Designs, Codes and Cryptography, 4, 381–394 (1994)

Hash Functions and Cayley Graphs

GILLES ZÉMOR
Network Dept., ENST, 46 Barrault, 75634 Paris 13, France

Editor: T. Beth

Received August 20, 1991; Revised September 23, 1993.

Abstract. We introduce cryptographic hash functions that are in correspondence with directed Cayley graphs, and for which finding collisions is essentially equivalent to finding short factorisations in groups. We show why having a large girth and a small diameter are properties that are relevant to hashing, and illustrate those ideas by proposing actual easily computable hash functions that meet those requirements.

1. Introduction

We focus on the problem of designing easily computable cryptographic hash functions, for integrity purposes. Such a function H should map the set of variable length texts over an alphabet $\mathcal{A}$, to a set of (short) fixed length texts.

$$H : \mathcal{A}^* \longrightarrow \mathcal{A}^n$$

A hash function should have the following properties:

- It should easily (i.e. quickly) computable.
- It should be computationally difficult to find "collisions", i.e. two texts having the same hashed value. (This is sometimes known as the strong collision criterion).

Many hashing schemes have been proposed and studied (see e.g. [5]), one of those, discussed by Godlewski and Camion in [6], is a Knapsack-type scheme based upon error-correcting codes, with the attractive property that the modification of any set of less than d characters of text will necessarily yield a modification of the hashed value, where d is the minimum distance of an appropriately chosen code. Unfortunately, such schemes are based upon linear computations which are well-known for their cryptographic weakness. In this paper we have tried to devise a hashing scheme with improved cryptographic strength which retains something of the features of the coding-based scheme. For that purpose, we substituted the original tool, i.e. the minimum distance of a code, by the girth of a Cayley graph. We will elaborate on this in the next section.

Our purpose in this paper is twofold. First we wish to make evident a correspondence between some hash functions and certain classes of Cayley graphs, and that the study of certain graph-theoretic parameters are relevant to hashing. Secondly, we illustrate the potential of this correspondence by studying actual, easily computable hash functions, and discuss some of their attractive features.

The paper is organised as follows. In section 2 we show how to construct hash functions from Cayley graphs, we also show why it is desirable that we choose those graphs with a large girth and small diameter. We also informally interpret the difficulty of finding collisions in group-theoretic terms. In section 3, we propose an actual hashing scheme based on a Cayley graph $\mathcal{G}(p)$ over $SL_2(\mathbf{F}_p)$. We show that hashing is fast and that $\mathcal{G}(p)$ has a large girth. In section 4 we discuss diameter issues. Finally section 5 gives concluding remarks and practical considerations.

2. A Design Strategy for Hashing

We address the problem of hashing variable length texts over an alphabet $\mathcal{A}$ (we will mainly focus on the case $\mathcal{A} = \{0, 1\}$). The hash functions we wish to consider will be constructed as follows.

2.1. *Construction*

Choose a group G and a set $\mathcal{S}$ of generators of G with $|\mathcal{S}| = |\mathcal{A}|$, together with a one-to-one mapping f between $\mathcal{A}$ and $\mathcal{S}$. The hash function $H_{G,\mathcal{S},f}$ (H for short) associated to G, $\mathcal{S}$, and f, is defined as follows. To any text x, i.e. string of elements of $\mathcal{A}$, associate the corresponding string of elements of $\mathcal{S}$, and compute the product in G to obtain the hashed value (after a suitable identification between elements of G and $\mathcal{A}^n$).

$$x = x_1x_2 \dots x_k \mapsto H(x) = f(x_1)f(x_2)\dots f(x_k)$$

Let us restate this definition of H in graph-theoretic terms. Denote by $\mathcal{G}(G, \mathcal{S})$ (or simply $\mathcal{G}$ when no confusion can arise) the directed Cayley graph associated with G and $\mathcal{S}$. This means that $\mathcal{G}$ has G as its set of vertices, and there is a directed edge between vertices v and w iff $w = vs$ with s belonging to $\mathcal{S}$. In this setting, a text x can be considered as a directed path in the graph $\mathcal{G}$, with the identity vertex as starting point, and its endpoint is precisely the hashed value $H(x)$. Now two texts yielding the same hashed value correspond to two paths with the same starting and endpoints. We would like those two paths to differ necessarily by a "minimum amount"; this can be guaranteed, if the graph $\mathcal{G}$ is chosen without short "cycles". Let us make this slightly more formal.

2.2. *The Directed Girth of a Graph*

Definition. We call the *"directed girth"* of a graph $\mathcal{G}$, the largest integer ∂ such that given any two vertices v and w, any pair of distinct directed paths joining v to w will be such that one of those paths has length (i.e. number of edges) ∂ or more.

