When left with this level of ambiguity, formal security analysis is impossible.

A more recent proposal was the work of Anshel *et al.* [4], which can use essentially any non-abelian group as the platform. In their original paper, the authors adopted braid groups. However this choice made the protocol susceptible to various attacks, some of them quite successful (*e.g.,* [13]; see also [33] for a survey).

Perhaps it is not surprising that many of the early attempts to employ non-abelian groups in cryptographic protocols were lacking in precision. The transition from finite abelian groups to non-abelian (possibly infinite) groups for cryptographic purposes is not a small step. Very little is known regarding problems in the theory of non-abelian groups with high average-case complexity, let alone about problems that additionally could support public-key operations. To move a discussion of security to the setting of infinite groups is more difficult still. To begin with, many of the fundamental definitions of security (*e.g.,* [19]) are phrased in terms of probability. Probabilistic analysis for finite groups is readily manageable because the uniform distribution over, say, a finite cyclic group is easy to sample given just a generator and an estimate of the order. For infinite groups, it is even unclear what the corresponding concept of the uniform distribution is, let alone how one goes about sampling it. An attempt toward defining a suitable analogue of the uniform distribution on infinite groups has been recently made by Lee [28], who proposed the notion of *right invariance*, and observed that all previous concrete cryptographic constructions on infinite groups have failed to achieve it.

LPN/LWE. Roughly speaking, the LPN and LWE problems are about learning a certain function by sampling a "noisy" oracle[3] for its input / output behavior. Early research on these problems appears in [10] and [36], respectively. Both problems exhibit attractive self-reducibility properties, giving strong evidence to support the hypothesis that natural randomized versions of these problems are intractable. For LWE, there is more evidence still: the works of [36,35] demonstrate reductions to LWE from worst-case lattice problems. The self-reducibility arguments for these problems are very algebraic, which perhaps suggests that the generalizations we propose may enjoy similar properties when instantiated with other classes of groups. Such a development could produce an exciting new source of problems in group theory which are difficult on average.

Our Contributions. Our main result is the generalization of the *learning parity with noise* (LPN) and *learning with errors* (LWE) problems to an abstract class of learning problems. At high level, we generalize the LWE setting of linear functions over vectors spaces to the context of homomorphisms between groups. This yields conjectured hard problems where the computational task is the recognition of noisy samples of (*preimage*, *image*) pairs for a hidden homomorphism versus random pairs of elements from the relevant domain and codomain.

The resulting abstract class of group-theoretic learning problems contains the LPN and LWE problems as special cases, but is much more general. It allows, for example, instantiations based on non-abelian groups: Another important component of our work is the development of a learning assumption based on free Burnside groups of exponent 3.

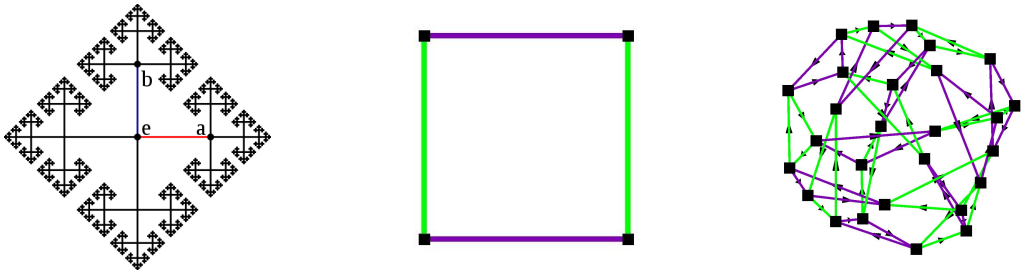
As an application, we propose a symmetric cryptosystem whose provable security can be rigorously analyzed and established based on the conjectured hardness of our Burnside learning problem. This is, to the best of our knowledge, the first time that the computational properties of Burnside groups have been employed for cryptographic purposes.

Organization. Section 2 provides a brief review of basic group-theoretic notions. The proposed generalized learning problem is described in Section 3. Section 4 develops a combinatorial instantiation from free Burnside groups of exponent 3. A symmetric cryptosystem based on Burnside is reported in Section 5.1. Attaining asymmetric encryption is substantially more involved: a possible approach toward this goal is outlined in Section 5.2.

2 Review of Relevant Group-Theoretic Notions

Free groups and presentations. If X is a subset of a group G, let $X^{-1} = \{x^{-1} \mid x \in X\}$. An expression w of the form $a_1 \dots a_n$ $(n \geq 0,\ a_i \in X \cup X^{-1})$ is termed a **word** or an X**-word**. Such an X-word is said to be **reduced** if $n > 0$ and no subword $a_i a_{i+1}$ takes either of the forms xx^{-1} or $x^{-1}x$. If F is a group and X is a subset of F such that X generates F and every reduced X-word is different from 1_F, then one says that F is a **free group**, freely generated by the set X, and refers to X as a **free set** of generators of F, and writes F as $F(X)$. A key property of a free group F freely generated by a set X is that for every group H, every mapping θ from X into H can be extended uniquely to a homomorphism θ_* from F into H. If θ_* is a surjection, and if K is the kernel of θ_*, then the quotient group F/K is isomorphic to H. If R is a subset of F, then in the event that K is generated by all of the conjugates of the elements of R, we express this by writing $H = \langle X; R\rangle$ and term the pair $\langle X; R\rangle$ a **presentation** of H (notice that the mapping θ is usually implicit). The elements of R are termed **relators** of H and R itself is referred to as a set of **defining relators** of H. H is said to be finitely presented if X and R can be chosen to be finite.

[3] Here, "noisy" refers to the fact that the oracle may perturb the correct output according to some random variable whose probability distribution is known.



(a) Free group, $F(a, b)$ (b) $B(2, 2) \cong (\mathbb{F}_2^2, +)$ (c) $B(2, 3) = B_2$

Fig. 1: Cayley graphs for various groups

Notice that if $r \in R$, then $\theta_*(r) = 1$ and we sometimes say that r takes the value 1 in H or that r is a **relation** in H. Of course a given group H can have many different presentations. It is also worth noticing that every pair $\langle X; R\rangle$ consisting of a free set of generators of a free group F and a subset R of F can be made a presentation of a group, namely the factor group $H = F/K$, where K is the **normal closure** in F of R, *i.e.,* the subgroup of F generated by the conjugates of the elements of R. We need only take θ to be the mapping sending elements $x \in X$ to their coset xK.

