On Families of Hash Functions via Geometric Codes and Concatenation

Jürgen Bierbrauer[1], Thomas Johansson[2],
Gregory Kabatianskii[3], Ben Smeets[2]

[1] Mathematisches Institut der Universität,
Im Neuenheimer Feld 288, 69 Heidelberg, Germany

[2] Dept. of Information Theory, University of Lund,
Box 118, S-221 00, Lund, Sweden †

[3] Inst. for Problems of Information Transmission, Russian Academy of Sciences,
Ermolovoy 19, Moscow, GSP-4, Russia

Abstract. In this paper we use coding theory to give simple explanations of some recent results on universal hashing. We first apply our approach to give a precise and elegant analysis of the Wegman-Carter construction for authentication codes. Using Reed-Solomon codes and the well known concept of concatenated codes we can then give some new constructions, which require much less key size than previously known constructions. The relation to coding theory allows the use of codes from algebraic curves for the construction of hash functions. Particularly, we show how codes derived from Artin-Schreier curves, Hermitian curves and Suzuki curves yield good classes of universal hash functions.

1 Introduction

The concept of universal classes of hash functions was introduced by Carter and Wegman in [1]. It has found numerous applications of which we mention only cryptography, complexity theory, search algorithms and associative memory (see the Introduction in [2]). Three essentially different applications of universal hashing to authentication codes, [3], have been described in [4],[5] and [6]. Two of them are concerned with authentication without secrecy, the third (in [5]) is a novel use of universal classes of hash functions for error detection and information reduction in a system which guarantees integrity and secrecy.

In this paper, we present a detailed analysis of constructions of families of almost strongly universal hash functions proposed by Wegman and Carter [4] and recently, by Stinson [7],[6]. Our analysis is based on a recently discovered relationship between families of hash functions (or authentication codes (A-codes)) and error-correcting codes [8].

In Section 2 we give a simple explanation of previous results on the above mentioned constructions using the theory of concatenated codes, [10]. In the next section we present various improvements by using the concatenation of the

† These authors was supported by the TFR grant 222 92-662

D.R. Stinson (Ed.): Advances in Cryptology - CRYPTO '93, LNCS 773, pp. 331-342, 1994.

well-known Reed-Solomon codes (RS-codes) and by using the powerful algebraic geometry codes (AG-codes) derived from algebraic curves. Finally we present some numerical results.

2 Universal hash functions and codes

In this section we recall some definitions and results from [7], [6]. We start by reformulating some of the results given by Stinson in a coding theoretic language and then proceed with introducing some additional notation.

Definition 1. Let $\epsilon > 0$. A multiset H of n functions from a set A to a q-set B is $\epsilon-$*almost universal*$_2$ (short: AU_2) if for every pair $a_1, a_2 \in A$, $a_1 \neq a_2$ the number $d_H(a_1, a_2) = |\{h \in H\ ;\ h(a_1) = h(a_2)\}| \leq \epsilon n$.

Consider now a q-ary code V of length n, ($q = |B|$, $n = |H|$), whose codewords have the form $v = (h_1(a), ..., h_n(a))$, $a \in A$. It is rather clear (see also [8], [9]) that the property of ϵ-almost universal$_2$ is equivalent to the property that the minimal code distance $d(V)$ of the code V is not less than $n(1-\epsilon)$. So we have

Lemma 2. *Let $\varepsilon > 0$, $q = |B|$ and $n = |H|$. Then the following are equivalent:*

(i) H is an $\epsilon - AU_2$ class of hash functions from A to B.
(ii) The words $v = (h_1(a), ..., h_n(a))$, $a \in A$ form a q-ary code of length n with minimum distance d, where $1 - \frac{d}{n} \leq \epsilon$.

Stinson's analysis and improvements of the Wegman and Carter construction are based on two constructions, composition 1 and 2, which can be considered as the construction of concatenated codes (see [10]). The corresponding theorems on the performance of these constructions can be reformulated now in a more familiar manner.