Our purpose is to search for hash functions among Cayley graphs with large girths. If the Cayley graph $\mathcal{G}$ has a large girth ∂, then the corresponding hash function will have the property that "local" modifications of a text will necessarily modify the hashed value; more precisely

Proposition 1 *If a substring of k consecutive symbols of a text is replaced by a string of h consecutive symbols, then* $\sup(k, h) \geq \partial$ *i.e. one of those strings has more than* ∂ *symbols.*

Proof: Let $x = x_1 x_2 \dots x_i \boxed{x_{i+1} \dots x_{i+k}} x_{i+k+1} \dots x_t$ and $x' = x_1 x_2 \dots x_i \boxed{y_{i+1} \dots y_{i+h}}$ $x_{i+k+1} \dots x_t$ be two texts that differ only in the framed substrings. Then they correspond in $\mathcal{G}$ to two directed paths

$$v \to vf(x_{i+1}) \to vf(x_{i+1})f(x_{i+2}) \to \cdots \to vf(x_{i+1}) \cdots f(x_{i+k}) = w$$

$$v \to vf(y_{i+1}) \to vf(y_{i+1})f(y_{i+2}) \to \cdots \to vf(y_{i+1}) \cdots f(y_{i+h}) = w'$$

where $v = f(x_1)f(x_2) \dots f(x_i)$. $H(x) = H(x')$ iff $w = w'$, implying the proposition by definition of ∂. ■

Of course this property in itself will not guarantee a good hash function, but it is sufficiently attractive to motivate investigating further. Note, by way of illustration, that it forbids the use of commutative groups, because any Cayley graph over such a group, with two distinct nonmutually inverse generators, has girth at most 4.

2.3. On the Difficulty of Finding Collisions

In the Cayley graph setting, the difficulty of finding a collision can be expressed in the following way. Find two strings of generators (elements of $\mathcal{S}$) such that the corresponding products coincide in G; i.e. find $s_1, s_2, \dots s_n, \sigma_1, \sigma_2, \dots \sigma_m \in \mathcal{S}$ such that

$$s_1 s_2 \dots s_n = \sigma_1 \sigma_2 \dots \sigma_m$$

equivalently

$$s_1 s_2 \dots s_n \sigma_m^{-1} \sigma_{m-1}^{-1} \dots \sigma_1^{-1} = 1. \tag{1}$$

So we see that finding a collision is equivalent to finding factorisations of the form (1). Now it can be argued that there are always trivial factorisations of the form (1) in any finite group (e.g. $s^{|G|} = 1$, for any $s \in \mathcal{S}$). But we must note that only $n \approx \log |G|$ bits are needed to express hashed values, so that we can choose groups of large cardinality (e.g. $|G| = 2^{500}$) for which trivial factorisations involving $N \sim |G|$ elements are useless as actual forgeries (because no text has 2^{500} bits !). This means, broadly speaking, that the strong collision criterion is satisfied whenever it is computationally difficult to find short factorisations of the form (1). Recall that the general problem of finding the shortest factorisation of an arbitrary element of an arbitrary group over some set of generators is Pspace-complete [7]. This does not formally prove that some proper choice of a group G and generators $\mathcal{S}$ will yield an actual hash function satisfying the strong collision criterion, but it also motivates further investigation.

2.4. *Diameter Issues*

A lot of work has been devoted to the search for Cayley graphs with a small diameter, see e.g. [2]. Recall that the diameter of a directed graph is the largest distance $d(v, w)$ between two vertices v and w, $d(v, w)$ being the smallest number of edges of a directed path joining v to w). This is also relevant to our hashing scheme because a relatively small diameter is necessary to ensure that every element of G is the hashed value of some reasonably-sized text (clearly a desirable feature of a hash function). Existing studies concern non-directed Cayley graphs though, (for which $\mathcal{S} = \mathcal{S}^{-1}$). This does not suit us, because we should not have both an element s and its inverse s^{-1} in $\mathcal{S}$, otherwise the factorisation $ss^{-1} = 1$ yields trivial collisions. We will therefore draw upon existing techniques for estimating the diameter of Cayley graphs, but also adapt them to the directed case for our purposes.

To summarize, we wish to look for hash functions among directed Cayley graphs with a large directed girth and a small diameter. Note that when $|G|$ goes to infinity and $|\mathcal{S}|$ is kept constant, the girth of Cayley graphs over G may not grow faster than $O(\log |G|)$, while the diameter may not grow slower than $O(\log |G|)$. In the rest of this paper we will investigate hash functions based on families of directed Cayley graphs whose directed girth and diameter are both in $O(\log |G|)$.