Relatively Free Groups. If F is a free group and K a normal subgroup of F, then the factor group F/K is called **relatively free** if K is **fully invariant**, *i.e.,* if $\alpha(K) \leq K$ for any endomorphism α of F. If $x_1, \ldots, x_n$ are free generators of F, then $x_1K, \ldots, x_nK$ are called relatively free generators of F/K, and typically denoted simply by $x_1, \ldots, x_n$ when there is no risk of confusion. Let E_n denote a relatively free group of rank n, *i.e.,* $F_n = F(x_1, \ldots, x_n)$ and $E_n = F_n/K$ for some fully invariant K. One key property of such a group is that any set map on its generators into E_n can be extended to an endomorphism of E_n. Hence, one is immediately equipped with an exponential number of homomorphisms, provided that the image is non-trivial.

Cayley distance. Finitely generated groups can also be viewed as geometric objects via the notion of the **Cayley graph**. The Cayley graph of a group G relative to a particular set of generators has the group elements as vertexes, and an edge between two vertexes if and only if multiplication by a generator (or its inverse) translates one to the other. Figure 1 depicts Cayley graphs for few simple groups, including the 27-element *Burnside* group $B(2, 3)$ of exponent 3 with 2 generators. (Burnside groups are discussed in Section 4.) The **Cayley distance** between two group elements is defined as the length of the shortest path between the corresponding nodes in the Cayley graph. The maximum Cayley distance between any two elements in the graph is the **diameter** of the Cayley graph. The **Cayley norm** of an element x, denoted $\|x\|$, is its distance from the identity element in the Cayley graph. We remark that $\max_{x \in G}(\|x\|)$ corresponds precisely to the diameter.

Commutators. In non-abelian groups, the **commutator** of two group elements a, b, denoted $[a, b]$, is the group element satisfying the identity $ab = ba[a, b]$, that is, $[a, b] = a^{-1}b^{-1}ab$. Starting with the generators $x_1, \ldots, x_n$ of the group as the recursive basis, one obtains an ordered sequence of **formal commutators** by combining two formal commutators a, b into the formal commutator $[a, b]$. The **weight** of a formal commutator is defined by assigning weight 1 to the generators, and defining the weight of $[a, b]$ as the sum of the weights of a and b. The weight imposes a partial order on formal commutators, which is typically made total by assuming an arbitrary ordering among formal commutators of any given weight greater than 1, and by adopting the lexicographical order among the generators.

Center. The **center** of a group G, denoted $Z(G)$, is the set of all elements that commute with every element of G.

3 Generalized Learning Problems

We begin by reviewing the learning with errors problem, and we then generalize it to a novel abstract group-theoretic problem concerning learning with respect to a "noisy" oracle.

3.1 Learning with Errors (LPN/LWE)

The problem of *learning from noisy examples* has been considered by Angluin and Laird [3], and subsequently by Kearns [26] and Blum, Kalai, and Wasserman [10]. Informally, the problem is to deduce a particular function by sampling the input / output behavior in the presence of noise (*i.e.,* some of the outputs are incorrect). Of particular interest is the problem of *learning vectors from parity with noise* (LPN) [10], which may be stated as follows. Let Ψ be a distribution on $\mathbb{F}_2$. Let $\mathbf{s}$ and $\{\mathbf{a}_i\}_{i=1}^m$ be randomly chosen vectors, $\mathbf{s}, \mathbf{a}_i \in \mathbb{F}_2^n$, and let $\{e_i\}_{i=1}^m$ be independent samples from Ψ. Define $b_i = \mathbf{s} \cdot \mathbf{a}_i + e_i$ for $i = 1, \ldots, m$, where $\cdot$ denotes the inner product. The problem is then to determine $\mathbf{s}$ given $\{(\mathbf{a}_i, b_i)\}_{i=1}^m$. In general, this problem is believed to be computationally intractable. The best known algorithm is only slightly sub-exponential ($2^{\mathcal{O}(n/\log n)}$, due to [10]).

More generally, one may consider the same problem on vector spaces over finite fields other than $\mathbb{F}_2$. The case of $\mathbb{F}_p$ under zero-mean/low-variance discrete Gaussian noise was considered by Regev and termed *learning with errors* (LWE) problem [36]. Therein, Regev showed a quantum reduction from worst-case lattice problems (*e.g.,* the shortest vector problem), which gives further support to the conjecture that these problems are intractable. When the noise parameter is greater than $\sqrt{n}$, the best known algorithm for solving this problem was demonstrated in [10] and requires $2^{\mathcal{O}(n)}$ time. When the noise parameter is smaller than $\sqrt{n}$, the recent work of [6] has demonstrated a subexponential time algorithm using certain linearization techniques.

We also mention a variant of the LWE problem, recently proposed by Lyubashevsky *et al.* in [30] to improve the ratio of the entropy of the noisy images over that of their preimages. In the setting of [30], termed *ring-LWE*, the noisy samples have the form $(a, b) \in R \times R$, where R is a ring of algebraic integers in a suitable number field, $b \approx a \cdot s$ for a secret random ring element s, and $\cdot$ denotes multiplication in R.

3.2 Learning Homomorphisms with Noise (LHN)

The class of functions at play in the LWE problem is the class $\{\lambda_{\boldsymbol{s}}\}_{\boldsymbol{s} \in \mathbb{F}_p^n}$ of linear functionals from $\mathbb{F}_p^n$ into $\mathbb{F}_p$. By algebraic abstraction, we may replace arbitrary homomorphisms between groups for the linear functionals thus translating the learning problem from the setting of vector spaces to that of arbitrary groups. We describe the resulting generalization below.

For every $n \in \mathbb{Z}^+$, let G_n and P_n be groups (with the operation written multiplicatively). Let Γ_n, Ψ_n, and Ξ_n be distributions on G_n, P_n, and $G_n \times P_n$, respectively. Intuitively, Γ_n determines how preimages are sampled, and will usually be uniform; Ψ_n is the error distribution on the codomain; and Ξ_n is a sort of "base" distribution which is independent of any homomorphism and will also be uniform in most finite cases. Finally, let Φ_n be a distribution on the set $\hom(G_n, P_n)$ of homomorphisms from G_n to P_n. Furthermore, assume that Γ_n, Ψ_n, Ξ_n and Φ_n are efficiently sampleable. Let $\varphi \xleftarrow{\$} \Phi_n$ and define a distribution $A_\varphi^{\Psi_n}$ on $G_n \times P_n$ whose samples are preimage / distorted image pairs (a, b) where $a \xleftarrow{\$} \Gamma_n$ and $b = \varphi(a)e$ for $e \xleftarrow{\$} \Psi_n$. Figure 2 depicts the above generalization.

