Left-Right Cayley Hashing: A New Framework for Provably Secure Hash Functions

Yusuke Aikawa[1], Hyungrok Jo[2], and Shohei Satake[3]

[1] The Graduate School of Information Science and Technology, The University of Tokyo
[2] Institute of Advanced Sciences, Yokohama National University
[3] Research and Education Institute for Semiconductors and Informatics, Kumamoto University

Received: 22nd June 2023 | Accepted: 15th July 2023

Abstract *Cayley hash function is a significant cryptographic primitive known to be provably secure and expected to be universal due to the expansion property of Cayley graphs on finite groups. In this paper we propose a new Cayley-type hash function based on certain non-backtracking walks on an left-right Cayley complex. Then we discuss its collision-resistance and prove the universality, provided that the underlying complex is constructed from high-girth Cayley graph expanders. We also present instantiations of the proposed hash functions by using known explicit Cayley graph expanders on special linear groups over finite fields.*

Keywords: Left-right Cayley hash functions, Left-right Cayley complexes, Cayley graphs, Expander graphs
2010 Mathematics Subject Classification: 94A60, 05C48, 20F10

1 INTRODUCTION

Hash functions are ubiquitous in various fields, especially for cryptographic schemes as substantial gadgets; ensuring the integrity of data, constructing digital signature schemes, encryption schemes, message authenticated codes and so on. For details see e.g. [4]. The standardized hash functions, namely SHA family, currently in use are practical and expected to be sufficiently secure, while SHA-1 has been broken [39]. It is hopeful that the security of a cryptographic primitive is reduced to the hardness of some mathematical problems because we can analyze the security of such a primitive by investigating the mathematical problem. However, the security of the standardized hash functions is not related to the hardness of some mathematical problems. On the other hand, CGL hash function [10] is known to be one of the provably secure hash functions, which is constructed by random walks on supersingular isogeny graphs. With respect to the efficiency, CGL hash function is very slow because of iteration of computing isogenies. So building hash functions having both provable security and practical efficiency seems to be a difficult problem.

For this problem, one of the important research directions is to build *group-theoretic hash functions*, which are initially proposed by Zémor [40]. The novelty of this proposal is to build hash functions from non-backtracking random walk on high-girth Cayley graph expanders. The idea of the construction is as follows: Let G be a finite group and $A = \{g_0, g_1, \ldots, g_{d-1}\} \subset G$ a subset of order d. And let Γ be the Cayley graph defined by $Cay(G, A)$. Then we define a function $H : [d]^* \to G$, where $[d]^* = \cup_{k=1}^{\infty} \{0, 1, \ldots, d-1\}^k$ by the following procedure. An input message is written as a string $m = m_1 m_2 \ldots m_k$ where $m_i \in \{0, 1, \ldots, d-1\}$. Then the output is computed by the group product $g_{m_k} \ldots g_{m_2} g_{m_1} \in G$. This function compresses a string with arbitrary length into a string fixed length. We call a hash function constructed in this way a *Cayley hash function*. As described below, Cayley hash functions have a number of interesting properties.

The security of these functions can be reduced to the hardness of group-theoretic problems. For example, detecting collisions is equivalent to find cycles on a Cayley graph. Finding cycles on a Cayley graph is restated into several group-theoretic problems, so-called Balance Problem and Representation Problem on a group. So a Cayley graph with high-girth may provide a hash function resistant to collision search. Note that these group-theoretic problems used in the context of the security of a Cayley hash function are considered to be resistant to cryptanalysis by quantum computers. Moreover, the uniform distribution of the outputs is expected by the expanding property on Cayley graphs. So if the underlying Cayley graph is an expander graph, the outputs of the function tend to be uniformly distributed rapidly as the length of inputs tends to be long. Since Zémor used the special linear group as a platform group, computation of a function is done efficiently by computing products of matrices. Moreover, we can compute functions in parallel due to its malleability. Scalability in outputs size also enhances convenience for use of these schemes. Therefore, after Zémor's proposal, various Cayley hash functions are suggested, although very specific cases have already been broken.

*Corresponding Author: aikawa@mist.i.u-tokyo.ac.jp, jo-hyungrok-xz@ynu.ac.jp, shohei-satake@kumamoto-u.ac.jp

As above, Cayley hash functions have favorable properties. However, certain parameters of Cayley hash functions are susceptible to known fatal attacks, which poses a significant concern as discussed below. Despite recent advancements in research on parameter selection to withstand these attacks, the secure construction of hash functions based on Cayley graphs is still a work in progress. Consequently, the question arises: Can we devise a methodology that enables the development of robust Cayley-type hash functions while preserving the advantages offered by Cayley hash functions?

Our Contribution In this paper, we propose a new framework for building group-theoretic hash functions called *left-right Cayley hash functions* (*LR-Cayley hash functions*, for short) based on *left-right Cayley complexes* (*LR-Cayley complexes*, for short). LR-Cayley complexes are recently introduced by Dinur et al. [15] who explicitly constructed locally testable codes with constant code rate, relative distance, and locality, which resolves a long-standing open problem in theoretical computer science. A crucial point to define an LR-Cayley complex over a finite (non-abelian) group G is to observe that G has two natural actions to itself, say left-action and right-action. Accordingly distinguishing these two actions one can obtain two types of Cayley graphs, left- and right-Cayley graphs. Then a LR-Cayley complex on G is defined as a 2-dimensional complex with the skeleton graph that is the union of left- and right-Cayley graphs on the same group G, where each 2-dimensional face is realized as a square with one pair of parallel edges are in the left-Cayley graph and the another is in the right-Cayley graph, see Section 2.2 for details. Our proposal comes from a similar idea of Cayley hash function, but non-backtracking walks defining the hashing process are chosen on an LR-Cayley *complex* (not a single Cayley graph).

It should be remarked that an LR-Cayley hash function is expected to be more collision-resistant than the Cayley hash function based on the Cayley graph on the same underlying group. One reason is that the collision-resistance of an LR-Cayley hash function would closely rely on the hardness of finding *two distinct* factorizations of identity *simultaneously* in the underlying group, while the hardness of a (single) factorization of identity is a basis of the collision-resistance of the original Cayley hash function.

Also we prove that the LR-Cayley hash function is universal provided that the underlying left- and right-Cayley graphs are expanders. The proof is based on introducing a new random walk, called *left-right random walk* (*LR-RW*), on the LR-Cayley complex and careful analysis of the mixing time of LR-RW via the spectral properties of the underlying left- and right-Cayley graphs of the complex.

As instantiations we provide several instances of LR-Cayley hash functions. Note that Cayley graphs on any abelian group cannot form expander families [1] and hence it is required to choose non-abelian groups. One nice candidate of such groups for instantiating LR-Cayley hash functions is the special linear group $\mathrm{SL}_n(\mathbb{F}_p)$ over the finite field. Indeed random Cayley graphs on $\mathrm{SL}_n(\mathbb{F}_p)$ are high-girth expanders with high probability (e.g. [5], [7]), and furthermore some explicit high-girth Cayley graph expanders on $\mathrm{SL}_n(\mathbb{F}_p)$ are presented in the literature (see e.g. [24], [22], [23], [21], [5], [6], [8], [3], [34]) whereas it is known to be very difficult to construct high-girth Cayley graph expanders over a finite group in general. In the security aspect there is no known efficient algorithm factorizing identity in $\mathrm{SL}_n(\mathbb{F}_p)$ (except for very limited cases), which provides us a certain confidence to believe that the implemented LR-Cayley hash functions are collision-resistant. Our instances from $\mathrm{SL}_n(\mathbb{F}_p)$ are in fact extensions of the Cayley hash functions designed by Tillich-Zémor [37] and Coz et al. [13], and it is expected that our instances could be more collision-resistant than the original ones.

Therefore, our contributions are summarized as follows:

1. We show a recipe for constructing hash functions from LR-Cayley complexes, which we call LR-Cayley hash function.
2. We introduce new group-theoretic problems, we call the Simultaneous Balance Problem and Simultaneous Representation Problem. Then we reduce the security of LR-Cayley hash functions to the hardness of these problems. In other words, we prove that LR-Cayley hash functions are provably secure.
3. We discuss the mixing property of LR-Cayley complexes. By giving an estimation on mixing time, we prove that LR-Cayley hash functions have universality, which means that the distribution of the outputs of LR-Cayley hash functions approaches the uniform distribution rapidly.
4. By using results on constructing high-girth Cayley graph expanders [5], [7], [3], we give instantiations of LR-Cayley hash functions and discuss their security.