3. Hashing Schemes Based on Computations in $SL_2(F_p)$

From now on we will deal with the alphabet $\mathcal{A} = \{0, 1\}$. Let p be a large prime number, (e.g. of about 150 bits). Denote by $\mathbf{Z}$ the set of integers and by $\mathbf{F}_p$ the finite field with p elements. We shall be dealing with Cayley graphs over the group $G = SL_2(\mathbf{F}_p)$ of 2×2-matrices of determinant 1 over $\mathbf{F}_p$. We shall consider severall possible sets of generators $\mathcal{S} = \{A, B\}$, among which three basic sets:

1. $\mathcal{S}_1 = \{A_1, B_1\}$ with $A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ $B_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

2. $\mathcal{S}_2 = \{A_2, B_2\}$ with $A_2 = A_1^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ $B_2 = B_1^2 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$

3. $\mathcal{S}_3 = \{A_3, B_3\}$ with $A_3 = A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ $B_3 = A_1B_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

Denote by $\mathcal{G}_i(p)$, $i = 1, 2, 3$, the corresponding Cayley graphs (from now on will shall often ommit the p to lighten notation).

Define the corresponding hash functions H_i associated with $\mathcal{G}_i$ as in section 2.1 where f will simply be the correspondence: $f : 0 \mapsto A_i, 1 \mapsto B_i$. When the actual choice of generator set $\{A_i, B_i\}$ is indifferent, we shall simply write A and B.

3.1. A Few Remarks

(i) Note that the function thus constructed hashes texts over the alphabet $\{0, 1\}$, and that the hashed values have a (fixed) length close to $3 \log p$ bits, because $|SL_2(\mathbf{F}_p)| = p(p^2 - 1)$. Note also that the multiplication of an arbitrary matrix of $SL_2(\mathbf{Z})$ by A or B requires just integer additions (between two and four), so that in $SL_2(\mathbf{F}_p)$, a multiplication by A or B requires essentially a few additions of $\log p$-bit integers; so computing the hashed values is therefore reasonably fast.

(ii) Hash functions based on associating an alphabet with a basic set of matrices and multiplying them in $GL_2(\mathbf{F}_p)$ have been proposed before, but with basic matrices of arbitrary size, so that a large p could not be chosen without damaging the speed of computation, and forging could be achieved with probabilistic methods of factoring in $GL_2(\mathbf{F}_p)$, see Camion (1987), which do not seem to apply here.

(iii) As mentioned in section 2.3, The problem of devising a forgery, i.e. finding collisions, involves factoring elements of $SL_2(\mathbf{F}_p)$ into products of A's and B's, for instance finding factorisations of the unit element. Some trivial factorisations can easily be found, e.g. $A^p = B^p = 1 \ldots$ but they have a length comparable to p (in $O(p)$), so are useless as an actual forgery provided p is large enough; (no text has 2^{150} bits !). What is needed is a method for finding short factorisations of elements of $SL_2(\mathbf{F}_p)$ into products of A's and B's; this is really the problem whose difficulty our scheme relies on.

3.2. The Girth of the Graphs $\mathcal{G}_i$

Suppose we are using a hash function of the above type, i.e. associated to some Cayley graph $\mathcal{G}(G, \{A, B\})$ with $G = SL_2(\mathbf{F}_p)$. Suppose that a subset of k consecutive bits $x_1 x_2 \ldots x_k$ of a text is changed into h consecutive bits $y_1 y_2 \ldots y_h$, with $x_1 \neq y_1$ and $x_k \neq y_h$. We wish to show that if $\max(k, h)$ is small enough, then the hashed value is necessarily changed.

If the hashed value is unchanged, then the corresponding products of A's and B's are equal in $SL_2(\mathbf{F}_p)$:

$$X_1 X_2 \ldots X_k = Y_1 Y_2 \ldots Y_h \qquad \bmod p \tag{2}$$

Now suppose we have chosen our generators A, B in such a way that equality

$$X_1 X_2 \ldots X_k = Y_1 Y_2 \ldots Y_h \tag{3}$$

where $X_i = A$ or B, $Y_i = A$ or B, holds in $SL_2(\mathbf{Z})$ (over the integers) if and only if $h = k$ and $X_i = Y_i$ for $1 \leq i \leq k$. In this case we will say that $\{A, B\}$ satisfies property $(\star)$.

Then equality (2) will occur only if the matrix

$$X_1 \ldots X_k - Y_1 \ldots Y_k = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

evaluated over the integers has at least one of its entries a, b, c, d equal to a non-zero multiple of p, which implies

$$\|X_1X_2\dots X_k - Y_1Y_2\dots Y_h\| = \left\|\begin{bmatrix} a & b \\ c & d\end{bmatrix}\right\| \geq p \tag{4}$$

(With the classical notation $\|M\| = \sup_{\xi\neq 0} \|M\xi\|/\|\xi\|$ and $\|\xi\| = \sqrt{\xi_1^2+\xi_2^2}$ for $\xi = \begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$).