We now formulate search and decision versions of a general problem which we term *learning homomorphisms with noise* (or for brevity, LHN).

Definition 1 (LHN-Search). *Given an $A_\varphi^{\Psi_n}$-oracle, the* LHN-search problem *is to recover φ.*

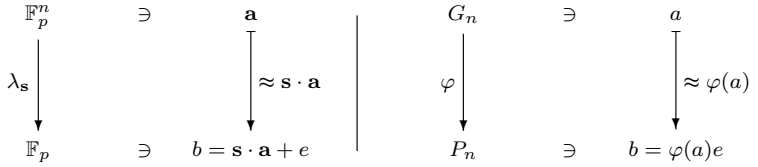


Fig. 2: Generalizing learning problems from vector spaces (LWE, left) to arbitrary groups (LHN, right).

Definition 2 (LHN-Decision). *The* LHN-decision problem *is to distinguish* $A_\varphi^{\Psi_n}$ *from* Ξ_n.

For the search problem, the corresponding assumption is that for all probabilistic polynomial time algorithms W and for every polynomial p, we have:

$$\Pr\left[\varphi' = \varphi \mid \varphi' \leftarrow W^{A_\varphi^{\Psi_n}}(1^n)\right] < \frac{1}{p(n)}$$

where the probability is over the random choices of $\varphi \xleftarrow{\$} \Phi_n$ and over the random coins of the attacker W and of the oracle $A_\varphi^{\Psi_n}$. The corresponding assumption for the decision problem is simply that $A_\varphi^{\Psi_n} \underset{\text{PPT}}{\approx} \Xi_n$.

Note that this is a proper generalization of the standard LWE problem [36], with $G_n = \mathbb{F}_p^n$, $P_n = \mathbb{F}_p$, Φ_n uniform on the linear functionals from $\mathbb{F}_p^n$ into $\mathbb{F}_p$, Γ_n uniform on $\mathbb{F}_p$, Ξ_n uniform over $G_n \times P_n$, and where Ψ_n corresponds to a zero-mean discrete Gaussian over $\mathbb{F}_p$ of suitable variance. At the same time, casting the assumption into abstract terms facilitates the formulation of new learning problems that leverage the potential hardness of group-theoretic settings other than the usual ones of cyclic groups and vector spaces. In particular, we will discuss an instantiation from combinatorial group theory in Section 4.

3.3 Looking for Instantiations of LHN: What Makes LPN/LWE Hard?

To gain insight as to what ingredients are required in the more general context, and to understand what properties one might need of a candidate group-theoretic setting to serve as a platform for the abstract LHN problem, we begin with some general observations on the standard LWE problem. First, we note that part of what makes LWE difficult in the standard vector space case is that $\mathbb{F}_p^n$ is a *free module*. Not only does this afford one with an exponential space of secret keys ($|\hom(\mathbb{F}_p^n, \mathbb{F}_p)| = p^n$); in some sense, it also maximizes the difficulty of learning with errors: Given a single noisy image $\varphi(\mathbf{a}_i) + e_i$, *every* choice of noise e_i produces a value that can be plausibly explained as the true image $\bar{\varphi}(\mathbf{a}_i)$ of some homomorphism $\bar{\varphi}$. Consequently, one must collect many samples in order to rule out any given potential value of the hidden homomorphism φ. Even once enough equations have been obtained to uniquely constrain φ, it is not clear which path to take to distill this large set of equations down to φ, leading to an essentially exponential number of choices to be considered. This is in sharp contrast with the setting of arbitrary finite groups, where $|\hom(G_n, P_n)|$ may not be exponential, and furthermore, one could potentially detect the presence of error from but a single sample (a, b) if, for example, the order of b does not divide the order of a.

From the above discussion, the setting of free groups arises as a seemingly natural alternative to vector spaces. As for the case of $\mathbb{F}_p^n$, instantiating LHN over free groups results in a huge space of possible keys (homomorphisms). Other similarities with vector spaces, however, are not easy to

derive. First, free groups are infinite, which adds non-trivial complications to the sampling process, and makes it cumbersome to even formally state the abstract learning problem in this case. Second, multiplication in free groups is a rather transparent operation. For example, the analogue of the subset sum problem (a crucial ingredient that is often paired with the LPN/LWE assumptions, and used *e.g.,* in the cryptosystem of [36]) admits an efficient algorithm in the setting of free groups (see *e.g.,* [29, Proposition I.2.21]), which makes it rather unsuitable for cryptographic applications.

We contend, however, that suitable analogues of $\mathbb{F}_p^n$ might be found by restricting attention to certain sub-classes of groups, like *relatively free groups*. As mentioned in Section 2, these groups enjoy many of the desirable properties that free groups exhibit: they are, for instance, equipped with exponentially many homomorphisms into any non-trivial group, and thus provide adequate key space for the LHN problem. In contrast to free groups, they can also be chosen to be finite, thus avoiding many of the complications that come with free groups. In the next section, we describe an infinite class of finite relatively free groups: The free Burnside groups of exponent 3.

4 An Instantiation from Combinatorial Group Theory

We now put forth a new intractability assumption by instantiating LHN with a certain class of finite non-abelian groups. We begin with some background and basic facts on the class of groups in question, and then discuss their computational properties and choice of parameters suitable for instantiating the LHN problem.

4.1 Burnside Groups

For a positive integer k, consider the class of groups for which all elements x satisfy $x^k = 1$. Such a group is said to be of *exponent* k. We will be interested in a certain family of such groups called the *free Burnside groups of exponent* k, which are in some sense the "largest." The free Burnside groups are uniquely determined by two parameters: the number of generators n, and the exponent k. We will denote these groups by $B(n, k)$:

Definition 3 (Free Burnside group). *For any $n, k \geq 0$, the* Burnside group of exponent k with n generators *is defined as*

$$B(n,k) = \langle \{x_1, \ldots, x_n\}; \{w^k \mid \textit{ for all words } w \textit{ over } x_1, \ldots, x_n\} \rangle.$$

The question of whether $B(n, k)$ is finite or not is known as the *bounded Burnside problem*. For sufficiently large k, $B(n, k)$ is generally infinite [25]. For small exponents, it is known that $k \in \{2, 3, 4, 6\}$ yields finite groups for all n. (We remark that with the exception of $k = 2$, these are non-trivial results.) For other small values of k (most notably, $k = 5$), the question remains open.