Previous works Since the seminal work of Zémor [40], Cayley hash functions have attracted the interest because of several favorable properties, e.g. provably secure, efficiency, universality, scalability, detection of small modifications of massages and so on. In the first instance of Cayley hash functions by Zémor [40], the special linear group $\mathrm{SL}_2(\mathbb{F}_p)$ is used as a platform group with the following generator: $g_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $g_2 = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. Although this scheme was soon broken [37], Tillich and Zémor proposed another instance by using a Cayley graph of $\mathrm{SL}_2(\mathbb{F}_{2^n})$ with generators $g_1 = \left(\begin{smallmatrix} \alpha & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $g_2 = \left(\begin{smallmatrix} \alpha & \alpha+1 \\ 1 & 1 \end{smallmatrix}\right)$, where α is a root of the polynomial defining $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$. This scheme is

now called the Zémor-Tillich hash function (ZT hash function, for short). In addition to being provably secure, this proposal has practical computational efficiency [32], [14]. Moreover, the notable feature of the ZT hash function is that the distribution of the outputs rapidly approaches the uniform distribution due to an expansion property of Cayley graphs of the special linear group, which is confirmed experimentally [20], [27]. A hash function having such a property is called universal.

Zémor's idea of building hash functions from random walks on a graph paved the way for further research. One of the subsequent notable studies is the construction of hash function from Ramanujan graphs [10], the supersingular isogeny graphs [33] and the LPS graphs [23], by Charles, Lauter and Goren. After that an analogue of the latter construction was proposed in [29] by using the Morgenstern graph [25], although both are fully broken in [28] by Petit, Lauter and Quisquater. Moreover, an analogue of the LPS hash function based on the Chiu graph [12], which is a cubic case of the LPS graphs, has been analyzed in [19]. On the other hand, the former opened the door to the area of isogeny-based cryptography using supersingular elliptic curves and is still considered to be sufficiently secure. Moreover, Castryck, Decru and Smith investigate a genus 2 analogue using superspecial abelian surfaces of the hash function based on isogeny graphs in [9].

Cryptanalysis for the ZT-hash function has also made progress. Firstly, in [11], [35], attacks restricted to the ZT hash functions with specific defining polynomials were proposed. The general attack on the ZT hash function was proposed in [17], but this method is considered to be not practical. A decisive attack on the collision-resistance of the ZT hash function was made by Grassl et al. [18]. This attack was soon extended to a preimage attack in [30]. However, it should be noted that these attacks are against special generators of Cayley graphs on $\mathrm{SL}_2(\mathbb{F}_{2^n})$. Although studies on general attacks on Cayley hash functions have been progressed [26], [27], they are not practical for a sufficiently large finite field. So the above attacks on the ZT hash function are very limited. Indeed, recently new parameters for Cayley hash functions were proposed by Bromberg, Shipilrain and Vdovina [8], which is considered to be robust to previous attack due to choice of generators as a monoid. Moreover, a method using the larger group $\mathrm{GL}_2(\mathbb{F}_{p^n})$, instead of $\mathrm{SL}_2(\mathbb{F}_{p^n})$ has also been proposed by Tomkins, Nevins and Salmasion [38]. This proposal offers the possibility of wider instantiation, but some research issues remain, such as expansion properties of Cayley graphs of $\mathrm{GL}_n(\mathbb{F}_{p^n})$, computational analysis of the efficiency.

2 PRELIMINARIES

Notation For a positive integer $d \in \mathbb{Z}$, we set $[d] = \{0, 1, \ldots, d-1\}$. For a finite set S, $s \leftarrow_R S$ denote the process of uniformly sampling elements from S. For a graph Γ let $V(\Gamma)$ and $E(\Gamma)$ denote its vertex set and edge set, respectively. For a group G and its subset A let $A^{-1} := \{a^{-1} \mid a \in A\}$. For a prime power q let $\mathbb{F}_q$ denote the finite field of order q. The *special linear group* $\mathrm{SL}_n(\mathbb{F}_p)$ *of rank* n over the finite field $\mathbb{F}_p$ is a multiplicative matrix group consisting of all $n \times n$ matrices with elements from $\mathbb{F}_p$ and determinant 1 (hence its identity is the identity matrix I_n).

2.1 HASH FUNCTIONS

Hash functions are functions that compresses a string of arbitrary length into a shorter fixed-length:

$$H : [d]^* \to \{0,1\}^s,$$

where $\ell \in \mathbb{N}$ and we set $[d]^* = \cup_{k=1}^{\infty} [d]^k$. Often written $H : \{0,1\}^* \to \{0,1\}^s$, considering the case $d = 2$. Scalability of hash functions means that we can control the length of the output of functions. In this case, we can construct a family of hash functions: $\mathcal{H} := \{H_i : [d]^* \to \{0,1\}^{s(i)}\}$, where $s(i)$ is a function with respect to i.

Let $H : [d]^* \to \{0,1\}^s$ be a hash function. The pair (x, x') of distinct inputs of H is said to be a collision pair of H if we have $H(x) = H(x')$. In cryptographic use, hash functions are typically required to meet the following strong conditions.

Definition 1 (Collision-resistance)**.** *Let d and s be positive integers. A hash function $H : [d]^* \to \{0,1\}^s$ is said to be collision resistant if finding a collision pair (x, x') of H is a computationally infeasible task. Moreover, the hash function H is provably secure if we can prove that H is collision resistant under the assumption that solving a mathematical problem is to be hard.*

The term "computationally infeasible" means that no probabilistic algorithm successfully solves the task in polynomial time with respect to s. Among various proposals for hash functions, there are some construction methods that prioritize computational efficiency. However, in mathematical cryptology and its applications, it is worth investigating provable security, as it establishes valuable connections between cryptography and mathematics. Therefore, in this paper, we concentrate on provably secure collision-resistant hash functions.

Moreover, the outputs of a hash function should be unbiased, so we introduce a definition of the asymptotic behavior of the distribution of the output of hash functions.

Definition 2 (Universality)**.** *Let d and s be positive integers. We call a hash function $H : [d]^* \to \{0,1\}^s$ ϵ-universal if the distribution of the outputs of H approaches the uniformly distribution on $\{0,1\}^s$ with respect to the error time ϵ. More precisely, we define random variables $\{X_n\}_{n\in\mathbb{N}}$ where X_n denotes the outputs of the function H with inputs of n-bits string $x \in [d]^n$. Then the hash function H is ϵ-universal if for $\epsilon \in \mathbb{R}$, there exists a positive integer n_ϵ such that for any $n \geq n_\epsilon$ and outputs $y \in \{0,1\}^s$, we have*

$$\left|\Pr[X_n = y] - \frac{1}{2^\ell}\right| < \epsilon.$$

In this paper, we aim to construct provably secure universal hash functions based on group-theoretic techniques. That is, we focus on hash functions whose security is ensured from the hardness of group-theoretic problems, and universality follows from the fact that graphs based on some groups carry expander families.

2.2 CAYLEY GRAPHS AND LEFT-RIGHT CAYLEY COMPLEXES

We introduce the definitions of Cayley graphs and left-right Cayley complexes.

Definition 3 (Left- and right-Cayley graphs)**.** *Let G be a finite group and let A and B be subsets of $G \setminus \{1\}$ and suppose that $A^{-1} = A$ and $B^{-1} = B$. Then the* left-Cayley graph $Cay_L(G, A)$ *is a graph with vertex set G and edge set consisting of $\{g, ag\}$ with $g \in G$ and $a \in A$. Similarly let B be a subset of G and suppose that $B^{-1} = B$. Then the* right-Cayley graph $Cay_R(G, B)$ *is a graph with vertex set G and edge set consisting of $\{g, gb\}$ with $g \in G$ and $b \in B$.*

For the sake of simplicity we use the notation $Cay(G, S)$ to mean the Cayley graph on G with respect to a subset $S \subset G \setminus \{1\}$ when we do not have to specify whether it is a left- or right-Cayley graph.