Inequality (4) implies

$$\max(\|X_1X_2\dots X_k\|, \|Y_1Y_2\dots Y_h\|) \geq p/2$$

and puttting $\alpha = \max(\|A\|, \|B\|)$, we obtain, by submultiplicativity of the norm of matrices,

$$\max(k, h) \geq \log_\alpha \frac{p}{2}.$$

Next we prove that the 3 sets of generators $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ mentioned above satisfy property ($\star$). To see this, note that it suffices to show the property for $\mathcal{S}_1 = \{A_1, B_1\}$, because $A_2 = A_1^2, B_2 = B_1^2$ and $A_3 = A_1, B_3 = A_1B_1$. Recall now that $SL_2(\mathbf{Z})$ is classically generated by the matrices

$$S = A_1 = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix} \text{ and } T = \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix} = -S^{-1}B_1S^{-1} = (S^{-1}BS^{-1})^3$$

in the modular group $\Gamma = SL_2(\mathbf{Z})/\{1, -1\}$, we have $T^2 = 1$ and $(TS)^3 = 1$, furthermore it is well known that Γ is isomorphic to the free product of a cyclic group generated by an x of order 2, and a cyclic group generated by a y of order 3, where x corresponds to T and y corresponds to TS (see e.g. [8] or [13]). From this it can readily be deduced that an equality of the form:

$$X_1X_2\dots X_kY_h^{-1}\dots Y_2^{-1}Y_1^{-1} = 1 \tag{5}$$

with $X_i = A_1$ or B_1, and $Y_j = A_1$ or B_1, cannot hold in Γ; to see this, express the lefthandside of (5) as a string of S's and T's and of S^{-1} and T's, and observe that the formal simplifications $T^2 = (TS)^3 = 1$ do not suffice to formally collapse this product, unless $X_1, X_2, \dots, X_k$ and $Y_1, Y_2, \dots, Y_h$ are identical strings.

By elementary linear algebra, $\|A\| = \sqrt{\|{}^tAA\|} = \sqrt{\lambda}$ where λ is the largest eigenvalue of ${}^tAA$. Hence $\|A_1\| = \|B_1\| = \|A_3\| = \phi = \frac{1+\sqrt{5}}{2} \approx 1.62$. $\|A_2\| = \|B_2\| = \alpha_2 = 1+\sqrt{2} \approx 2.41$. $\|B_3\| = \alpha_3 = \frac{3+\sqrt{5}}{2} \approx 2.61$.

We have proved:

Proposition 2 *The girths of $\mathcal{G}_1(p), \mathcal{G}_2(p), \mathcal{G}_3(p)$ satisfy respectively*

$$\partial_1 \geq \log_\phi \frac{p}{2} \tag{6}$$

$$\partial_2 \geq \log_{\alpha_2} \frac{p}{2} \tag{7}$$

$$\partial_3 \geq \log_{\alpha_3} \frac{p}{2} \tag{8}$$

The idea of using $SL_2(\mathbf{F}_p)$ for constructing Cayley graphs with large girths originates in [10], from which the above proof is inspired.

3.3. *Choosing a Set of Generators*

In a preliminary version of this work, we thought of using the pair of generators $\mathcal{S}_1 = \{A_1, B_1\}$ for a hashing scheme: however, it was pointed out to us by J-P Tillich that the set $\mathcal{S}_1$ should not actually be used because of the following property. Any integer matrix of $SL_2(\mathbf{Z})$ with non-negative coefficients can be decomposed as a product of A_1's and B_1's in $SL_2(\mathbf{Z})$ (essentially by applying Euclid's algorithm). To obtain collisions, it suffices therefore to find a matrix C, different from the identity matrix, with non-negative coefficients, such that $C \equiv Id \bmod p$. If enough care is taken to ensure that C has coefficients of the same order of magnitude (J-P Tillich found a nice trick to do this [14]), then factoring C over the integers as a product of A_1's and B_1's will yield a short factorisation of unity in $SL_2(\mathbf{F}_p)$ that can be used for constructing collisions.

To avoid this type of attack, one should take care to choose generators A, B with the property that the set of matrices of $SL_2(\mathbf{Z})$ that can be expressed as a product of A's and B's is sufficiently scarce. The sets $\mathcal{S}_2 = \{A_2, B_2\}$ and $\mathcal{S}_3 = \{A_3, B_3\}$ are such that choosing (by random search methods) matrices of $SL_2(\mathbf{Z})$ that are equivalent modulo p to a given matrix, will yield with very small probability (exponential in $\log p$) a matrix that decomposes over $\mathcal{S}_2$ or $\mathcal{S}_3$ in $SL_2(\mathbf{Z})$.

4. The Diameter of $\mathcal{G}_3(p)$

As mentioned in section 2.4, it is desirable that our Cayley graphs have a relatively small diameter. Although we suspect that $\mathcal{G}_2(p)$ does have a small diameter, we know of no way to prove this. However if we change the set of generators to $\mathcal{S}_3$, we can prove, and this will be the purpose of this section, that $\mathcal{G}_3(p)$ has a diameter in $O(\log p)$, with an acceptable constant, so that hashed values of megabyte-texts will range over all of $SL_2(\mathbf{F}_p)$. Of course the proof will be completely nonconstructive; (a constructive method would be equivalent to breaking the scheme).