For the purposes of this paper, we will be interested primarily in groups of exponent 3; hence in what follows we will denote $B(n, 3)$ simply by B_n for brevity. Next, we review some important facts about B_n (see [24,22] for a fuller account).

Normal form of B_n. Although B_n is non-abelian, an interesting consequence of the order law $w^3 = 1$ for $w \in B_n$ is that B_n has a simple normal form: Each B_n-element can be written uniquely as an ordered sequence of (a subset of) generators (or their inverses[4]), appearing in lexicographical order, followed by (a subset of) the commutators of weight 2 (or their inverses), and finally by (a

subset of) the commutators of weight 3 (or their inverses):

$$x_1^{\alpha_1} \cdots x_i^{\alpha_i} \cdots x_n^{\alpha_n} [x_1, x_2]^{\beta_{1,2}} \cdots [x_i, x_j]^{\beta_{i,j}} \cdots [x_{n-1}, x_n]^{\beta_{n-1,n}} [x_1, x_2, x_3]^{\gamma_{1,2,3}}$$
$$\cdots [x_i, x_j, x_k]^{\gamma_{i,j,k}} \cdots [x_{n-2}, x_{n-1}, x_n]^{\gamma_{n-2,n-1,n}} = \prod_{i=1}^{n} x_i^{\alpha_i} \prod_{i<j} [x_i, x_j]^{\beta_{i,j}} \prod_{i<j<k} [x_i, x_j, x_k]^{\gamma_{i,j,k}}$$

where all $\alpha_i, \beta_{i,j}, \gamma_{i,j,k} \in \{0, 1, -1\}$ for all $1 \leq i < j < k \leq n$, and $[x_i, x_j, x_k] = [[x_i, x_j], x_k]$.

ORDER OF B_n. From the above normal form, it follows that B_n has exactly $3^{n+\binom{n}{2}+\binom{n}{3}}$ elements.

HOMOMORPHISMS FROM B_n TO B_r. There are $3^{n\left(r+\binom{r}{2}+\binom{r}{3}\right)}$ homomorphisms from $B_n \to B_r$. This follows immediately from the order of B_r and from the fact that B_n is a free object in the category of groups of exponent 3 with generating set of size n.

We also have the following lemma regarding the diameter of B_n:

Lemma 1. *$\exists \tau_n \in B_n$ such that $\|\tau_n\| \in \Omega(\frac{n^3}{\log n})$.*

Proof. Let $d_n = \max_{x \in B_n}(\|x\|)$, and recall that $|B_n| = 3^{n+\binom{n}{2}+\binom{n}{3}}$. Since all elements of the group can be written with at most d_n symbols taken from $x_1^{\pm 1}, \ldots, x_n^{\pm 1}$:

$$(2n)^{d_n} \geq 3^{n+\binom{n}{2}+\binom{n}{3}} \iff d_n \log_3(2n) \geq n + \binom{n}{2} + \binom{n}{3} \iff d_n \geq \left\lceil \frac{n + \binom{n}{2} + \binom{n}{3}}{\log_3 2n} \right\rceil \qquad \square$$

4.2 Computational Aspects of Burnside Groups

In order for the Burnside groups to be of use in cryptography, at a minimum, they must have a concise representation, and the group operation must be efficiently computable. We demonstrate here that both criteria are met. First, we note that as described above, each element of B_n has a unique normal form as a product of the generators and certain commutators. Hence by storing an array of the exponents (each of which is in the set $\{0, 1, -1\}$) we can uniquely represent an element. The size of the array is cubic in n.

As for the group operation, this can be computed simply by concatenating two normal forms, and then reducing the resulting word back into normal form. This process, referred to as the *collection process*, takes cubic time (see [24], chap. 11) in the length of the input (which is itself cubic in n). However, all commutators of weight 3 are in the center $Z(B_n)$ of B_n, and hence there is no need to expand them and apply the collection process—one can simply add the corresponding exponents modulo 3. Furthermore, since all commutators of weight 4 are trivial (see [24], chap. 18), we know that $[B_n, B_n]$ is commutative. Hence, we can again avoid the collection process when moving the weight-2 commutators amongst themselves, and in cubic time, we can reduce the expression to a "nearly" normal form consisting of a product of at most $2n$ generators (or their inverses) followed by commutators in normal form. Therefore we need only to apply the collection process on linear input, and so the overall running time of computing the product is indeed $\mathcal{O}(n^3)$. Inverses can also be computed over B_n in at most cubic time by a similar (yet somewhat simpler) collecting process.

The last and most challenging computational aspect of B_n relates to its *geodesics*—the computation of distances in the Cayley graph. For the applications we introduce here, it will suffice to compute the *norm* (*i.e.,* the distance to the identity of the group).

[4] Note that $x^{-1} = x^2$ in B_n, as B_n has exponent 3.

In general, geodesics in the Cayley graph is a difficult problem. In some cases, it is known to be NP-hard [32].[5] However, this is not as troubling as it seems. We need only to compute norms in the codomain group P_n, which is generally small, and does not necessarily grow with the security parameter (although it may grow with a correctness parameter). For the case of the free Burnside group B_r, one possible solution is to perform a breadth-first search of the Cayley graph, storing the norm of every element in a table. This process will begin to become infeasible around $r = 5$. However, even with this small number of generators, the diameter is large enough to properly decode for many interesting error distributions Ψ_n. For the general case, geodesics in the Cayley graph of B_n might be efficiently computable (perhaps up to small approximation factors) making use of a number of commutator identities, for instance:

$$[a, b] = [b, a]^{-1};\ a[a, b] = [a, b]a;\ [a, b]^{-1} = [a^{-1}, b];\ [a, b, c]^{-1} = [a^{-1}, b, c].$$

This might allow one to either shorten the normal forms, or to first re-order the generators and then notice that the resulting words are shorter. We shall not concern ourselves with this here, but will consider this problem separately.