In [7, Theorem 5.5] the following is shown:

Theorem 3 (Composition 1). *Let H_1 be an $\varepsilon_1 - AU_2$ from A_1 to B_1 and let H_2 be an $\varepsilon_2 - AU_2$ from B_1 to B_2. Then $H = H_1 \times H_2$ is an $\varepsilon - AU_2$ from A_1 to B_2 with $\varepsilon \leq \varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2$.*

In our language, this is a concatenation of two codes. If D is the distance of the concatenated code, then it is a well known fact that $D \geq d_1 d_2$. This gives $D \geq (1-\varepsilon_1)(1-\varepsilon_2)n_1 n_2$ and $\varepsilon \leq \varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2$. Thus this result is only a reformulation of the distance property of concatenated codes!

We recall from [7] also the notion of almost strongly universal hash functions.

Definition 4. Let $\epsilon > 0$. A multiset H of n functions from a set A to a q-set B is $\epsilon-$*almost strongly universal*$_2$ (short: ASU_2) if

1. for every $a \in A$ and $y \in B$ the number of elements of H mapping $a \mapsto y$ is n/q,

2. for every pair $a_1, a_2 \in A$, $a_1 \neq a_2$, and every pair $y_1, y_2 \in B$ the number of elements of H affording the operation $a_1 \mapsto y_1$,$a_2 \mapsto y_2$ is $\leq \epsilon \cdot n/q$.

The notion of ASU_2 is clearly a generalization of orthogonal arrays of strength 2. In fact, an orthogonal array is obtained if in 2. of Definition 4 we always have equality. Condition 1 then follows automatically. Hence ASU_2 may be described as *generalized orthogonal arrays*. This link between ASU_2−classes and orthogonal arrays has been observed in earlier work. The relation between ASU_2−classes (or equivalently authentication codes), and error-correcting codes is already described in [8].

Now, also Theorem 5.6 in [7] can be reformulated as the product of "distances":

Theorem 5 (Composition 2). *Let H_1 be an $\varepsilon_1 - AU_2$ from A_1 to B_1 and let H_2 be an $\varepsilon_2 - ASU_2$ from B_1 to B_2. Then $H = H_1 \times H_2$ is an $\varepsilon - ASU_2$ from A_1 to B_2 with $\varepsilon \leq \varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2$.*

In our terms the theorem states that the concatenation of two codes, where the inner code additionally has the A-code properties, gives a code that satisfies the A-code properties with $\varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2$. Using our coding theoretic notation, we give a new proof, found in the appendix.

The families of hash functions which are used in Wegman&Carter and in Stinson's constructions can also be described using well-known codes. For example, it is obvious that the codes of Theorem 5.1, [7], are RS-codes with two information symbols. The codes of Theorems 5.2 and 5.3 in the same paper can be obtained in the same manner, see our Lemma 10, which is a generalization of these two.

Stinson's construction consists of two ingredients: the AU_2 classes (or error-correcting codes) and the ASU_2 classes.

For the first ingredient consider any linear code over a finite field. We fix the ground-field $\mathbb{F}_q$ and the relative minimum distance d/n of such a q-ary code. In fact the minimum distance has to be extremely large, as $\epsilon = 1 - \frac{d}{n}$ should be small. For a fixed number Q of codewords we ask for the minimum length of such a code. That is, we want a code with the highest possible rate.

Definition 6. Let natural numbers q, Q, and the real number $\epsilon, 0 < \epsilon < 1$ be given. Define $m(\epsilon, q, Q)$ as the minimum length n of a q-ary code with Q code-words and minimum distance d satisfying $d/n \geq 1 - \epsilon$.

This is a rather unusual question in coding theory. Unusual is also the fact that we are only interested in q-ary codes with relatively large q.

Similarly for the ASU_2 classes:

Definition 7. Define $m_A(\epsilon, q, Q)$ as the minimum number of functions of an $\epsilon - ASU_2$ class of hash functions from a Q-set to a q-set.

Using the above terminology, Stinson's compositions give:

Lemma 8. *With $m(\varepsilon, q, Q)$ and $m_A(\varepsilon, q, Q)$ as defined above we have*

(i) For composition 1: $m(\epsilon_1+\epsilon_2,q,Q)\le m(\epsilon_2,q,Q_1)\cdot m(\epsilon_1,Q_1,Q)$.
(ii) For composition 2: $m_A(\epsilon_1+\epsilon_2,q,Q)\le m_A(\epsilon_2,q,Q_1)\cdot m(\epsilon_1,Q_1,Q)$.
(iii) Cartesian product, [6]: $m(\epsilon,q^i,Q^i)\le m(\epsilon,q,Q)$.