Several techniques have been developped for studying the diameter of a Cayley graph $\mathcal{G}(G, \mathcal{S})$ see e.g. [2], [1], [4]; they concern however nondirected graphs, (with nondirected edges), corresponding to generating sets $\mathcal{S}$ satisfying $\mathcal{S} = \mathcal{S}^{-1}$. So we will need to adapt those techniques to directed graphs. As mentioned in section 2.4, note once more that we could not have chosen for our hash function a nondirected Cayley graph, if only to avoid the existence of trivial factorisations of the kind $AA^{-1} = 1$.

4.1. *Notation and Plan of Proof*

A graph (directed or not) with vertex set V and edge set E will be denoted by (V, E). If $\mathcal{G}$ is the Cayley graph $\mathcal{G}(G, \mathcal{S})$, then $\mathcal{G}^*$ will denote the corresponding nondirected graph (obtained from $\mathcal{G}$ by suppressing the orientation of the edges) i.e. the Cayley graph $\mathcal{G}(G, \mathcal{S} \cup \mathcal{S}^{-1})$. Therefore $\mathcal{G}_i^*(p)$ will denote $\mathcal{G}(SL_2(\mathbf{F}_p), \{A_i, B_i, A_i^{-1}, B_i^{-1}\})$.

If X is a subset of vertices of a graph, denote by $N_+(X)$ ($N_-(X)$) the set of vertices not in X, that are the endpoints of an edge with its initial point in X (that are the initial points of an edge with its endpoint in X). In the Cayley graph case, $N_+(X) = X\mathcal{S} \setminus X$, $N_-(X) = X\mathcal{S}^{-1} \setminus X$. Let $N(X)$ denote $N(X) = N_+(X) \cup N_-(X)$.

A method is indicated in [2] to prove that $\mathcal{G}_1^*(p)$ has a diameter in $O(\log p)$; it cannot be deduced from it directly, however, that $\mathcal{G}_1(p)$ and $\mathcal{G}_3(p)$ have a diameter in $O(\log p)$. We will need to use the "expansion" properties of those graphs. Following [1], call c a *magnifying coefficient* of a nondirected graph with vertex set V whenever

$$\text{for all subsets } X \text{ of } V \text{ such that } |X| \leq \frac{|V|}{2}, \quad |N(X)| \geq c|X|$$

For a directed graph we will also call c a magnifying coefficient when

$$\text{for all subsets } X \text{ of } V \text{ such that } |X| \leq \frac{|V|}{2}, \left\{ \begin{array}{l} |N_+(X)| \geq c|X| \\ |N_-(X)| \geq c|X| \end{array} \right.$$

It is reasonably straightforward to obtain that if the graphs $\mathcal{G}_3(p)$ and $\mathcal{G}_3^*(p)$ (in directed and nondirected cases) have "good expansion properties" (i.e. magnifying coefficients independent of p) then they have a diameter in $O(\log p)$; it is also possible to prove that if the nonoriented versions $\mathcal{G}_3^*(p)$ of the graphs $\mathcal{G}_3(p)$ have good expansion properties, then the oriented versions $\mathcal{G}_3(p)$ also have good expansion properties. This is essentially the object of the next lemmas, namely to reduce the study of the diameter of $\mathcal{G}_3(p)$ to the study of the expanding properties of a nondirected graph. Actually the nondirected graph we will reduce our problem to will not be $\mathcal{G}_3^*(p)$ but rather $\mathcal{G}(PSL_2(\mathbf{F}_p), \{S, S^{-1}, T\})$ (S and T being defined as in (3.2)), at which point arithmetic considerations can be brought in, following [2].

4.2. Reduction to the Study of $\mathcal{G}(PSL_2(F_p), \{S, S^{-1}, T\})$

LEMMA 4.1 *Let (V, E) be a directed graph, and suppose it has magnifying coefficient c. Then the diameter D verifies*

$$D \leq 2 \log_{(1+c)} \frac{|V|}{2} + 1 \leq \frac{2}{c} \ln \frac{|V|}{2} + 1$$

Proof: Let v and w be any two vertices of V. Denote by $N_+^{[k]}(v)$ ($N_-^{[k]}(w)$) the subset of vertices of V reachable from v by paths of length k or less (from which w can be reached by paths of length k or less). In other words define inductively

$$N_+^{[0]}(v) = v; \qquad N_+^{[k+1]}(v) = N_+^{[k]}(v) \cup N_+(N_+^{[k]}(v))$$

$$N_-^{[0]}(w) = w; \qquad N_-^{[k+1]}(w) = N_-^{[k]}(w) \cup N_-(N_-^{[k]}(w))$$

That c is a magnifying coefficient of (V, E) means that as long as $|N_+^{[k-1]}(v)| \leq \frac{|V|}{2}$ and $|N_-^{[h-1]}(v)| \leq \frac{|V|}{2}$, we have:

$$\begin{cases} |N_+^{[k]}(v)| \geq (1+c)^k \\ |N_-^{[h]}(w)| \geq (1+c)^h \end{cases}$$

take $k > \log_{(1+c)} \frac{|V|}{2}$ and $h \geq \log_{(1+c)} \frac{|V|}{2}$, then $N_+^{[k]}(v)$ and $N_-^{[h]}(w)$ necessarily have a common vertex z, and there is therefore a path joining v to w,passing through z, with length $k + h$ or less. ■

In the nondirected graph case, some other methods can be brought in to improve on the constant $\frac{2}{c}$, (see [1]; [4]) but by methods that do not seem to generalise to the directed case, (at least not for noncommutative Cayley graphs), and that do not represent a substantial improvement when c is small, both of which are the case here.