4.3 Instantiating LHN over Burnside Groups

Here we propose a concrete instantiation from Burnside groups, which we subsequently denote by B_n-LHN. Set $G_n \doteq B_n$ and $P_n \doteq B_r$, where $2 \leq r \leq 4$. Let $\Gamma_n \doteq \mathbf{U}(B_n)$ and $\Xi_n \doteq \mathbf{U}(B_n \times B_r)$. The error distribution Ψ_n on B_r is constructed by taking a randomly ordered product of the generators, raised to random exponents. More precisely, its probability mass function is:

$$\forall e \in B_r, \qquad \Pr_{E \xleftarrow{\$} \Psi_n}[E = e] = \Pr_{\mathbf{v} \xleftarrow{\$} \mathbb{F}_3^r, \sigma \xleftarrow{\$} S_r}\left[e = \prod_{i=1}^{r} x_{\sigma(i)}^{v_i}\right] \tag{1}$$

where the x_i's are the generators of B_r, the v_i's are the components of $\mathbf{v}$, and S_r denotes the symmetric group on r letters. Since $x^2 = x^{-1}$ in B_r, the norm $\|e\|$ of a Ψ_n-sample e is at most r. (Some intuition for this choice of Ψ_n is discussed at the end of this Section.) For any given secret homomorphism φ, the above choices completely describe the distribution $A_\varphi^{\Psi_n}$. As for the distribution Φ_n from which φ is drawn, we simply let $\Phi_n \doteq \mathbf{U}(\hom(B_n, B_r))$. Note that since B_n is a relatively free group, any mapping of its n generators uniquely extends to a homomorphism. Hence, to sample Φ_n, it suffices to select random B_r-images for the n generators of B_n. Note that for $2 \leq r \leq 4$, elements of B_r take at most 3 bytes, and thus storing φ requires just linear space.

Figure 3 summarizes the choice of groups and distributions for the B_n-LHN problem.

G_n	P_n	Γ_n	Ξ_n	Ψ_n	Φ_n
B_n	B_r	$\mathbf{U}(B_n)$	$\mathbf{U}(B_n \times B_r)$	$\left[\mathbf{v} \xleftarrow{\$} \mathbf{U}(\mathbb{F}_3^r),\ \sigma \xleftarrow{\$} S_r : \prod_{i=1}^{r} x_{\sigma(i)}^{v_i}\right]$	$\mathbf{U}(\hom(B_n, B_r))$

Fig. 3: Choice of groups and distributions for the B_n-LHN problem.

Choice of Parameters. To determine suitable choice of parameters for the B_n-LHN instantiation described above, here we consider known approaches to attacking the assumption. First, observe that the key space is rather large: $|\hom(B_n, B_r)| = 3^{\Theta(nr^3)}$, and so even small choices of n and r

[5] One entertaining example is that of the Rubik's cube group, whose diameter was demonstrated to be 20 in 2010 via a distributed computing project which required 35 CPU-years.

will defeat a brute-force attack. In terms of a distinguishing attack, we derive below an interesting connection to LWE with $p = 3$, based on the projection onto the commutator-factor (*cf.* Figure 4). Computationally, ρ_n amounts to just retaining the exponent-tuple corresponding to the generators

$$\rho_n : B_n \to B_n/[B_n, B_n] \cong (\mathbb{F}_3^n, +)$$
$$\rho_r : B_r \to B_r/[B_r, B_r] \cong (\mathbb{F}_3^r, +)$$

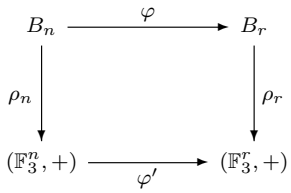


Fig. 4: Action over B_n and B_r of the projection onto the commutator-factor.

in the normal form of a B_n-element. One easily verifies that ρ_n and ρ_r transform the distribution $A_{\varphi}^{\Psi_n}$ from an B_n-LHN instance to a new distribution, $A_{\varphi'}^{\Psi'_n}$ over $\mathbb{F}_3^n \times \mathbb{F}_3^r$, which presents us with a problem very similar to the standard LWE with $p = 3$. (Even with $r > 1$, the resulting problem is polynomial-time equivalent to the standard version; see *e.g.,* [5], Lemma 4.2.) Notably, the resulting noise distribution Ψ'_n for the LWE-like instance is just the abelianization of Ψ_n, which by construction amounts to a random r-tuple of $\mathbb{F}_3$-exponents (*cf.* Equation (1)). Thus, applying the commutator-factor transformation yields an LWE-like instance where the noisy distribution is *identical* to the random one, and so the instance is impossible to break. Nevertheless, in light of this connection with LWE, it seems prudent to pick values for n that would also make LWE hard. The best algorithm for this setting is currently the one of [10], and requires time $2^{\mathcal{O}(n/\log n)}$, which suggests values of n in the few hundreds.

Remark 1. Regarding the error distribution, we remark that the support of Ψ_n should never be contained in a proper normal subgroup of P_n, else an adversary will always be able to "factor out" the noise to mount a distinguishing attack. Note that for the standard LWE/LPN problems, as well as the error distribution we propose for B_n-LHN in Equation (1), this issue does not arise because the support of Ψ_n generates all of P_n.

Regarding the choice of r, as discussed above, $r = 4$ will suffice, as this permits an exhaustive, breadth first search of the Cayley graph. For each element of B_r, a geodesic representative (or just the norm) can be stored in a moderately-sized table ($\approx$ 14MB) for future use. We stress that this computation needs only to be done once for the lifetime of the system.

We note also that there is still much flexibility in the choice of Ψ_n; random walks of variable length, perhaps according to a similar distribution as that of [9], may also be appropriate. Additionally, we remark that one may consider altogether different metrics on the group, *e.g.,* taking a normal form for the elements and then using the Hamming metric on the resulting vector of exponents. The distribution Ψ_n could then correspond to explicitly corrupting part of the description of $\varphi(a)$. However, the former approach using the Cayley graph seems to have much more promise for application to an asymmetric setting—we discuss this further in the following section.

5 Applications

5.1 A Group-Based Symmetric Cryptosystem

In this section, we present a symmetric cryptosystem based on the hardness of learning Burnside homomorphisms with noise (B_n-*LHN cf.* Section 4.3).

Precomputation: Run breadth-first search on the Cayley graph of B_r, recording the norm of each element. We stress that this procedure need only be done once for the lifetime of the system.

Key-Gen(n): Run the setup algorithm for B_n-LHN to select a random homomorphism φ from the set of homomorphisms from B_n into B_r, and set the shared key $\mathsf{SK} \doteq \varphi$. Using the table generated in the precomputation phase, select an element $\tau \in B_r$ of maximal norm. Given the lower bound from Lemma 1, we know that $\|\tau\| = \Omega(r^3/\log r)$.