Remark 1. *Notice that $Cay_L(G, A)$ and $Cay_R(G, B)$ are well-defined undirected graphs since $A^{-1} = A$ and $B^{-1} = B$. Also $Cay_L(G, A)$ (resp. $Cay_R(G, B)$) is a $|A|$-regular (resp. $|B|$-regular) graph.*

Remark 2. *The reason why $Cay_L(G, A)$ (resp. $Cay_R(G, B)$) is called a left-Cayley graph (resp. right-Cayley graph) is that it is invariant (as a graph) under the left- (resp. right-) action of G to itself.*

Remark 3. *Throughout this paper we deal with non-abelian groups to distinguish left- and right-Cayley graphs since if G is abelian then $Cay_L(G, A)$ is nothing but $Cay_R(G, A)$.*

The left-right Cayley complexes are recently introduced by Dinur et al. [15] who explicitly constructed locally testable codes with constant code rate, relative distance, and locality. Before introducing the definition, we define a equivalence relation on $A \times G \times B$, where, again, A and B are subsets of $G \setminus \{1\}$ that $A^{-1} = A$ and $B^{-1} = B$. For $(a, g, b) \in A \times G \times B$ define an equivalence relation denoted by $\sim$ as

$$(a, g, b) \sim (a^{-1}, ag, b) \sim (a^{-1}, agb, b^{-1}) \sim (a, gb, b^{-1}).$$

Accordingly let $[a, g, b]$ denote the equivalence class of (a, g, b), which consists of the four elements in $A \times G \times B$. Note that $[a, g, b]$ can be realized as a square with 4 vertices g, ag, gb, agb and two pairs of parallel edges, one consists of edges in $Cay_L(G, A)$ connected by $a, a^{-1} \in A$ and the another consists of edges in $Cay_R(G, B)$ connected by $b, b^{-1} \in B$, see Figure 1.

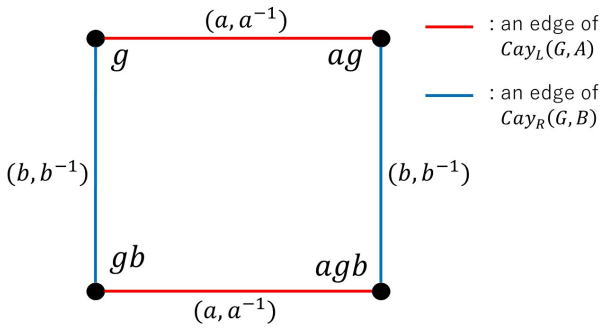

Figure 1: A square corresponding to $[a, g, b]$

Definition 4 (Left-right Cayley complex)**.** *Let G be a finite group and let A and B be subsets of $G \setminus \{1\}$ and suppose that $A^{-1} = A$ and $B^{-1} = B$. Then define the* left-right Cayley complex (LR-Cayley complex) $Cay^2(G, A, B)$ *as a 2-dimensional complex with vertex set G, edge set $E(Cay_L(G, A)) \cup E(Cay_R(G, B))$ and the set of squares $\{[a, g, b] \mid (a, g, b) \in A \times G \times B\}$.*

2.3 EXPANDER GRAPHS

For $d \geq 3$ let Γ be a d-regular graph. Let $A(\Gamma)$ denote the adjacency matrix of Γ and $\lambda(\Gamma)$ denote the largest eigenvalue in modulus other than d.

Definition 5. *Let $0 < \lambda < d$ be a constant. A d-regular graph Γ is said to be a* λ-expander *if $\lambda(\Gamma) \leq \lambda$.*

For a fixed $d \geq 3$ we say an infinite family $\{\Gamma_i\}_{i\geq 1}$ of d-regular graphs with n_i vertices is an *λ-expander* if $\lambda(\Gamma_i) \leq \lambda$ for every $i \geq 1$. By Cheeger's inequality a (sufficiently large) d-regular expander is a highly-connected sparse graph. Expanders are extensively studied in mathematics and theoretical computer science. In particular it is well-known that random walk on expanders exhibits several nice and elegant behaviors, in particular, rapid mixing. More details of mixing are explained in the next subsection.

2.4 MIXING

We introduce the definition of non-backtracking random walks on regular graphs and their mixing time.

Definition 6 (Non-backtracking random walk, NBRW)**.** *Let $d \geq 3$ and $\Gamma = (V, E)$ be a non-bipartite connected d-regular n-vertex graph and $u \in V$ be any fixed vertex. A* non-backtracking random walk (NBRW) *of length k on Γ is a finite Markov chain $(Y_0, Y_1, \ldots, Y_k)$, where Y_i is a random variable such that*

- *$Y_0 \in \{(u, x) \mid ux \in E\}$, and for each x with $ux \in E$ $\Pr[Y_0 = (u, x)] = 1/d$,*
- *if $i \geq 1$ and $Y_{i-1} = (y, v)$ with $yv \in E$ then we have*

$$Y_i \in \{(v, w_1), (v, w_2), \ldots, (v, w_{d-1}) \mid \text{for all } 1 \leq j \leq d-1,\ vw_j \in E,\ w_j \neq y\},$$

and for each j with $vw_j \in E$ and $w_j \neq y$ $\Pr[Y_i = (v, w_j)] = 1/(d-1)$.

It is known that for the Markov chain corresponding to non-backtracking random walk on a connected non-bipartite regular graph converges to the uniform distribution on its vertex set, which this phenomenon is called *mixing*. Hence one can define the mixing time of non-backtracking random walk.

Definition 7 (Mixing time)**.** *Let $\tilde{P}^{(k)}_{v,u}$ be the transition probability of non-backtracking random walks of length k on Γ starting from u and ending in v. Then the mixing time $\tilde{\tau}(\Gamma)$ (w.r.t. L^∞-distance) is defined as follows.*

$$\tilde{\tau}(\Gamma) := \min\left\{t \in \mathbb{N} \mid \forall k \geq t \max_{u,v\in V}\left|\tilde{P}^{(k)}_{v,u} - \frac{1}{n}\right| \leq \frac{1}{n^2}\right\}.$$

It is known that the mixing time of NBRW on expanders can be estimated as follows, which plays a key role in the analysis of universality of the proposed hash functions, see Section 3.3.

Theorem 1 ([2])**.** *For a fixed $d \geq 3$ suppose that an infinite family $\{\Gamma_i\}_{i\geq 1}$ of d-regular graphs is a λ-expander for some $0 < \lambda < d$ Let n_i be the number of vertices of Γ_i. Then we have*

$$\tilde{\tau}(\Gamma_i) = O(\log n_i).$$

Note that this upper bound is optimal up to constant.

3 A HASH FUNCTION BASED ON A LEFT-RIGHT CAYLEY COMPLEX

In this section, we propose a hash function based on left-right Cayley graphs. Moreover, we discuss the collision-resistance and the distribution of the outputs of the hash functions. In discussion on the distribution, we define left-right random walks on the graphs and prove expansion properties.

3.1 CONSTRUCTION

We introduce the following hash function as a candidate of cryptographic hash functions. First for a finite group G and its subsets $A, B \subset G$, we say the *total no-conjugacy condition* (*TNC*) holds if for any $h \in G$, $a \in A$ and $b \in B$, $h^{-1}ah \neq b$. This property prevents trivial collisions occurring by a certain backtrack (see Remark 7).

Parameters Let G be a finite group and fix an element $g \in G$ where $g \neq 1$. For an integer $d \geq 2$, suppose that each of $(d+1)$-element-subsets $A, B \subset G$ generates G and TNC holds for G, A and B. We fix an element $b_0 \in B$. We decide labelling maps $\pi_A : [d] \times A \to A$, $\pi_B : [d] \times B \to B$ so that for any $a \in A$ and $b \in B$ we have $\pi_A([d] \times \{a\}) = A \setminus \{a\}$ and $\pi_B([d] \times \{b\}) = B \setminus \{b\}$. For each integer $\ell \geq 2$, we decide how to divide ℓ into two positive integers ℓ_L and ℓ_R, that is, $\ell_L + \ell_R = \ell$ as parameters. For example, $\ell_L = \lfloor \ell/2 \rfloor$ and $\ell_R = \lceil \ell/2 \rceil$.

Hashing process Then for a given d-ary text $t \in [d]^\ell$, we divide t into the string $t_1t_2\dots t_{\ell_L}t'_1t'_2\dots t'_{\ell_R}$ where $t_i, t'_j \in [d]$ according to the above parameter. At last for each $1 \le i \le \ell_L$ and $1 \le j \le \ell_R - 1$ set

$$a_i = \pi_A(t_i, a_{i-1}) \in A, \quad b_j = \pi_B(t'_j, b_{j-1}) \in B$$

where define $a_0 = g$. Its hash value is

$$a_{\ell_L} \cdots a_2a_1 \cdot g \cdot b_1b_2 \cdots b_{\ell_R}$$

We call this hash function the *left-right Cayley hash function with respect to G, A and B*, denoted by $H_{G,A,B}$.