LEMMA 4.2 *Suppose $\mathcal{G}(PSL_2(\mathbb{F}_p), \{S, S^{-1}, T\})$ has a magnifying coefficient c, then $\frac{c}{6}$ is a magnifying coefficient for $\mathcal{G}(PSL_2(\mathbb{F}_p), \{A_3, B_3\})$.*

Proof: suppose the contrary, then there is a subset $X \subset PSL_2(\mathbb{F}_p)$, with $|X| \leq |PSL_2(\mathbb{F}_p)|/2$ such that in $\mathcal{G}(PSL_2(\mathbb{F}_p), \{A_3, B_3\})$, either $N_+(X)$ or $N_-(X)$ has cardinality less than $\frac{c}{6}|X|$. Suppose, for example, that it is the case for $N_+(X)$ (the other case being analoguous), then

$$|XA_3 \setminus X| < \frac{c}{6}|X| \text{ and } |XB_3 \setminus X| < \frac{c}{6}|X|$$

therefore (recall that $A_3 = S$ and $B_3 = S^2TS$)

$$|XS \cap X| > \left(1 - \frac{c}{6}\right)|X|, \text{ equivalently,} \tag{9}$$

$$|XS^2 \cap XS| > \left(1 - \frac{c}{6}\right)|X|, \text{ and} \tag{10}$$

$$|XS^2TS \cap X| > \left(1 - \frac{c}{6}\right)|X| \tag{11}$$

(11) is equivalent to

$$|XS^2T \cap XS^{-1}| > \left(1 - \frac{c}{6}\right)|X| \tag{12}$$

but (9) means that $|XS^{-1} \cap X| > (1 - \frac{c}{6})|X|$, and applied to (12) this yields

$$|XS^2T \cap X| > \left(1 - \frac{2c}{6}\right)|X|$$

or equivalently, since $T = T^{-1}$ in $PSL_2(\mathbf{F}_p)$,

$$|XS^2 \cap XT| > \left(1 - \frac{2c}{6}\right)|X|$$

applying successively (10) and (9), we obtain:

$$|XS \cap XT| > \left(1 - \frac{3c}{6}\right)|X| \tag{13}$$

$$|X \cap XT| > \left(1 - \frac{4c}{6}\right)|X| \text{ i.e.} \tag{14}$$

$$|XT \setminus X| < \frac{4c}{6}|X| \tag{15}$$

(9) yields also

$$|XS \setminus X| < \frac{c}{6}|X| \tag{16}$$

$$|XS^{-1} \setminus X| < \frac{c}{6}|X| \tag{17}$$

and since in $\mathcal{G}(PSL_2(\mathbf{F}_p), \{S, S^{-1}, T\})$ we have

$$|N(X)| \leq |XT \setminus X| + |XS \setminus X| + |XS^{-1} \setminus X|$$

adding (15), (16), and (17) we obtain

$$|N(X)| < c|X|$$

a contradiction. ■

Note that since -1 commutes with every matrix, every magnifying coefficient of $\mathcal{G}(PSL_2(\mathbf{F}_p), \{A_3, B_3\})$ is a magnifying coefficient for $\mathcal{G}(SL_2(\mathbf{F}_p), \{A_3, B_3\})$, our problem is therefore now reduced to finding a magnifying coefficient for the nondirected graph $\mathcal{G}(PSL_2(\mathbf{F}_p), \{S, S^{-1}, T\})$.

4.3. *The Expanding Properties of $\mathcal{G}(PSL_2(F_p), \{S, S^{-1}, T\})$*

The method we are about to describe to obtain a magnifying coefficient for the graph $\mathcal{G}(PSL_2(\mathbf{F}_p), \{S, S^{-1}, T\})$ is hinted at in [2]; we will present it here in more detail for the sake of completeness. Denote by Γ the modular group as in section 3.2; for more detailed information on Γ and related arithmetic, see for example [13], [8].