Encrypt(SK, t): To encrypt a bit t, select $(a, b) \xleftarrow{\$} A_\varphi^{\Psi_n}$, compute $b' \doteq b\tau^t (= \varphi(a)e\tau^t)$, and output the ciphertext $c \doteq (a, b')$.

Decrypt($\mathsf{SK}, (a, b')$): Compute $e' = \varphi(a)^{-1} \cdot b'$ and output $t = 0$ if and only if $\|e'\| \leq r$.

Theorem 1 (Correctness). *If $(a, b') \xleftarrow{\$}$ Encrypt(SK, t), then Decrypt($\mathsf{SK}, (a, b')$) $= t$.*

Proof. First note that for any group G, the norm in the Cayley metric is well-behaved with respect to the group product: in particular we have that for all $a, b \in G$,

$$|\|a\| - \|b\|| \leq \|ab\| \leq \|a\| + \|b\| .$$

Combining this fact with Lemma 1, we see that as r grows, correctness follows easily, since for any $e \xleftarrow{\$} \Psi_n$, we have $\|e\| \leq r$, and $\|\tau\| \in \Omega(r^3/\log_3(2r))$ so that $\|e\tau\| \geq \|\tau\| - \|e\| \gneq r$. □

Remark 2. For the case of small r, we must take more care. Note that from the proof of Lemma 1, we have more precisely that $\|\tau\| \geq \left\lceil \frac{r+\binom{r}{2}+\binom{r}{3}}{\log_3 2r} \right\rceil$. Hence if $r = 4$, then our lower bound for $\|\tau\|$ is 8, which presents a small problem, since the maximal norm element from the support of Ψ_n is of norm 4. Such an element will be sampled from Ψ_n with probability $\frac{16}{81}$, and hence in this case, we simply remark that correctness can be amplified by sending multiple encryptions. The Decrypt algorithm will then output 0 if $\left\|\varphi(a)^{-1} \cdot y\right\|$ is ever less than 4, 1 if it is ever greater, and $\perp$ if it is always 4. We also remark that the elementary lower bounds from Lemma 1 are likely not tight, in which case there is no need for the amplification. Even for $r = 4$, if $\|\tau\| = 9$ rather than 8, the scheme above would attain correctness with probability 1, making amplification unnecessary.

Theorem 2 (Security). *If the B_n-LHN-Decision problem is hard, then the above cryptosystem is IND-CPA secure.*

Proof. We structure the proof following the sequence-of-game approach of [38]. We define a sequence of "indistinguishable" games G_0, G_1, G_2, and G_3, all operating over the same underlying probability space. Starting from the actual adversarial game G_0, we make incremental modifications to the behavior of the encryption and challenge oracles. This changes the way in which the adversary's view is computed, but the changes are controlled so as to ensure that the views' distributions across games are indistinguishable. Additionally, the last game (game G_3) is defined so that the adversary's goal is clearly impossible; by the indistinguishability of any pair of consecutive games, it will follow that the adversary's advantage in the original game is negligible.

Game G_0. Game G_0 models the standard notion of symmetric-key IND-CPA [7] security for the case of bit encryption. On input the security parameter 1^n, the adversary $\mathcal{A}$ interacts with an encryption oracle $\mathcal{O}_{\mathrm{ENC}}(\cdot)$ and with a challenge oracle $\mathcal{O}_{\mathrm{CH}}()$. The oracles (reported in Figure 5) run with the same random secret key φ, sampled according to the B_n-LHN-Decision instance generator from Section 4.3. At the end of the interaction, $\mathcal{A}$ outputs a bit t' as her best guess to the bit t^* drawn by the challenge oracle in step C_0.

$E_1.\ a_i \xleftarrow{\$} B_n$
$E_2.\ e_i \xleftarrow{\$} \Psi_n$
$E_3.\ b_i \leftarrow \varphi(a_i)e_i$
$E_4.\ d_i \leftarrow b_i\tau^{t_i}$
return(a_i, d_i)

(a) $\mathcal{O}_{\text{ENC}}(t_i)$

$C_0.\ t^* \xleftarrow{\$} \{0,1\}$
$C_1.\ a^* \xleftarrow{\$} B_n$
$C_2.\ e^* \xleftarrow{\$} \Psi_n$
$C_3.\ b^* \leftarrow \varphi(a^*)e^*$
$C_4.\ d^* \leftarrow b^*\tau^{t^*}$
return(a^*, d^*)

(b) $\mathcal{O}_{\text{CH}}()$

Fig. 5: Oracles in Game G_0.

Game G_1. Game G_1 is identical to game G_0 except for a notational change: steps E_1–E_3 and C_1–C_3 are replaced with:

$$E'_{123}.\ (a_i, b_i) \xleftarrow{\$} A_\varphi^{\Psi_n} \qquad C'_{123}.\ (a^*, b^*) \xleftarrow{\$} A_\varphi^{\Psi_n}$$

Let S_1 be the event that $t' = t^*$ at the end of game G_1. By definition of $A_\varphi^{\Psi_n}$, it is clear that the difference between games G_0 and G_1 is just syntactic, and so $\boxed{\Pr[S_0] = \Pr[S_1]}$.

Game G_2. In game G_2, the distribution *Adist* is replaced with $\Xi_n \doteq \mathbf{U}(G_n \times P_n)$:

$$E''_{123}.\ (a_i, b_i) \xleftarrow{\$} \Xi_n \qquad C''_{123}.\ (a^*, b^*) \xleftarrow{\$} \Xi_n$$

In Lemma 2, games G_1 and G_2 are proven indistinguishable under the B_n-LHN-Decision assumption. In particular, denoting with S_2 the event that $t' = t^*$ at the end of game G_2, and with $\nu_{B_n\text{-}LHN}$ the maximum success probability against the B_n-LHN-Decision problem by any probabilistic polynomial time distinguisher W, it holds that $\boxed{|\Pr[S_1] - \Pr[S_2]| \leq \nu_{B_n\text{-}LHN}}$.

Game G_3. Game G_3 is identical to game G_3 except that step C_4 is replaced with:

$$C'''_4.\ d^* \leftarrow b^*$$

Let S_3 be the event that $t' = t^*$ at the end of game G_3. We claim that events S_2 and S_3 are equivalent, and so $\boxed{\Pr[S_2] = \Pr[S_3]}$. To see this, note that the only difference between the two games is how d^* is computed. Since in game G_3 d^* is uniform and independent of anything else in the adversary's view, it suffices that this is also the case in game G_2. Indeed, for any fixed $c \in B_r$:

$$d^* = c\,[\text{in } G_2] \quad \Leftrightarrow \quad b^*\tau^{t^*} = c \quad \Leftrightarrow \quad b^* = c(\tau^{t^*})^{-1}$$

Thus,

$$\Pr_{G_2}[d^* = c] = \Pr_{G_2}[b^* = c(\tau^{t^*})^{-1}] = \frac{1}{|B_r|}$$

because b^* is uniform in B_r.