Example 1. *Let $d = 3$ and for a given $\ell \ge 2$ set $\ell_L = \lfloor \ell/2 \rfloor$ and $\ell_R = \lceil \ell/2 \rceil$ in the construction above. A ternary text $t = 012 \in [3]^3$ is divided into the string $t_1t'_1t'_2$ with $t_1 = 0$, $t'_1 = 1$ and $t'_2 = 2$. Then $a_1 = \pi_A(0, g)$, $b_1 = \pi_B(1, b_0)$ and $b_2 = \pi_B(2, b_1)$. Hence we have*

$$H_{G,A,B}(012) = a_1 \cdot g \cdot b_1b_2.$$

A ternary text $t = 2102 \in [3]^*$ of length 4 is divided into the string $t_1t_2t'_1t'_2$ with $t_1 = 2$, $t_2 = 1$, $t'_1 = 0$ and $t'_2 = 2$. Then $a_1 = \pi_A(2, g)$, $a_2 = \pi_A(1, a_1)$, $b_1 = \pi_B(0, b_0)$ and $b_2 = \pi_B(2, b_1)$ and thus

$$H_{G,A,B}(2102) = a_2a_1 \cdot g \cdot b_1b_2.$$

Remark 4. *Since A and B are generating sets of G, the hash value space is the group G.*

Remark 5. *Each hash value is determined by choosing two non-backtracking walks on vertices of the LR-Cayley complex $Cay^2(G, A, B)$. First choose a walk from g of length ℓ_L on edges in $Cay_L(G, A)$ according to $t_1t_2\dots t_{\ell_L}$. Then choose another walk of length ℓ_R from the terminal vertex of the first walk according to $t'_1t'_2\dots t'_{\ell_R}$. The second walk is on edges in $Cay_R(G, B)$. As a result the hash value is the terminal vertex of the second walk. This walk will rigorously be defined later (see Definition 8).*

Remark 6. *In the construction above two maps π_A and π_B have the role of choosing direction in each step of the first and second walks, respectively. Note that the definitions of π_A and π_B is for the sake of forbidding backtracks.*

Remark 7. *TNC prevents a backtrack occurring when $a_{\ell_L} \cdots a_2a_1 \cdot g \cdot b_1 = a_{\ell_L-1} \cdots a_2a_1 \cdot g$ which is equivalent to $h^{-1}a_{\ell_L}h = b_1^{-1}$ where $h = a_{\ell_L-1} \cdots a_2a_1g$.*

Remark 8. *Our proposed hash function does not allow the malleability, which means*

$$H_{G,A,B}(m_1||m_2) \ne H_{G,A,B}(m_1) \cdot g^{-1} \cdot H_{G,A,B}(m_2),$$

unlikely those of usual Cayley hash functions.

3.2 SECURITY : COLLISION-RESISTANCE

Recall that a collision pair (shortly, a collision) is a pair of two distinct texts with same hash value. Below we discuss the collision-resistance of LR-Cayley hash functions. Recall that for a finite group G and its $(d+1)$-element subset $S \subset G \setminus \{1\}$ a *word* of length ℓ is a product of ℓ elements in S. We say a word $s_1s_2 \cdots s_\ell$ is *reduced* if $s_{i+1} \ne s_i^{-1}$ for all $1 \le i \le \ell - 1$.

Let $t_1t_2\dots t_{\ell_L}t'_1t'_2\dots t'_{\ell_R} \in \{0, 1, \dots, d\}^\ell$ and $u_1u_2\dots u_{m_L}u'_1u'_2\dots u'_{m_R} \in \{0, 1, \dots, d\}^m$ be two distinct (divided) texts. Similarly as a_i and b_j, for each $1 \le i \le m_L$ and $1 \le j \le m_R$ let $a'_i = \pi_A(u_i, a'_{i-1})$, $b'_j = \pi_B(u'_j, b'_{j-1})$ where $a'_0 = g$ and $b'_0 = b_0$. Then these form a collision of $H_{G,A,B}$ if and only if

$$a_{\ell_L} \cdots a_2a_1 \cdot g \cdot b_1b_2 \cdots b_{\ell_R} = a'_{m_L} \cdots a'_2a'_1 \cdot g \cdot b'_1b'_2 \cdots b'_{m_R}. \tag{1}$$

It can easily verfied that the equation (1) holds provided that

$$\begin{cases} a_{\ell_L} \cdots a_2a_1 = a'_{m_L} \cdots a'_2a'_1 \\ b_1b_2 \cdots b_{\ell_R} = b'_1b'_2 \cdots b'_{m_R} \end{cases} \tag{2}$$

In other words, to find a collision of the left-right Cayley hash function $H_{G,A,B}$, it suffices to solve the following problem (see also Remark 9).

Problem 1 (Simultaneous Balance Problem)**.** *For given positive integers ℓ_L, ℓ_R, m_L and m_R find four reduced words satisfying the following system.*

$$\begin{cases} \alpha_{\ell_L} \cdots \alpha_2 \alpha_1 = \alpha'_{m_L} \cdots \alpha'_2 \alpha'_1 \\ \beta_1 \beta_2 \cdots \beta_{\ell_R} = \beta'_1 \beta'_2 \cdots \beta'_{m_R} \end{cases} \tag{3}$$

where $\alpha_i, \alpha'_{i'} \in A$ and $\beta_j, \beta'_{j'} \in B$ denote the variables of the words.

Each equation in (3) is a particular case of the *balance problem*. Note that the collision-resistance of Cayley hash functions relies on its hardness.

One can check that since A and B are inverse-closed, the system (2) is equivalent to

$$\begin{cases} \alpha_1 \alpha_2 \cdots \alpha_{\ell_L + m_L} = 1 \\ \beta_1 \beta_2 \cdots \beta_{\ell_R + m_R} = 1 \end{cases}$$

where $\alpha_i \in A, \beta_j \in B$ denote variables of the words, and both of $\alpha_1 \alpha_2 \cdots \alpha_{\ell_L + m_L}$ and $\beta_1 \beta_2 \cdots \beta_{\ell_R + m_R}$ are reduced. Thus a solution of the following problem forms a collision of $H_{G,A,B}$ as well.

Problem 2 (Simultaneous Representation Problem)**.** *For given positive integers n_L and n_R find two reduced words of length less than or equal to n_L and n_R, respectively, such that*

$$\begin{cases} \alpha_1 \alpha_2 \cdots \alpha_{n_L} = 1 \\ \beta_1 \beta_2 \cdots \beta_{n_R} = 1 \end{cases} \tag{4}$$

where $\alpha_i \in A$ and $\beta_j \in B$ denote the variables of the words.

Note that each equation in (4) is a particular case of the *representation problem* which its hardness is a reason to believe the collision-resistance of Cayley hash functions (based on undirected Cayley graphs).

Remark 9. *It should be emphasized that the collision-resistance of $H_{G,A,B}$ are related to Problems 1 and 2 which both require to simultaneously solve the balance or representation problem with respect to two disjoint generating sets A and B. Hence, it may be reasonable to expect that finding a collision of $H_{G,A,B}$ would in general be harder than the Cayley hash function corresponding to a single Cayley graph, say $Cay_L(G, A)$ or $Cay_R(G, B)$.*

On the other hand, Problems 1 and 2 provide sufficient *conditions for finding collisions, whereas at least we are not aware of any other good reformulation of* (1) *to find collisions efficiently.*

In addition it is possible to discuss the collision-resistance of $H_{G,A,B}$ based on the girth of $Cay_L(G, A)$ and $Cay_R(G, B)$. For a graph Γ let $\mathrm{girth}(\Gamma)$ denote the girth of Γ, that is, the length of the shortest cycle in Γ. Then we have the following propositions, which immediately follow from the construction of LR-Cayley hash functions and the equation (1).