For a prime number p denote by Γ_p the congruence subgroup of Γ

$$\Gamma_p = \left\{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma \ ; \ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \bmod p\right\}$$

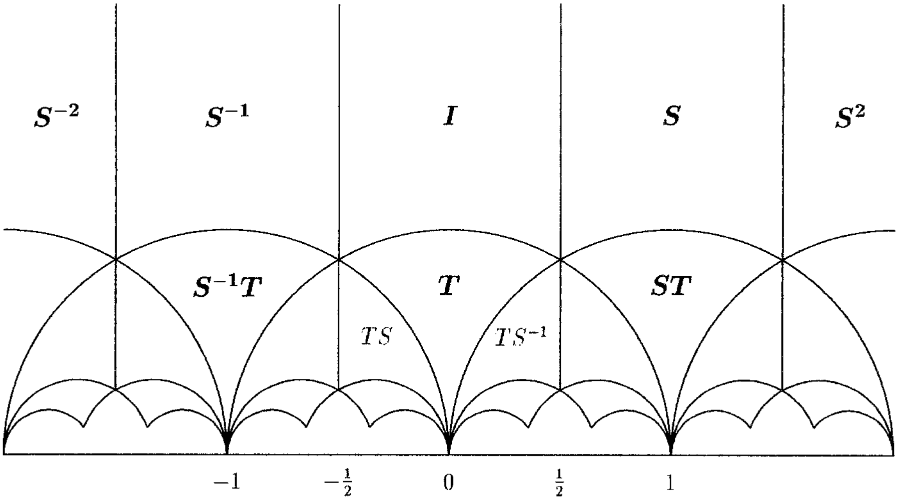

Figure 1.

we have

$$\Gamma/\Gamma_p \simeq PSL_2(\mathbf{F}_p) \simeq SL_2(\mathbf{F}_p)/\{1, -1\}$$

Let us denote by $\mathcal{H}_p$ the Cayley graph $\mathcal{G}(PSL_2(\mathbf{F}_p), \{S, S^{-1}, T\})$. Let $\boldsymbol{H}$ denote the upper complex half-plane $\{z \mid Imz > 0\}$. Recall that Γ acts on $\boldsymbol{H}$ in the following way:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} z = \frac{az+b}{cz+d}$$

A fundamental domain for this action is the region

$$\boldsymbol{D} = \left\{ z = x + iy \ ; \ -\frac{1}{2} \leq x \leq \frac{1}{2}, \ x^2 + y^2 \geq 1 \right\}$$

For any $z \in \boldsymbol{H}$, identify z with all Mz when $M \in \Gamma_p$. We obtain a (Riemann) surface $\Sigma_p = \boldsymbol{H} \backslash \Gamma_p$ on which $PSL_2(\mathbf{F}_p) = \Gamma/\Gamma_p$ acts naturally. Σ_p provides us with the following geometric representation of the Cayley graph $\mathcal{H}_p$: the vertices of $\mathcal{H}_p$ can be seen as the domains $g\boldsymbol{D}$ for $g \in PSL_2(\mathbf{F}_p)$, and any two domains $g_1\boldsymbol{D}$ and $g_2\boldsymbol{D}$ are adjacent in $\mathcal{H}_p$ iff they intersect in a curve in Σ_p. See *fig. 1* for an illustration.

The result that is brought in at this point is the following [12]:

Proposition 3 *For any real-valued function f defined on $\boldsymbol{H}$ and invariant under Γ_p, and satisfying the following properties:*

α) *f is continuously differentiable*

β) $\int\int_{D_p} f(x, y) \frac{dxdy}{y^2} = 0$

γ) $\quad \iint_{D_p} f^2(x,y)\frac{dxdy}{y^2} = 1$

we have

$$\frac{3}{16} \leq \iint_{D_p} \left(\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 \right) dx\,dy$$

where D_p denotes a union of regions $g\boldsymbol{D}$ when g describes a set of representatives of $PSL_2(\mathbf{F}_p)$ (i.e. D_p is a fundamental domain for the action of Γ_p).

Now choose any subset X of vertices of $\mathcal{H}_p$ such that $|X| \leq |PSL_2(\mathbf{F}_p)|/2$. A lower bound on $|N(X)|/|X|$ will be achieved through a lower bound on $|E(X)|/|X|$ where $E(X)$ denotes the set of edges joining vertices of X to vertices not in X. The idea is to consider the function g_X defined by $g_X(z) = 1$ if z is in the interior of $g\boldsymbol{D}$ with $g \in X$, and $g_X(z) = -e$ if z is in the interior of $g\boldsymbol{D}$ with $g \notin X$, e being an appropriately chosen positive constant. Then the function g_X is modified so as to provide us with a function f_X satisfying conditions α), β), γ) above; the point is that the function f_X will essentially be constant (with $grad f_X = 0$) except for regions corresponding to frontiers between domains $g_1\boldsymbol{D}$ and $g_2\boldsymbol{D}$ with $g_1 \in X$ and $g_2 \notin X$. Then the application of proposition 3 will give us an inequality of the form

$$\frac{3}{16} \leq \frac{1}{c_1|X|}(c_2|E(X)| + c_3|X|)$$

where c_1, c_2, c_3 are constants, and hence yield a constant lower bound on $|E(X)|/|X|$. From this, the fact that $\frac{|E(X)|}{|X|} \leq 3\frac{|N(X)|}{|X|}$, and lemmas 4.1 and 4.2 we obtain a constant κ such that

Proposition 4 *The diameter D of $\mathcal{G}_p$ verifies $D \leq \kappa \ln p$*

With some tedious but reasonably straightforward estimations of the above constants c_1, c_2, c_3 that we wish to spare the reader, it can be grossly estimated that the constant κ is of the order 500,000 or less, which is just about satisfactory since hash functions are supposed to hash texts of several megabytes.