Example 1. Starting from the 2-dimensional RS-code and using composition 1 and the Cartesian product recursively one obtains

$$m(\frac{i}{q},q,q^{2^i})\le q^i,$$

for every prime-power q and every $i\ge 1$. This is a construction of Stinson's, [6, Theorem 6.1], expressed in different words. □

The k-dimensional Reed-Solomon code yields $m(\frac{k-1}{q},q,q^k)\le q$. The Singleton bound shows that we actually have equality:

Theorem 9.

$$m(\frac{k-1}{q},q,q^k)=q$$

for every prime power q and $k\ge 2$.

What the ASU_2−classes are concerned we may use the following lemma:

Lemma 10 "projection hashing". *Let π be some $\mathbb{F}_{q_0}$-linear map from $\mathbb{F}_Q$ on $\mathbb{F}_q$, where $Q=q_0^n, q=q_0^m$ and q_0 a prime power. Then the following family of hash functions $H=\{h_{a,b}\,;\,h_{a,b}(x)=\pi(ax)+b\}$, where $a,x\in\mathbb{F}_Q, b\in\mathbb{F}_q$ is $\epsilon-ASU_2$ with $\epsilon=1/q$.*
Consequently $m_A(\frac{1}{q_0^m},q_0^m,q_0^n)\le q_0^{m+n}$ for every prime-power q_0 and $n\ge m$.

Proof. Proof follows from Theorem 11 below. □

Remark 1: We obtain the same ASU_2's if we take the family of orthogonal arrays constructed in [12, page 363].
Remark 2: The case n=m=1 stems from a 2-dimensional Reed-Solomon code.
Remark 3: This lemma can be generalized as is done in Theorem 11. The first author gave a generalization via orthogonal arrays $OA_{q^{(t-1)(n-m)}}(t,q^n,q^m)$, with $t\ge 2$.

3 The evaluation of parameters of Wegman&Carter's construction

Wegman and Carter proposed in [4] the following method for constructing an authentication code. Let A be a set binary words of length a' and B a set of binary words of length b'. Divide a word $a\in A$ in segments of length s, where the parameter s will be chosen later in a proper way, and apply to each segment some (but the same) hash function from family $\widetilde{H}$ of [7, Theorem 5.2], where $q=2^s$. As a result one has again binary words but only halve as long. Repeat this procedure

ν times with arbitrary hash functions $h_1, ..., h_\nu$ until we get a word of length s. And the last step consists in taking some (e.g. the low-order) b' bits from this word. Wegman&Carter and later Stinson investigated the important parameters of the A-code thus obtained: the probability of successful impersonation[5] P_I, the probability of successful substitution P_S and the size of the key (logarithm of the number of hash functions). Wegman&Carter's construction was interesting because of their basic observation that by increasing P_S beyond P_I ($P_S \leq 2 \cdot P_I$ say), the source space can be dramatically enlarged. In fact, in [8], it was shown that when $P_S > P_I$ the source space grows exponentially in the key size! In what follows we always have $P_I = 2^{-b'}$.

For our purpose it is more convenient to represent the construction by two "stages". The first stage consists of $\nu - 1$ concatenations. As a result of this stage we have a family of hash functions from A to B^*, where B^* is a set of binary words of length $2s$. Or, in other words, we have q^*-ary code ($q^* = 2^{2s}$), which is a result of $\nu - 1$ concatenations. According to Theorem 3 we have got the following inequality for the corresponding value of ϵ for this code $\varepsilon_1 = \epsilon^* \leq 1 - (1 - \widetilde{\epsilon})^{\nu-1}$ where $\widetilde{\epsilon} = 1/q$, see [7, Theorem 5.2].

The second stage consists of application of a hash function from $\widetilde{H}$ to a $2s$-bit word and then taking b'-bits from the resulting word. The performance of such family of hash functions is given by Lemma 10.

Combining these two stages we have got exactly(!) the parameters of the Wegman&Carter construction. Namely $\epsilon = \epsilon_1 + \epsilon_2 - \epsilon_1 \cdot \epsilon_2$, where $\epsilon_1 = 1 - (1 - (1/2^s))^{\nu-1}$, $\epsilon_2 = 1/2^{b'}$, $\nu = \log_2 a'/s$. The number of hash functions, n, (or length of the corresponding code) equals $2^{(\nu-1)3s+2s+b'}$ (this slightly better than in the original paper as the authors used a rough estimate, i.e., Q^2 instead of Qq as we have from Lemma 10). Thus we can confirm the correctness of [4] and refute the remark in [7, page 83].