Proposition 1. *For $d \geq 2$ we have the followings.*

(1) *Distinct two d-ary texts $t_1 t_2 \ldots t_{\ell_L} t'_1 t'_2 \ldots t'_{\ell_R}$ and $t_1 t_2 \ldots t_{\ell_L} u'_1 u'_2 \ldots u'_{\ell_R}$ of the same length never form a collision provided that $\ell_R < \mathrm{girth}(Cay_R(G, B))/2$.*

(2) *Distinct two d-ary texts $t_1 t_2 \ldots t_{\ell_L} t'_1 t'_2 \ldots t'_{\ell_R}$ and $u_1 u_2 \ldots u_{\ell_L} t'_1 t'_2 \ldots t'_{\ell_R}$ of the same length never form a collision provided that $\ell_L < \mathrm{girth}(Cay_L(G, A))/2$.*

Proposition 2. *We have the following claims.*

(1) *Problem 1 has no solutions provided that $\ell_L + m_L < \mathrm{girth}(Cay_L(G, A))$ and $\ell_R + m_R < \mathrm{girth}(Cay_R(G, B))$.*

(2) *Problem 2 has no solutions provided that $n_L < \mathrm{girth}(Cay_L(G, A))$ and $n_R < \mathrm{girth}(Cay_R(G, B))$.*

Remark 10. *By Proposition 1, we can also protect against small modifications of texts (of certain type, at least). In other words, any small modification through $H_{G,A,B}$ deduce to change its hash value. (i.e. The small modification property)*

3.3 UNIVERSALITY : MIXING OF RANDOM WALK ON AN LR-CAYLEY COMPLEX

In this subsection we aim to prove the following theorem.

Theorem 2. *Let G be a finite group and $A, B \subset G \setminus \{1\}$ be equally-sized subsets of constant size (as $|G| \to \infty$) that TNC holds. Assume that $Cay_L(G, A)$ and $Cay_R(G, B)$ are λ-expanders for some $0 < \lambda < |A| = |B|$. Then the LR-Cayley hash function $H_{G,A,B}$ is $1/|G|^2$-universal with respect to any text of length at least $\Omega(\log |G|)$ (that is, $n_{1/|G|^2} = O(\log |G|)$ in the term of Definition 2).*

To prove the theorem we first introduce the following random walk on an LR-Cayley complex which is crucially related to the distribution of hash values of the LR-Cayley hash function.

Definition 8 (Left-right random walk, LR-RW)**.** *Let $Cay_L(G,A)$ and $Cay_R(G,B)$ be a left-Cayley graph and a right-Cayley graph, respectively, on the same finite group G. Assume that TNC holds for G, A and B. Let $g \in G$ be a fixed element. Then a* (non-backtracking) left-right random walk (LR-RW) *of length-type (ℓ_L, ℓ_R) is a finite Markov chain $(\mathcal{L}_0, \mathcal{L}_1, \dots, \mathcal{L}_{\ell_L-1}, \mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_{\ell_R-1})$, where $\mathcal{L}_i$ and $\mathcal{R}_i$ are random variables such that*

- *$\mathcal{L}_0 \in \{(g, ag) \mid a \in A\}$, and for each $a \in A$ $\Pr[\mathcal{L}_0 = (g, ag)] = 1/|A|$,*
- *if $1 \le i \le \ell_L - 1$ and $\mathcal{L}_{i-1} = (x, ax)$ with $a \in A$ then we have*

$$\mathcal{L}_i \in \{(ax, a'ax) \mid a' \in A, a' \ne a^{-1}\},$$

  and for each $a' \in A$ with $a' \ne a^{-1}$ we have $\Pr[\mathcal{L}_i = (ax, a'ax)] = 1/(|A| - 1)$,
- *if the terminal vertex of $\mathcal{L}_{\ell_L-1}$ is h', then $\mathcal{R}_0 \in \{(h', h'b) \mid b \in B\}$, and for each $b \in B$ we have $\Pr[\mathcal{R}_0 = (h', h'b)] = 1/|B|$ (note that for every $b \in B$ no backtracks occurs by TNC),*
- *if $1 \le i \le \ell_R - 1$ and $\mathcal{R}_{i-1} = (y, yb)$ with $b \in B$ then we have*

$$\mathcal{R}_i \in \{(yb, ybb') \mid b' \in B, b' \ne b^{-1}\},$$

  and for each $b' \in B$ with $b' \ne b^{-1}$ we have $\Pr[\mathcal{R}_i = (yb, ybb')] = 1/(|B| - 1)$.

Also let $\tilde{Q}_{h,g}^{(\ell_L,\ell_R)}$ be the probability of LR-RW of length-type (ℓ_L, ℓ_R) starting from g and ending in h. Then we define the mixing time $\tilde{\mu}(G; A, B)$ (w.r.t. L^∞-distance) as follows:

$$\tilde{\mu}(G; A, B) := \min\left\{ t \in \mathbb{N} \mid \forall \ell_L, \ell_R \ge 1 \text{ s.t. } \ell_L + \ell_R = t, \max_{g,h \in G} \left| \tilde{Q}_{h,g}^{(\ell_L,\ell_R)} - \frac{1}{|G|} \right| \le \frac{1}{|G|^2} \right\}.$$

Intuitively LR-RW is just a walk obtained by connecting NBRWs on $Cay_L(G,A)$ and $Cay_R(G,B)$ where the walk on $Cay_L(G,A)$ precedes the one on $Cay_R(G,B)$. Hence recalling Remark 5 the mixing of LR-RW immediately implies the universality of the LR-Cayley hash function.

By Definition 2 Theorem 2 immediately follows from the following proposition.

Proposition 3. *Let G be a finite group and $A, B \subset G \setminus \{1\}$ be equally-sized subsets of constant size (as $|G| \to \infty$) that TNC holds. Assume that $Cay_L(G,A)$ and $Cay_R(G,B)$ are λ-expanders for some $0 < \lambda < |A| = |B|$. Then*

$$\tilde{\mu}(G; A, B) = O(\log |G|) \ (|G| \to \infty).$$

To prove the proposition, we introduce the following two elementary lemmas.

Lemma 1. *Let $\tilde{Q}^{(\ell_L,\ell_R)}$ be the transition probability matrix of LR-RW of length-type (ℓ_L, ℓ_R). Let $\tilde{P}_L^{(\ell_L)}$ and $\tilde{P}_R^{(\ell_R)}$ be the transition probability matrices of NBRWs on $Cay_L(G,A)$ and $Cay_R(G,B)$, respectively. Then we have*

$$\tilde{Q}^{(\ell_L,\ell_R)} = \tilde{P}_R^{(\ell_R)} \tilde{P}_L^{(\ell_L)}.$$

Lemma 2. *Let $\tilde{P}_L^{(\ell_L)}$ and $\tilde{P}_R^{(\ell_R)}$ be the transition probability matrices of NBRWs of length ℓ_L and ℓ_R on $Cay_L(G,A)$ and $Cay_R(G,B)$, respectively. Let $\mathbf{u} = (u(x))_{x \in G} \in \mathbb{R}^{|G|}$ denote the stationary distribution on G, that is, for any $x \in G$ we have $u(x) = 1/|G|$. Then for any integers $\ell_L, \ell_R \ge 1$ and any probability distribution $\mathbf{p}$ on G we have*

$$\|\tilde{P}_R^{(\ell_R)} \tilde{P}_L^{(\ell_L)} \mathbf{p} - \mathbf{u}\|_\infty \le \|\tilde{P}_L^{(\ell_L)} \mathbf{p} - \mathbf{u}\|_\infty. \tag{5}$$

Proof. Notice that $\mathbf{u}$ is an eigenvector of $\tilde{P}_R^{(\ell_R)}$ corresponding to the largest eigenvalue 1 since $\tilde{P}_R^{(\ell_R)}$ is doubly stochastic. Hence we have

$$\begin{aligned} \|\tilde{P}_R^{(\ell_R)} \tilde{P}_L^{(\ell_L)} \mathbf{p} - \mathbf{u}\|_\infty &= \|\tilde{P}_R^{(\ell_R)} \tilde{P}_L^{(\ell_L)} \mathbf{p} - \tilde{P}_R^{(\ell_R)} \mathbf{u}\|_\infty \\ &= \|\tilde{P}_R^{(\ell_R)} \cdot (\tilde{P}_L^{(\ell_L)} \mathbf{p} - \mathbf{u})\|_\infty \\ &\le \|\tilde{P}_R^{(\ell_R)}\|_\infty \cdot \|\tilde{P}_L^{(\ell_L)} \mathbf{p} - \mathbf{u}\|_\infty, \end{aligned}$$

where for a matrix $M \in \mathbb{R}^{|G| \times |G|}$ recall that $\|M\|_\infty$ denotes the matrix norm induced by L^∞-norm of vectors in $\mathbb{R}^{|G|}$. Hence the inequality (5) immediately follows from the fact that $\|\tilde{P}_R^{(\ell_R)}\|_\infty \le 1$ (since, again, $\tilde{P}_R^{(\ell_R)}$ is doubly stochastic). □

Now we are ready to prove Proposition 3.