Next, observe that $\boxed{\Pr[S_3] = 1/2}$: Indeed, $\mathcal{A}$'s view in game G_3 is computed independently of the bit t^*. Putting it altogether we get:

$$|\Pr[S_0 - 1/2]| = |\Pr[S_0 - S_3]| = |\Pr[S_1 - S_2]| \leq \nu_{B_n\text{-}LHN}.$$

□

Lemma 2. $\Pr[S_1] - \Pr[S_2] \leq \nu_{B_n\text{-}LHN}$.

Proof. We describe a reduction from the B_n-LHN-decision problem to distinguishing between games G_1 and G_2. In other words, given an IND-CPA adversary $\mathcal{A}$ against our scheme, we construct a distinguisher W that, given oracle access to a distribution $\mathcal{D} \in \{A_\varphi^{\Psi_n}, \Xi_n\}$, uses $\mathcal{A}$ to discriminate between the two cases.

Distinguisher W runs a copy of $\mathcal{A}$ internally. Whenever $\mathcal{A}$ asks for a ciphertext on $t_i \in \{0,1\}$, W obtains from its oracle $\mathcal{D}$ a sample (a_i, b_i), and returns (a_i, d_i) to $\mathcal{A}$, where $d_i \doteq b_i \tau^{t_i}$. When $\mathcal{A}$ asks for the challenge ciphertext, W picks a random bit t^*, obtains a sample (a^*, b^*) from its oracle $\mathcal{D}$, computes $d^* \doteq b^* \tau^{t^*}$, and returns (a^*, d^*) as challenge to $\mathcal{A}$. Eventually, $\mathcal{A}$ will stop and output a bit t'. At that point, if $t' = t^*$, then W will output 1, guessing that $\mathcal{D}$ was in fact running as $A_\varphi^{\Psi_n}$. Otherwise, W will output 0, guessing that $\mathcal{D}$ was instead Ξ_n.

Clearly, W is roughly as efficient as $\mathcal{A}$. Furthermore, W essentially "interpolates" between games G_1 and G_2, in the sense that:

$$\Pr[W^{A_\varphi^{\Psi_n}}(1^n) = 1] = \Pr[S_1] \qquad \Pr[W^{\Xi_n}(1^n) = 1] = \Pr[S_2]$$

It follows that:

$$|\Pr[S_1] - \Pr[S_2]| = |\Pr[W^{A_\varphi^{\Psi_n}}(1^n) = 1] - \Pr[W^{\Xi_n}(1^n) = 1]| \leq \nu_{B_n\text{-}LHN}$$

□

5.2 Towards Group-Based Asymmetric Cryptosystems

There are several remaining obstacles to basing asymmetric cryptography on B_n-LHN. The primary issue is in providing a means of sampling the distribution $A_\varphi^{\Psi_n}$ without knowledge of the secret φ. In cryptosystems like that of [36], this was accomplished via computing the sum over a random subset of known samples from the distribution. However, note that commutativity seems critical for this to be effective:[6] if $\{(a_i, b_i)\}_{i=1}^m$ are samples (so $b_i = \varphi(a_i) + e_i$, where e_i are "small") and $S \subset [m]$, then $\sum_{i \in S} b_i = \sum_{i \in S} (\varphi(a_i) + e_i) = \varphi\left(\sum_{i \in S} a_i\right) + \sum_{i \in S} e_i$. It follows that, if $|S|$ is not too large, $\sum_{i \in S} b_i$ will remain close to the true image $\varphi(\sum_{i \in S} a_i)$. In the non-abelian case, $\prod(\varphi(a_i)e_i)$ is not generally equal to $\prod \varphi(a_i) \prod e_i$, and so it is not necessarily true that $\prod(\varphi(a_i)e_i)$ remains close to $\varphi(\prod a_i)$ just because the norm of $\prod e_i$ is small.

We briefly mention some possible approaches toward bypassing this issue. A first workaround might be to consider only abelian P_n. However, this makes the problem somewhat less interesting, since applying the factor-commutator transformation to $A_\varphi^{\Psi_n}$ would then produce a new distribution over abelian groups which likely is not any more difficult to distinguish from uniform as the original (assuming that $\Gamma_n = \mathbf{U}(G_n)$ and $\Xi_n = \mathbf{U}(G_n \times P_n)$ as usual).[7] So it would seem that to consider only abelian P_n is to rule out non-abelian groups altogether.

[6] We adopt below additive notation, as it is more natural for the LPN/LWE setting.

[7] This follows primarily from the fact that for any epimorphism $\psi : G \to P$ of groups, $\psi(\mathbf{U}(G)) = \mathbf{U}(P)$, but also requires the assumption that the commutator subgroup $[G_n, G_n]$ can be efficiently sampled, or that some other means exist for sampling the fibers of the projection $G_n \to G_n/[G_n, G_n]$.

A more promising approach might be to place additional constraints on the distribution Ψ_n. By careful selection of the error terms, one might be able to guarantee that the resulting product behaves well in the sense that commutators involving $e \xleftarrow{\$} \Psi_n$ are small in comparison to the diameter of P_n. However, we remark that the naïve method of forcing the support of Ψ_n to be contained in $Z(P_n)$ is flawed: the commutator-factor transformation then produces a distribution without noise, which will typically be easy to distinguish from random via standard linear algebra techniques. More generally, as discussed in Section 4, the support of Ψ_n should never be contained in a proper normal subgroup of P_n. Note that in our instantiation of B_n-LHN using free Burnside groups, the support of Ψ_n generates all of P_n.

6 Conclusions and Future Work

In this paper, we put forth a generalization of the learning parity with noise and learning with errors problems, moving from linear functionals over vector spaces to homomorphisms between arbitrary (possibly non-abelian) groups. We also developed an instantiation of our abstract group-theoretic learning problem from the theory of Burnside groups, and proposed the first cryptographic applications of these groups in the form of a symmetric cryptosystem.