Example 2. (see [7]). Let $s = 23$, $b' = 20$, $\nu = 7$. Then the W&C construction gives $a' = 23 \cdot 2^7$, $\epsilon_1 \leq (3/4)2^{-20}$, $\varepsilon \leq \varepsilon_1 + \varepsilon_2 \leq 2^{-19}$, and the number of hash functions equals 2^{480}. □

The disadvantage of the original W&C construction is the usage of A-codes within the first stage as it is enough to use only ordinary codes. This observation immediately leads us to replacing the A-codes of [7, Theorem 5.2] by codes of [7, Theorem 5.1]. It decreases the number of function to $2^{(\nu-1)s+2s+b'}$ without decreasing the final ϵ. In particular, one gets 2^{204} as the number of hash functions for the considered example, like in [6]. However we can do better as we will show in the next section.

4 Construction of families of hash functions via RS codes

Before we describe our construction we first prove the following theorem:

[5] Stinson uses here P_{d_0} and P_{d_1}. We keep the original notation of Simmons [3].

Theorem 11. *Let Q, q, p, π be the same as in Lemma 10. Then the following family of hash functions*

$$H = \{h_{x,y} : h_{x,y}(a_1, ..., a_k) = \pi(xa_1 + x^2a_2 + ... + x^ka_k) + y \\ \text{where } x, a_1, a_2, ..., a_k \in \mathbb{F}_Q \text{ and } y \in \mathbb{F}_q\}$$

is $\epsilon - ASU_2$ with $\epsilon = k/q$ and $|A| = Q^k$, $|H| = qQ$. Thus we have $m_A(k/q, q, Q^k) \leq qQ$.

Proof. It is clear that for any $a \in \mathbb{F}_Q, z \in \mathbb{F}_q$, the number of hash functions $h : h(a) = z$ is the same and equals n/q, where $n = |H|$. We now calculate the maximal number of hash functions such that $h(a) = z, h(b) = z'$, where $a, b \in \mathbb{F}_q^k$, $z, z' \in \mathbb{F}_q$. Saying in other words, we are interested in the evaluation of the maximal number of solutions of the corresponding system of two algebraic equations. This system is equivalent to the following system $h(a) = z, \pi(c_1x + c_2x^2 + ... + c_kx^k) = w$, where $c = a - b, w = z - z'$. According to Bezout's Theorem the number of solutions of the second equation is not greater than $k|\text{Kern}\pi|$, where $|\text{Kern}\pi| = |\{u \in \mathbb{F}_q ; \pi(u) = 0\}| = Q/q$. □

Remark 1: For $k = 1$ one has Lemma 10. This theorem easily gives some of the results found in [6].

Remark 2: This construction can be explained in coding theoretic language starting with Reed-Solomon codes.

A natural application of RS-codes is their concatenation as inner codes together with ASU_2-codes of Lemma 10.

Proposed construction: We propose to construct $\epsilon - ASU_2$ classes of hash functions for authentication in the following way: Concatenate an $\epsilon_1 - AU_2$ class which is obtained from an RS-code over $\mathbb{F}_Q$ with an $\epsilon_2 - ASU_2$ class from Lemma 10. According to Theorem 5 we get an $(\epsilon_1 + \epsilon_2) - ASU_2$ class.

In detail, it can be described as follows: Let $q = 2^r$ and $Q = 2^{r+s}$. Choose an RS-code over $\mathbb{F}_Q$ with $n = Q$ and $k = 1 + 2^s$. The size of the message space is $|M| = Q^{1+2^s} = 2^{(r+s)(1+2^s)}$. This is the $\epsilon_1 - AU_2$ class and $\epsilon_1 = 1 - d/n = 1 - (2^{r+s} - 2^s)/2^{r+s} = 1/2^r$. From Lemma 10 we have an $\epsilon_2 - ASU_2$ class from $\mathbb{F}_Q$ to $\mathbb{F}_q$, with $\varepsilon_2 = 1/2^r$. The concatenation of these two gives the desired $\epsilon - ASU_2$ class, where $\epsilon \leq 2/2^r$. The size of the key space is then Q^2q. Note there is only one (!) concatenation in this construction. Note also that this works in any characteristic. The result is