5. Concluding Remarks

We conclude by two technical remarks, and some practical considerations.

1. The following generalisation of lemma 4.2 is not difficult to obtain.

   If a directed graph $\mathcal{G}$ has degree d (i.e. for any vertex v, $|N_+(v)| = |N_-(v)| = d$) *and if its non directed version $\mathcal{G}^*$ has a magnifying coefficient c, then $c/(d+1)$ is a magnifying coefficient for $\mathcal{G}$.*

2. Other Cayley graphs than the ones we have proposed can be envisaged, e.g. directed versions of the so-called “Ramanujan graphs” constructed in [9], [11], which have large girths and magnifying coefficients. The methods of lemmas 4.1, 4.2 and the

above remark will prove that they have small directed diameters and will yield a better estimate of the constant κ in proposition 4. The sets of generators we propose however have the practical advantage of being simpler and of cardinality two.

3. It has been pointed out to us, that choosing A and B to be each other's transpose (as in the set S_2), may be a potential weakness of the hash function. Essentially this is because given a text x, there is an obvious way of creating a text x', such that the hashed values of x and x' are matrices whose diagonal elements coincide. However it seems to us that making a third coordinate of those matrices coincide modulo p, in a more efficient way than by random search, (which must of course be prohibited by the size of p), is quite difficult. This remark, among other arguments, shows that the security parameter of our hash function should be considered to be the length $\log p$ of the prime number p rather than the length of the hashed value, i.e. $3 \log p$. We propose using a prime number of 150 bits, as a practical value (so that the hashed values are 450 bits long). In this case, proposition 2 shows that the girths ∂_i satisfy $\partial_2 \geq 118$ and $\partial_3 \geq 108$. It was also put to us by Marc Girault that some care should be taken in the choice of the prime number p, because finding simultaneously two texts and a prime number p such that those two texts collide for the hash function associated to p, is substantially easier than finding a collision for a given p. So p should be chosen in some manner allowing no leeway, to ensure that a collision has not been built in the hash function. For instance take p equal to the first prime that is bigger than 2^{150}.

   The above precautions being taken, we have a fast hashing scheme that should be challenging to break. We also hope to have convinced readers that the design strategy is worth investigating further.

Acknowledgements

We wish to thank M. Girault, P. Godlewski, J-J. Quisquater and particularly J-P. Tillich for numerous helpful discussions.

References

1. Alon, N., and Milman, V. D. 1985. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Comb. Theory Ser. B* 38:73–88.
2. Babai, L., Kantor, W. M., and Lubotsky, A. 1989. Small-diameter Cayley graphs for finite simple groups. *Europ. J. of Combinatorics* 10:507–522.
3. Camion, P. 1987. Can a fast signature scheme without secret key be secure? In *Proc. AAECC*, pp. 187–196. Springer-Verlag Lec. N. Comp. Sci. 228.
4. Chung, F. R. K. 1989. Diameters and eigenvalues. *J. Am. Math. Soc.* 2:187–196.
5. Damgård, I. B. 1989. Design principles for hash functions. In *Crypto*.
6. Godlewski, P., and Camion, P. 1988. Manipulations and errors, detection and localization. In *Advances in Cryptology, EUROCRYPT-88*, pp. 96–106. LNCS 330 Springer-Verlag.
7. Jerrum, M. R. 1985. The complexity of finding minimum length generator sequences. *Theoretical Computer Science* 36:265–289.
8. Koblitz, N. 1984. *Introduction to Elliptic Curves and Modular Forms*. Springer-Verlag.

9. Lubotsky, A., Philips, R., and Sarnack, P. 1988. Ramanujan graphs. *COMBINATORICA* 8(3):261–277.
10. Margulis, G. A. 1982. Explicit constructions of graphs without short cycles and low density codes. *COMBINATORICA* 2(1):71–78.
11. Margulis, G. A. 1988. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problemy Peredachi Informatsii* 24(1):51–60.
12. Selberg, A. 1965. On the estimation of Fourier coefficients of modular forms. *AMS Proc. Symp. Pure Math.* 8:1–15.
13. Serre, J-P. 1973. *A Course in Arithmetic*. Springer-Verlag.
14. Tillich, J-P., and Zémor, G. 1993. Group-theoretic hash functions. In *French-Israeli workshop in Algebraic coding*. LNCS 781 Springer-Verlag, to appear.