Proof of Proposition 3. We divide the proof into the following two cases.
Case 1 ($\ell_L \geq \ell_R$) Lemmas 1 and 2 imply that for any $\ell_R \geq 1$

$$\max_{g,h\in G}\left|\tilde{Q}_{h,g}^{(\ell_L,\ell_R)} - \frac{1}{|G|}\right| \leq \max_{g,h\in G}\left|(\tilde{P}_L^{(\ell_L)})_{h,g} - \frac{1}{|G|}\right|.$$

By Theorem 1 we have $\tilde{\tau}(Cay_L(G,A)) = O(\log|G|)$, which implies that if $\ell_L = c\log|G|$ with $c > 0$ that $\tilde{\tau}(Cay_L(G,A)) \leq c\log|G|$,

$$\max_{g,h\in G}\left|\tilde{Q}_{h,g}^{(\ell_L,\ell_R)} - \frac{1}{|G|}\right| \leq \frac{1}{|G|^2}.$$

Case 2 ($\ell_L \leq \ell_R$) Notice that $\tilde{P}_L^{(\ell_L)}\mathbf{p}$ is a distribution on G for any $\ell_L \geq 1$ because $\tilde{P}_L^{(\ell_L)}$ is doubly stochastic. By Theorem 1, we have $\tilde{\tau}(Cay_R(G,B)) = O(\log|G|)$. Indeed one can observe that for any distribution $\mathbf{q}$ on G,

$$\|\tilde{P}_R^{(\ell_R)}\mathbf{q} - \mathbf{u}\|_\infty \leq \|\tilde{P}_R^{(\ell_R)}\mathbf{q} - \mathbf{u}\|_2 \leq \lambda^{\ell_R} \tag{6}$$

since $Cay_R(G,B)$ is a λ-expander (see [2, eq.(10) and Claim 2.2]), which follows from the relation between L^p-distances and expressing $\mathbf{q}$ as a linear combination of orthonormal eigenvectors of $\tilde{P}_R^{(\ell_R)}$ (note that $\tilde{P}_R^{(\ell_R)}$ is a symmetric matrix). Hence if $\ell_R = c'\log|G|$ with $c' > 0$ that $\tilde{\tau}(Cay_R(G,B)) \leq c'\log|G|$, by (6) we have

$$\|\tilde{P}_R^{(\ell_R)}(\tilde{P}_L^{(\ell_L)}\mathbf{p}) - \mathbf{u}\|_\infty \leq \frac{1}{|G|^2}.$$

Hence by Lemma 1 we have

$$\max_{g,h\in G}\left|\tilde{Q}_{h,g}^{(\ell_L,\ell_R)} - \frac{1}{|G|}\right| \leq \frac{1}{|G|^2}.$$

In summary for any $t \geq \max\{c,c'\}\cdot\log|G| + 1$ and any pair of $\ell_L, \ell_R \geq 1$ that $\ell_L + \ell_R = t$ it holds that

$$\max_{g,h\in G}\left|\tilde{Q}_{h,g}^{(\ell_L,\ell_R)} - \frac{1}{|G|}\right| \leq \frac{1}{|G|^2},$$

completing the proof. □

Hence we complete the proof of Theorem 2.

4 INSTANTIATION

We present several instances of the LR-Cayley hash functions via the special linear group $\mathrm{SL}_n(\mathbb{F}_p)$ over the finite field $\mathbb{F}_p$.

4.1 VIA THE GROUP $\mathrm{SL}_2(\mathbb{F}_p)$

The special linear group $\mathrm{SL}_2(\mathbb{F}_p)$ of rank 2 with p an odd prime is a nice platform group for instantiating LR-Cayley hash functions. It is worth mentioning that multiplications of matrices in $\mathrm{SL}_2(\mathbb{F}_p)$ can be calculated very fast which implies the efficiency of computing the hash value of the LR-Cayley hash functions on $\mathrm{SL}_2(\mathbb{F}_p)$. In particular it is known by the following theorem that random Cayley graphs on $\mathrm{SL}_2(\mathbb{F}_p)$ have high-girth and are expanders with high probability, which both are desired to instantiate LR-Cayley hash functions.

Theorem 3 ([5], [16])**.** *Let S_p be a random $2d$-element subset (with no matrices of order 2) of $\mathrm{SL}_2(\mathbb{F}_p)$ with each element is chosen uniformly at random.*

(1) *There exists a constant $\tau > 0$ that $\mathrm{girth}(Cay(\mathrm{SL}_2(\mathbb{F}_p), S_p)) \geq \tau\log_{2d-1} p$ with probability 1 (as $p \to \infty$).*

(2) *There exists a constant $\lambda(\tau) > 0$ depending only on τ that $Cay(\mathrm{SL}_2(\mathbb{F}_p), S_p)$ is a $\lambda(\tau)$-expander with probability 1 (as $p \to \infty$).*

More concretely we discuss the generators of $\mathrm{SL}_2(\mathbb{F}_p)$ consisting of $M_1(k)$ and $M_2(k)$ defined as

$$M_1(k) := \begin{pmatrix}1 & k\\ 0 & 1\end{pmatrix}, M_2(k) := \begin{pmatrix}1 & 0\\ k & 1\end{pmatrix}, M_1'(k) := \begin{pmatrix}-1 & k\\ 0 & -1\end{pmatrix}, M_2'(k) := \begin{pmatrix}-1 & 0\\ k & -1\end{pmatrix}.$$

The following theorem shows that the Cayley graphs defined by these matrices are explicit high-girth Cayley graph expanders.

Theorem 4 ([5], [24]). *Let $k \geq 1$ be a positive integer. There exists $p_0 > 0$ such that for any prime $p > p_0$ we have the followings.*

(1) $Cay(\mathrm{SL}_2(\mathbb{F}_p), \{M_1(k)^{\pm}, M_2(k)^{\pm}\})$ *is a λ-expander for some constant $0 < \lambda = \lambda(k) < 4$.*

(2) $Cay(\mathrm{SL}_2(\mathbb{F}_p), \{M_1(k)^{\pm}, M_2(k)^{\pm}\})$ *has girth at least $\tau \log_3 p$ for some constant $\tau = \tau(k) > 0$.*

(3) $Cay(\mathrm{SL}_2(\mathbb{F}_p), \{M'_1(k)^{\pm}, M'_2(k)^{\pm}\})$ *is a λ'-expander for some constant $0 < \lambda' = \lambda'(k) < 4$.*

(4) $Cay(\mathrm{SL}_2(\mathbb{F}_p), \{M'_1(k)^{\pm}, M'_2(k)^{\pm}\})$ *has girth at least $\tau' \log_3 p$ for some constant $\tau' = \tau'(k) > 0$.*

Now we have the following implemented LR-Cayley hash function $H_{G,A,B}$ with

$$G = \mathrm{SL}_2(\mathbb{F}_p), A = \{M_1(k_1)^{\pm}, M_2(k_1)^{\pm}\}, B = \{M'_1(k_2)^{\pm}, M'_2(k_2)^{\pm}\}$$

with $k_1, k_2 \geq 3$, which forms a family $\mathcal{H} = \{H_{\mathrm{SL}_2(\mathbb{F}_p),A,B} \mid p > p_0\}$. Note that TNC holds for G, A and B since the order of any matrix of A is p while any matrix of B has order $2p$, see [36].

By Theorems 2 and 4 the LR-Cayley hash function $H_{G,A,B}$ is universal. It should be appealed that if $k_1, k_2 \geq 3$ the LR-Cayley hash function $H_{G,A,B}$ would currently be secure since there are at least no known attacks solving Problem 2. Indeed the lifting attacks in [37] and [8] cannot be applied for these choices of any of A and B as mentioned in [8] and [31].