$$m_A(\frac{2}{q^r}, q^r, q^{(r+s)(1+q^s)}) \leq q^{3r+2s}.$$

Example 9. Let us show how the construction works by giving a numerical example for the case considered in Example 2. Take a Q-ary RS-code with $Q = 2^{27}, k = 1 + 2^7 = 129$. This is an AU_2-code with $\epsilon = 2^{-20}$. Application of the concatenation construction with codes, the Lemma with $Q = 2^{27}$, $q = 2^{20}$ gives an ASU_2 code with $|A| = 2^{27 \cdot 129}$, $|H| = 2^{74}$, $\epsilon \leq 2^{-19}$. This is case $q = 2, r = 20$, $s = 7$ above. □

5 The use of geometric codes

We want to show how more sophisticated classes of linear codes, in particular of codes defined on algebraic curves, may be used to improve Stinson's bound considerably (see Example 1). It is natural in our context to use the machinery of geometric codes in the following form:

Theorem 12 (Canonical construction). *Let $q \geq 9$ be a quadratic prime power and let K be a function field of transcendence-degree 1 (equivalently: an algebraic curve) over the field $\mathbb{F}_q$ of constants, $P_0, P_1, \ldots P_n$ rational points of K. Consider the divisors $D = P_1 + P_2 \ldots + P_n, G = mP_0$. Let $m_1 = 0, m_2, \ldots, m_k, \ldots$ be the pole-orders of P_0. Consider the code*

$$C_k = C(D, m_k P_0)$$

of functions which are everywhere holomorphic except for a pole of degree $\leq m_k$ at P_0, evaluated at $P_1, \ldots P_n$ (this is the L-construction of [13]). Then C_k has dimension k and minimum distance $\geq n - m_k$. Hence

$$m(\frac{m_k}{n}, q, q^k) \leq n.$$

If moreover $m_k - 1$ is a Weierstraß gap, then

$$m(\frac{m_{k-1}}{n}, q, q^k) \leq n.$$

We need curves with many rational points and at least one rational Weierstraß-point whose gaps are as large as possible. In fact, Reed-Solomon codes result from the canonical construction when applied to the rational curve. In [14] and [15] a class $K_q^{(r)}$ of function fields defined over an arbitrary finite field $\mathbb{F}_q$ of constants is studied, where $r \geq 2$. Here $K_q^{(r)}$ is a tower of Artin-Schreier extensions of the rational function field. The following facts are to be found in [15]: The number N_1 of rational points of $K_q^{(r)}$ is $N_1 = q^r + 1$. There is a rational Weierstraß-point P_0 whose semigroup of pole-orders is

$$\sum_{i=1}^{r} q^{r-i}(q+1)^{i-1}\mathbb{N}_0.$$

This yields improvements upon the Stinson-bound valid for all sufficiently large prime-powers. In fact we can get a precise asymptotic statement. Upon using a well-known inequality between binomials and the binary entropy-function H:

$$2^{mH(l/m)}/(m+1)^2 \leq \binom{m}{l} \leq 2^{mH(l/m)}$$

(see [16]) the following is obtained:

Theorem 13. *Let q_0 be the unique positive solution of the equation*

$$H(q_0) = q_0.$$

For every $\epsilon > 0$ and sufficiently large i we have

$$m(\frac{i}{q}, q, q^{2^i}) \leq q^r,$$

where $r = \lfloor (i-1)(1-q_0)/q_0 - \epsilon \rfloor$ and q is an arbitrary prime-power, $q > (i-1)(r-1)$.

The numerical values are

$$q_0 = .7729\ldots, (1-q_0)/q_0 \approx .2938$$

We note that the same number q_0 appears in the theory of *Sperner capacity*, a recently discovered extension of the concept of *Shannon capacity* of a graph (see [17]).