Our work broadens the family of cryptographically useful intractability assumptions. It also raises several research questions, ranging from specific issues like estimating the most suitable choice of parameters, to broader problems like devising alternate instantiations of our learning problem. Other related lines of inquiry to be investigated in future work include: 1) applying the techniques of [34] to extend the symmetric scheme to efficiently encrypt multiple bits; 2) adapting our Burnside-based cryptosystem to the asymmetric setting; 3) improving existing algorithms for computing over Burnside groups (*e.g.,* to compute the Cayley norm); and 4) assessing the hardness of learning homomorphisms with noise over Burnside groups by designing sub-exponential distinguishing attacks.

Acknowledgement. We are grateful to Hugo Krawczyk for suggesting a cleaner acronym for our generalized learning assumption.

References

1. M. Ajtai. Generating hard instances of lattice problems (extended abstract). In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 99–108. ACM, 1996.
2. M. Ajtai and C. Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *STOC '97*, pages 284–293, 1997.
3. D. Angluin and P. Laird. Learning from noisy examples. *Machine Learning*, 2(4):343–370, 1988.
4. I. Anshel, M. Anshel, and D. Goldfeld. Non-abelian key agreement protocols. *Discrete Applied Mathematics*, 130(1):3–12, 2003.
5. Benny Applebaum, David Cash, Chris Peikert, and Amit Sahai. Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In *CRYPTO*, pages 595–618, 2009.
6. Sanjeev Arora and Rong Ge. New algorithms for learning in presence of errors. Manuscript, 2011.
7. M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science—FOCS '97*, pages 394–403, 1997.
8. J.C. Birget, S.S. Magliveras, and M. Sramka. On public-key cryptosystems based on combinatorial group theory. *Tatra Mountains Mathematical Publications*, 33:137–148, 2006.
9. A. Blass and Y. Gurevich. Matrix transformation is complete for the average case. *SIAM Journal on Computing*, 24(1):3–29, 1995.

10. A. Blum, A. Kalai, and H. Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *J. ACM*, 50:2003, 2003.
11. D. Boneh and M. Franklin. Identity-based encryption from the weil pairing. *SIAM J. of Computing*, 32(3):586–615, 2003.
12. C. Cocks. An identity-based encryption scheme based on quadratic residuosity. In *Cryptology and Coding*, pages 360–363, Heidelberg, 2001. Springer. LNCS 2260.
13. D. Garber, S. Kaplan, M. Teicher, B. Tsaban, and U. Vishne. Probabilistic solutions of equations in the braid group. *Advances in Applied Mathematics*, 35:323–334, 2005.
14. C. Gentry. Fully homomorphic encryption using ideal lattices. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 169–178, New York, NY, USA, 2009. ACM.
15. C. Gentry. Toward basing fully homomorphic encryption on worst-case hardness. In *CRYPTO*, pages 116–137, 2010.
16. C. Gentry, S. Halevi, and V. Vaikuntanathan. i-hop homomorphic encryption and rerandomizable Yao circuits. In *CRYPTO*, pages 155–172, 2010.
17. O. Goldreich. *Foundations of Cryptography, vol. 1.* Cambridge Univ. Press, 2001.
18. O. Goldreich. *Foundations of Cryptography, vol. 2.* Cambridge Univ. Press, 2004.
19. S. Goldwasser and S. Micali. Probabilistic encryption. *JCSS*, 28(2):270–299, 1984.
20. M.I. Gonzalez-Vasco, S. Magliveras, and R. Steinwandt. *Group Theoretic Cryptography.* Chapman and Hall/CRC, United States, 2012. To appear.
21. M.I. Gonzalez-Vasco and R. Steinwandt. Reaction attacks on public key cryptosystems based on the word problem. *Applicable Algebra in Engineering, Communication and Computing*, 14(5):335–340, 2002.
22. N. Gupta. On groups in which every element has finite order. *Amer. Math. Month.*, 96:297–308, 1989.
23. C. Hall, I. Goldberg, and B. Schneier. Reaction attacks against several public-key cryptosystem. In *In Proc. of ICICS'99, LNCS*, pages 2–12. Springer-Verlag, 1997.
24. M. Hall. *The Theory of Groups.* Macmillan Company, New York, 1959.
25. Sergei V. Ivanov. The free Burnside groups of sufficiently large exponents. *Internat. J. Algebra Comput.*, 4(1-2):ii+308, 1994.
26. M. Kearns. Efficient noise-tolerant learning from statistical queries. In *Journal of the ACM*, pages 392–401. ACM Press, 1993.
27. N. Koblitz. Elliptic curve cryptosystems. *Mathematics of Computation*, 48(177):203–209, 1987.
28. E. Lee. Right-invariance: A property for probabilistic analysis of cryptography based on infinite groups. In *ASIACRYPT*, pages 103–118, 2004.
29. R. Lyndon and P. Schupp. *Combinatorial Group Theory.* Classics in Mathematics. Springer, 2001.
30. Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. In *EUROCRYPT*, pages 1–23, 2010.
31. V. Miller. Use of elliptic curves in cryptography. In *Advances in Cryptology—Crypto '85*, pages 417–426, New York, 1985. Springer. LNCS 218.
32. A. Myasnikov, V. Roman'kov, A. Ushakov, and A. Vershik. The word and geodesic problems in free solvable groups. *Trans. Amer. Math. Soc.*, 362:4655–4682, 2010.
33. A. Myasnikov, V. Shpilrain, and A. Ushakov. *Group-Based Cryptography.* Birkhäuser Verlag, Switzerland, 2008.
34. C. Peikert, V. Vaikuntanathan, and B. Waters. A framework for efficient and composable oblivious transfer. In *CRYPTO*, pages 554–571, 2008.
35. Chris Peikert. Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In *STOC*, pages 333–342, 2009.
36. O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In *STOC*, pages 84–93. ACM Press, 2005.
37. A. Shamir. Identity based cryptosystems and signatures schemes. In *Advances in Cryptology—Crypto '84*, pages 47–53, Heidelberg, 1984. Springer. LNCS 196.
38. V. Shoup. Sequences of games: A tool for taming complexity in security proofs, 2004. Manuscript. Available at `http://www.shoup.net/papers/games.pdf`.
39. M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan. Fully homomorphic encryption over the integers. In *EUROCRYPT*, pages 24–43, 2010.
40. N. Wagner and M. Magyarik. A public key cryptosystem based on the word problem. In *Proceedings of CRYPTO 84 on Advances in cryptology*, pages 19–36, New York, NY, USA, 1985. Springer-Verlag New York, Inc.