Remark 11. *Setting the parameters $1 \leq k_1, k_2 \leq 2$ is not recommended. Indeed the implemented LR-Cayley hash function $H_{G,A,B}$ above with $1 \leq k_1, k_2 \leq 2$ is no longer collision-resistant since the lifting attacks designed in [37] and [8] enable the adversary to solve Problem 2 efficiently and so it is possible to find collisions consisting of two distinct texts of length $O(\log p)$.*

4.2 VIA THE GROUP $\mathrm{SL}_n(\mathbb{F}_p)$ WITH EVEN $n \geq 4$

Another recommended platform group for instantiation is the special linear group $\mathrm{SL}_n(\mathbb{F}_p)$ with $n \geq 4$ even and p an odd prime. As in the case of $\mathrm{SL}_2(\mathbb{F}_p)$ it is proved in [7] that random Cayley graphs on $\mathrm{SL}_n(\mathbb{F}_p)$ with fixed n are high-girth expanders with high probability. Also LR-Cayley hash functions on $\mathrm{SL}_n(\mathbb{F}_p)$ would be expected to be more secure than the ones based on $\mathrm{SL}_2(\mathbb{F}_p)$ (at the expense of the efficiency of computing the hash value) as discussed in the last of this subsection.

To see more concretely we deal with the generators of $\mathrm{SL}_n(\mathbb{F}_p)$ consisting of the matrices defined as follows. Let $a, b, k \geq 2$ be integers with the following conditions.

($*$) For even $n \geq 4$, there exists a prime q with $n \equiv a \equiv b \equiv 1 \pmod{q}$. Furthermore suppose that $k = q^{c+1} + 1$ for some integer c and $k \geq 3(n-1)$.

Then let

$$N_1(a) = \begin{pmatrix} 1 & a & 0 & \dots & 0 & 0 \\ 0 & 1 & a & \dots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & a \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \; N_2(b) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ b & 1 & 0 & \dots & 0 & 0 \\ 0 & b & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \dots & b & 1 \end{pmatrix}.$$

Theorem 5 ([3]). *Let $n \geq 4$ be a fixed even integer. Let a, b and k be positive integers satisfying ($*$). Then there exists $p_1 > 0$ such that for any prime $p > p_1$ we have the followings.*

(1) $Cay(\mathrm{SL}_n(\mathbb{F}_p), \{N_1(a)^{\pm k}, N_2(b)^{\pm k}\})$ *is a λ-expander for some $0 < \lambda = \lambda(a, b, k) < 4$.*

(2) $Cay(\mathrm{SL}_n(\mathbb{F}_p), \{-N_1(a)^{\pm k}, -N_2(b)^{\pm k}\})$ *is a λ'-expander for some $0 < \lambda' = \lambda'(a, b, k) < 4$.*

(3) $Cay(\mathrm{SL}_n(\mathbb{F}_p), \{N_1(a)^{\pm k}, N_2(b)^{\pm k}\})$ *has girth at least $\sigma \log_3 p$ with some constant $\sigma = \sigma(a, b, k) > 0$.*

(4) $Cay(\mathrm{SL}_n(\mathbb{F}_p), \{-N_1(a)^{\pm k}, -N_2(b)^{\pm k}\})$ *has girth at least $\sigma' \log_3 p$ with some constant $\sigma' = \sigma'(a, b, k) > 0$.*

To instantiate LR-Cayley hash function on $\mathrm{SL}_n(\mathbb{F}_p)$ for each even integer $n \geq 4$, set

$$G = \mathrm{SL}_n(\mathbb{F}_p), \; A = \{N_1(a_L)^{\pm k_L}, N_2(b_L)^{\pm k_L}\}, \; B = \{-N_2(a_R)^{\pm k_R}, -N_2(b_R)^{\pm k_R}\}.$$

where $a_L, b_L, k_L, a_R, b_R, k_R$ are integers satisfying ($*$). Notice that TNC holds for G, A and B since the order of any matrix of A is p by Lucas's theorem while any matrix of B has order $2p$. Accordingly for each even $n \geq 4$ we obtain a family which $\mathcal{H} = \{H_{\mathrm{SL}_n(\mathbb{F}_p),A,B} \mid p > p_1\}$. Then the implemented LR-Cayley hash function $H_{G,A,B}$ is universal by Theorems 2 and 5.

Now we discuss the collision-resistance of the LR-Cayley hash function $H_{G,A,B}$. Let

$$X_L = N_1(a_L)^{k_L}, Y_L = N_2(b_L)^{k_L}, X_R = -N_1(a_R)^{k_R}, Y_R = -N_2(b_R)^{k_R}.$$

In this case, Problem 2 can be deduced to the following systems with variables $x_1, x_2, \ldots, x_{n_L}$, $y_1, y_2, \ldots, y_{n_L}$, $z_1, z_2, \ldots, z_{n_R}$, $w_1, w_2, \ldots, w_{n_R}$ over $\mathbb{F}_p$.

$$\begin{cases} X_L^{x_1} Y_L^{y_1} X_L^{x_2} Y_L^{y_2} \cdots X_L^{x_{n_L}} Y_L^{y_{n_L}} = I_n \\ X_R^{z_1} Y_R^{w_1} X_R^{z_2} Y_R^{w_2} \cdots X_R^{z_{n_R}} Y_R^{w_{n_R}} = I_n \end{cases} \tag{7}$$

Here recall that I_n denotes the $n \times n$ identity matrix. Notice that the first (resp. second) equation in the system (7) corresponds to n^2 polynomial equations with variables $x_1, x_2, \ldots, x_{n_L}$, $y_1, y_2, \ldots, y_{n_L}$ (resp. $z_1, z_2, \ldots, z_{n_R}$, $w_1, w_2, \ldots, w_{n_R}$). Note that solving the system (7) is expected to in general be harder than the system (E_m) in [13] with $m = n_L + n_R$ since the system (7) contains more variables and more polynomial equations, which gives us certain confidence that $H_{G,A,B}$ is more collision-resistant than the Cayley hash function in [13] based on $Cay_L(G, A)$ or $Cay_R(G, B)$.

Also as mentioned in [13] solving systems of multivariate polynomial equations are known to be NP-hard (in worst-case) suggesting good level of security (whereas it does not necessarily imply post-quantum security). On the other hand solving (7) is finding two distinct factorization of identity in $\mathrm{SL}_n(\mathbb{F}_p)$. It should be noted that as far as we know there is no known efficient (theoretical or implemented) algorithm for such (even a single) factorization (except for very specific choices of generators) which is a reason that one can expect that $H_{G,A,B}$ is still collision-resistant against attacks by quantum computers. A detailed discussion can be found in Section 3.2 in [13].

5 CONCLUSION

We propose a new framework for building provably secure hash functions from group-theoretic techniques. Precisely we design a hash function on a left-right Cayley complex (LR-Cayley complex) with hashing process based on new random walk, left-right random walk (LR-RW), on the complex, whereas the series of studies initiated by Zémor have used random walks on high-girth Cayley graph expanders. By doing this, we can reduce the security of hash functions to the simultaneous balance problem and (equivalently) simultaneous representation problem. Moreover, by proving mixing properties of LR-RW, we show that our hash functions are universal if the underlying LR-Cayley graph is an expander graph. Employing the results [5], [3], we give some instantiation for our proposal. In summary, the security of our proposed LR-Cayley hash functions has reduction to the problem that seem to be harder than the usual word problem, and the distribution of the outputs is proved to tend to be uniform under some assumptions. However, there are several open problems to be addressed in the future:

- Our proposed functions seem to be not malleable, so we need to compare the efficiency of LR-Cayley hash functions with that of Cayley hash functions of Zémor-Tillich type.
- We introduce new assumptions of group-theoretic problems named the simultaneous balance problem and simultaneous representation problem. So, analysis for the hardness of these problems is an important problem.
- As with Cayley hash functions, the security and efficiency of our proposed LR-Cayley hash functions may highly rely on choice of a group and its generator. The construction of LR-Cayley complexes suitable for cryptography remains an open problem.

ACKNOWLEDGEMENTS

We appreciate the anonymous referees for their useful comments. We are grateful to Vladimir Shpilrain for his valuable comments. We thank Delaram Kahrobaei for letting us know the paper [13]. This work was supported by JST CREST Grant Number JPMJCR2113, Japan and JSPS KAKENHI Grant Number 23K13007, Japan. This work was also supported by Institute of Mathematics for Industry, Joint Usage/Research Center in Kyushu University (FY2022 Short-term Joint Research "Toward a new method for constructing expander graphs and their applications" (2022a017). This research was in part conducted under a contract of "Research and development on new generation cryptography for secure wireless communication services" among "Research and Development for Expansion of Radio Wave Resources (JPJ000254)", which was supported by the Ministry of Internal Affairs and Communications, Japan.