For small values of i and a quadratic ground-field we obtain improvements by means of *Hermitian codes*. Consider the *Hermitian curve* defined by the equation $X^{q+1}+Y^{q+1}+Z^{q+1}$ over the field $\mathbb{F}_{q^2}$ of constants. This curve has genus $\binom{q}{2}$ and q^3+1 rational points. These form the well-known Hermitian unital. They are all Weierstraß-points. The semigroup of pole-orders of any of them is $q\mathbb{N}_0+(q+1)\mathbb{N}_0$. In particular the integers between wq and w(q+1) are pole orders. Let us call w the *weight* of such a pole-order. If $w < q$, then a pole order of weight w doesn't have any other weight. The number of pole orders of weight $\leq w$ is then $1+2+\ldots+(w+1) = \binom{w+2}{2}$.

Lemma 14. *Let (m_k) be the pole orders of the Hermitian curve over $\mathbb{F}_{q^2}$. If $w < q$, then*

$$m_{\binom{w+2}{2}} = w(q+1),$$

$$m_{\binom{w+1}{2}+1} = w \cdot q > m_{\binom{w+1}{2}} + 1.$$

We may use the construction of the preceding section and replace the RS-codes by Hermitian codes. If we choose $k = q^s + 1$ and use the canonical construction in its strengthend form, the following is obtained:

Example 4. We want bounds on $m_A(2^{-19}, 2^{20}, 2^{2^{28}})$. This is case $q = 2$, $r = 20$ above. We have

$$m_A(2^{-19}, 2^{20}, 2^{2^{28}}) \leq m(\frac{2^{28}}{q^3}, q^2, 2^{2^{28}}) \cdot m_A(2^{-20}, 2^{20}, 2^{32}),$$

by Lemma 8, where $q = 2^{16}$. The second factor above is bounded by 2^{52} (Lemma 10). Use the canonical construction for the Hermitian curve, where $k = 2^{28}/32 =$

2^{23}, $w = 2^{12} - 1$. As $\binom{w+2}{2} > k$, it follows from Lemma 14 that $m_k < w(q+1) = (2^{12}-1)(2^{16}+1) < 2^{28}$. By the canonical construction

$$m(\frac{2^{28}}{q^3}, q^2, 2^{2^{28}}) \leq m(m_k/q^3, q^2, (q^2)^k) \leq q^3.$$

Thus

$$m_A(2^{-19}, 2^{20}, 2^{2^{28}}) \leq q^3 2^{52} = 2^{100}$$

and we may thus choose $s = 12$. □

In characteristic 2 we get further improvements by using a family of curves which admit the Suzuki groups as automorphism groups. This family is studied in [18]. Let $q = 2^{2f+1}, q_0 = 2^f$. The curve is defined over $\mathbb{F}_q$ by the homogeneous equation

$$X^{q_0}(Z^q + ZX^{q-1}) = Y^{q_0}(Y^q + YX^{q-1}),$$

has q^2+1 rational points and a Weierstraß-point whose semigroup of pole-orders is

$$q\mathbb{N}_0 + (q+q_0)\mathbb{N}_0 + (q+2q_0)\mathbb{N}_0 + (q+2q_0+1)\mathbb{N}_0.$$

The number of pole-orders of weight $w < q_0$ is $\binom{w+2}{2} + \binom{w+1}{2} = (w+1)^2$. Via the canonical construction we obtain:

Theorem 15. *Let $q = 2^{2f+1} \geq 128$. Then*

$$m(\frac{i}{q}, q, q^{2^i}) \leq q^2 \ (i = 3, 4, 5, 6, 7).$$

If we use Suzuki codes in the same spirit as we used RS-codes and Hermitian codes above, we get

$$m_A(\frac{1}{2^{r-1}}, 2^r, 2^{(r+s)\{1+(2^s+1)(2^s+2)(2^{s+1}+3)/6\}} \leq 2^{4r+3s} \text{ if } s < r, s+r \text{ odd}.$$

The last statement of Theorem 12 follows from a recent result of A. Garcia et.al. ([19]). We give an application of this strengthened form of the canonical construction when applied to Hermitian curves:

Theorem 16. *Let q be a quadratic prime-power. Then*

$$m(\frac{1}{q} + \frac{1}{q^{3/2}}, q, q^4) \leq q^{3/2}.$$

Observe that there is not much of a difference between probabilities $\frac{1}{q}$ and $\frac{1}{q} + \frac{1}{q^{3/2}}$. For practical purposes the statement above should therefore be interpreted as

$$m(\approx \frac{1}{q}, q, q^4) \leq q^{3/2}.$$

It is natural to conjecture that all the Deligne-Lusztig curves will yield good codes and good classes of hash functions. In the case of the Ree curves we have not yet been able to verify this as the Weierstraß-points and their pole orders seem to be unknown.

6 Numerical Results

In order to illustrate our results we proceed as in Section 6 of [6]: Let $|A| = 2^{a'}$, $|B| = 2^{b'}$. This means that we have an a'-bit source and we want to use b'-bit authenticators. The cases we tabulate include those given in [6], with improved values for the necessary length of key. Here $P_S = 2^{-19}$. In most cases the AU_2 class is produced by Reed-Solomon codes (Section 4). Only in the case $a' = 2^{28}$, $b' = 20$ we use a Hermitian code, Example 4, where we get 100 instead of 106 bits of key!

length of source	length of authenticator		length of key	
a'	b'	s	new	[6]
2^8	20	24	68	135
2^{12}	20	28	76	236
2^{16}	20	32	84	332
2^{20}	20	35	90	445
2^{24}	20	39	98	-
2^{28}	20	32	100	-
2^8	40	43	126	255
2^{12}	40	47	134	346
2^{16}	40	51	142	612
2^{20}	40	55	150	805

Table 1. Table with source and key size for A-codes with $P_I = 2^{-b}$ and $P_S \leq 2 \cdot P_I$.

We can compare these results with a lower bound based on the q-twisting technique and the Varshamov-Gilbert bound which gives a Varshamov-Gilbert type bound for A-codes. This bound is an existence result and tells us that there exist A-codes with given P_I and $P_S \geq P_I$ with a certain number of source states (as function of the number of keys). For example, using the results of another paper, [20], which discusses this in more detail we have that for the situation in Examples 2 and 3 with $P_I = 2^{-20}$ and $P_S \leq 2P_I$ there exist classes which require only 52 bits for the key size.

7 Conclusion

We have shown that using coding theory we can easily reformulate and prove results in [4] and [7]. In particular, the concepts of geometric codes and concatenated codes gives us powerful tools. We also gave a simple analysis of the Wegman&Carter construction by our approach and suggested some improvements. Finally the idea of concatenation was used to get a new type of construction which requires considerably less key size than previously known. The construction using RS-codes is surpassed by the one using AG-codes for very

large source sizes. In our table this happened when we authenticate 33MB(yte) source strings. Further development on algebraic geometry might improve this in favor of the AG-codes.

A Proof of Theorem 5

Proof. We have a Q-ary code ($Q = |B_1|$) V of length n_1 ($n_1 = |H_1|$) and with distance $d = (1-\varepsilon_1)n_1$ and a q-ary code W of size Q ($|H_2| = Q$) of length n_2 with the special property

$$\forall w \neq w' \in W, \forall \alpha, \beta \in \{0,1,\ldots,q-1\} \quad |\{i\,;\, w_i = \alpha, w'_i = \beta\}| \leq \varepsilon_2 \frac{n_2}{q}$$

and

$$\forall w \in W, \forall \alpha \in \{0,1,\ldots,q-1\} \quad |\{i\,;\, w_i = \alpha\}| = \frac{n_2}{q}.$$

We form the concatenated code C from V and W by replacing symbols from the Q-ary alphabet in the codewords of V by the corresponding codewords from code W. Let us now compute

$$\left|\left\{j\,;\, c_j = \alpha, c'_j = \beta\right\}\right|, \quad \alpha, \beta \in \{0,1,\ldots,q-1\},$$

where $\underline{c} = \phi(\underline{v}), \underline{c'} = \phi(\underline{v'})$,

$$\phi(\underline{v}) = (\rho(v_1), \rho(v_2), \ldots, \rho(v_{n_1}))$$

and where

$$\rho : \{0,1,\ldots,q-1\} \to W$$

is a bijective map. The index j of c_j can be considered as a pair (j_1, j_2), where

$$0 \leq j_1 < n_1, 0 \leq j_2 < n_2.$$

Let us first consider the case $\alpha = \beta$. Then the set $\left\{j\,;\, c_j = c'_j = \alpha\right\}$ consists of all $j = (j_1, j_2)$ such that

a) $v_{j_1} = v'_{j_1}$ and $\rho(v_{j_1})_{j_2} = \alpha$

b) $v_{j_1} \neq v'_{j_1}$ and $\rho(v_{j_1})_{j_2} = \rho(v'_{j_1})_{j_2} = \alpha$.