REFERENCES

[1] N. Alon and Y. Roichman. "Random Cayley graphs and expanders". In: *Random Structures & Algorithms* 5.2 (1994), pp. 271–284.

[2] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin. "Non–backtracking random walks mix faster". In: *Communications in Contemporary Mathematics* 9.04 (2007), pp. 585–603.

[3] G. Arzhantseva and A. Biswas. "Logarithmic girth expander graphs of $SL_n(\mathbb{F}_p)$". In: *Journal of Algebraic Combinatorics* 56.3 (2022), pp. 691–723.

[4] J.-P. Aumasson. *Serious Cryptography: A Practical Introduction to Modern Encryption*. No Starch Press, 2017.

[5] J. Bourgain and A. Gamburd. "Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(\mathbb{F}_p)$". In: *Annals of Mathematics* (2008), pp. 625–642.

[6] E. Breuillard and A. Gamburd. "Strong uniform expansion in $SL(2, p)$". In: *Geometric and Functional Analysis* 20.5 (2010), pp. 1201–1209.

[7] E. Breuillard, B.J. Green, R.M. Guralnick, and T. Tao. "Expansion in finite simple groups of Lie type". In: *Journal of the European Mathematical Society* 17.6 (2015), pp. 1367–1434.

[8] L. Bromberg, V. Shpilrain, and A. Vdovina. "Navigating in the Cayley graph of $\mathrm{SL}_2(\mathbb{F}_p)$ and applications to hashing". In: *Semigroup Forum* 94.2 (2017), pp. 314–324.

[9] W. Castryck, T. Decru, and B. Smith. "Hash functions from superspecial genus–2 curves using Richelot isogenies". In: *Journal of Mathematical Cryptology* 14.1 (2020), pp. 268–292.

[10] D.X. Charles, K.E. Lauter, and E.Z. Goren. "Cryptographic hash functions from expander graphs". In: *Journal of Cryptology* 22.1 (2009), pp. 93–113.

[11] C. Charnes and J. Pieprzyk. "Attacking the SL_2 hashing scheme". In: *Advances in Cryptology- ASIACRYPT'94: 4th International Conferences on the Theory and Applications of Cryptology Wollongong, Australia, November 28–December 1, 1994 Proceedings 4*. Springer. 1995, pp. 322–330.

[12] P. Chiu. "Cubic Ramanujan graphs". In: *Combinatorica* 12.3 (1992), pp. 275–285.

[13] C.L. Coz, C. Battarbee, R. Flores, T. Koberda, and D. Kahrobaei. "Post–quantum hash functions using $\mathrm{SL}_n(\mathbb{F}_p)$". In: *arXiv preprint arXiv:2207.03987* (2022).

[14] G. De Meulenaer, C. Petit, and J.-J. Quisquater. "Hardware Implementations of a Variant of the Zémor–Tillich Hash Function: Can a Provably Secure Hash Function be very efficient?" In: *Cryptology ePrint Archive* (2009).

[15] I. Dinur, S. Evra, R. Livne, A. Lubotzky, and S. Mozes. "Locally testable codes with constant rate, distance, and locality". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 357–374.

[16] A. Gamburd, S. Hoory, M. Shahshahani, A. Shalev, and B. Virág. "On the girth of random Cayley graphs". In: *Random Structures & Algorithms* 35.1 (2009), pp. 100–117.

[17] W. Geiselmann. "A note on the hash function of Tillich and Zémor". In: *IMA International Conference on Cryptography and Coding*. Springer. 1995, pp. 257–263.

[18] M. Grassl, I. Ilić, S. Magliveras, and R. Steinwandt. "Cryptanalysis of the Tillich–Zémor hash function". In: *Journal of cryptology* 24.1 (2011), pp. 148–156.

[19] H. Jo, C. Petit, and T. Takagi. "Full cryptanalysis of hash functions based on cubic Ramanujan graphs". In: *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 100.9 (2017), pp. 1891–1899.

[20] J.D. Lafferty and D.N. Rockmore. "Numerical investigation of the spectrum for certain families of Cayley graphs." In: *Expanding graphs* 10 (1992), pp. 63–73.

[21] A. Lubotzky. *Discrete Groups, Expanding Graphs and Invariant Measures*. Vol. 125. Springer Science & Business Media, 1994.

[22] A. Lubotzky, R. Phillips, and P. Sarnak. "Explicit expanders and the Ramanujan conjectures". In: *Proceedings of the eighteenth annual ACM symposium on Theory of computing*. 1986, pp. 240–246.

[23] A. Lubotzky, R. Phillips, and P. Sarnak. "Ramanujan graphs". In: *Combinatorica* 8.3 (1988), pp. 261–277.

[24] G.A. Margulis. "Explicit constructions of graphs without short cycles and low density codes". In: *Combinatorica* 2.1 (1982), pp. 71–78.

[25] M. Morgenstern. "Existence and explicit constructions of $q + 1$–regular Ramanujan graphs for every prime power q". In: *Journal of Combinatorial Theory, Series B* 62.1 (1994), pp. 44–62.

[26] C. Mullan. "Some Results in Group–based Cryptography". PhD thesis. University of London, 2011.

[27] C. Mullan and B. Tsaban. "SL_2 homomorphic hash functions: worst case to average case reduction and short collision search". In: *Designs, Codes and Cryptography* 81.1 (2016), pp. 83–107.

[28] C. Petit, K. Lauter, and J.-J. Quisquater. "Full cryptanalysis of LPS and Morgenstern hash functions". In: *International Conference on Security and Cryptography for Networks*. Springer. 2008, pp. 263–277.

[29] C. Petit, K. E Lauter, and J.-J. Quisquater. "Cayley hashes: A class of efficient graph-based hash functions". In: *preprint* (2007).

[30] C. Petit and J.-J. Quisquater. "Preimages for the Tillich–Zémor hash function". In: *Selected Areas in Cryptography: 17th International Workshop, SAC 2010, Waterloo, Ontario, Canada, August 12-13, 2010, Revised Selected Papers 17*. Springer. 2011, pp. 282–301.

[31] C. Petit and J.-J. Quisquater. "Rubik's for Cryptographers". In: *American Mathematical Society. Notices* 60.6 (2013), p. 733.

[32] C. Petit, N. Veyrat-Charvillon, and J.-J. Quisquater. "Efficiency and pseudo–randomness of a variant of Zémor–Tillich hash function". In: *2008 15th IEEE International Conference on Electronics, Circuits and Systems*. IEEE. 2008, pp. 906–909.

[33] A. Pizer. "Ramanujan graphs". In: *AMS/IP Studies in Advanced Mathematics* 7 (1998), pp. 159–178.

[34] I.D. Shkredov. "On a girth–free variant of the Bourgain–Gamburd machine". In: *Finite Fields and Their Applications* 90 (2023), p. 102225.

[35] R. Steinwandt, M. Grassl, W. Geiselmann, and T. Beth. "Weaknesses in the $SL_2(\mathbb{F}_{2^n})$ Hashing Scheme". In: *Advances in Cryptology—CRYPTO 2000: 20th Annual International Cryptology Conference Santa Barbara, California, USA, August 20–24, 2000 Proceedings*. Springer. 2000, pp. 287–299.

[36] O. Šuch. "The order of elements in $SL(2, p)$". In: *The Mathematical Gazette* 95.533 (2011), pp. 292–300.

[37] J.-P. Tillich and G. Zémor. "Group–theoretic hash functions". In: *Workshop on Algebraic Coding*. Springer. 1993, pp. 90–110.

[38] H. Tomkins, M. Nevins, and H. Salmasian. "New Zémor–Tillich type hash functions over $GL_2(\mathbb{F}_{p^n})$". In: *Journal of Mathematical Cryptology* 14.1 (2020), pp. 236–253.

[39] X. Wang, Y.L. Yin, and H. Yu. "Finding collisions in the full SHA–1". In: *Advances in Cryptology–CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005. Proceedings 25*. Springer. 2005, pp. 17–36.

[40] G. Zémor. "Hash functions and graphs with large girths". In: *Workshop on the Theory and Application of Cryptographic Techniques*. Springer. 1991, pp. 508–511.