We have that the number of indices satisfying a) is $(n_1 - d(\underline{v}, \underline{v'}))\,\frac{n_2}{q}$. The number of indices satisfying b) is $\leq d(\underline{v}, \underline{v'})\varepsilon_2 \frac{n_2}{q}$. Since a) and b) count disjoint situations the total number satisfies

$$\begin{aligned}\left|\left\{j\,;\, c_j = c'_j = \alpha\right\}\right| &\leq \frac{n_2}{q}(n_1 - d + d\varepsilon_2) \\ &= \frac{n_1 n_2}{q}(1 - (1-\varepsilon_1)(1-\varepsilon_2)) \\ &= \frac{n_1 n_2}{q}(\varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2).\end{aligned}$$

For the case $\alpha \neq \beta$ we can not have situation a) but only b) and the contribution is less for this case. Thus we have proved that $\varepsilon \leq \varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2$. □

References

1. J.L. Carter, M.N. Wegman, "Universal Classes of Hash Functions", *J.Computer and System Sci.*, Vol. 18, 1979, pp. 143-154.
2. D.R. Stinson, *Combinatorial techniques for universal hashing*, University of Nebraska-Lincoln. Department of Computer Science and Engineering, 1990.
3. G.J. Simmons, "A survey of Information Authentication", in *Contemporary Cryptology, The science of information integrity*, ed. G.J. Simmons, IEEE Press, New York, 1992.
4. M.N. Wegman, J.L. Carter, "New hash functions and their use in authentication and set equality", *J. Computer and System Sciences*, Vol. 22, 1981, pp. 265-279.
5. C.H. Bennett, G. Brassard, J-M. Roberts, " Privacy amplification by public discussion", *SIAM J.Comput.*, Vol. 17:2, 1988, pp. 210-229.
6. D.R. Stinson, "Universal Hashing and Authentication Codes", to appear in IEEE Transactions on Information Theory. This is a final version of [7].
7. D. R. Stinson, "Universal hashing and authentication codes" *Proceedings of Crypto 91*, Santa Barbara, USA, 1991, pp. 74-85.
8. T. Johansson, G. Kabatianskii, B. Smeets, "On the relation between A-codes and codes correcting independent errors" *Proceedings Eurocrypt'93*, to appear.
9. J. Bierbrauer, "Universal hashing and geometric codes", manuscript.
10. G.D. Forney, Jr., *Concatenated Codes*, M.I.T. Press, Cambridge, MA., 1966.
11. J. Bierbrauer, "Construction of orthogonal arrays", to appear in *Journal of Statistical Planning and Inference.*
12. T. Beth, D. Jungnickel, H. Lenz, *Design Theory*, Bibliographisches Institut, Zürich 1985.
13. M.A. Tsfasman, S.G. Vlăduţ , *Algebraic-Geometric codes*, Kluwer Academic Publ., Dordrecht/Boston/London, 1991.
14. R. Pellikaan, B.Z. Shen, and G.J.M. van Wee, " Which linear codes are algebraic-geometric?", *IEEE Trans. Information Theory*, Vol. 37, 1991, pp. 583-602.
15. B.H. Matzat, "Kanonische Codes auf einigen Überdeckungskurven", *Manuscripta Mathematica*, Vol. 77, 1992, pp. 321-335.
16. L. Gargano, J. Körner, U. Vaccaro, "Sperner capacities", to appear in *Graphs and Combinatorics.*
17. L. Gargano, J. Körner, U. Vaccaro, "Capacities: from information theory to extremal set theory", to appear in *Journal of the AMS.*
18. J.P. Hansen, H. Stichtenoth, "Group Codes on Certain algebraic curves with many rational points",*AAECC*, Vol. 1, 1990, pp. 67-77.
19. A. Garcia, S.J. Kim, R.F. Lax, "Consecutive Weierstrass gaps and minimum distance of Goppa codes", *Journal of Pure and Applied Algebra*, Vol. 84, 1993, pp. 199-207.
20. G. Kabatianskii, B. Smeets, T. Johansson, "Bounds on the size of a-codes and families of hash functions via coding theory", manuscript.