Universal hashing and geometric codes

Jürgen Bierbrauer
Mathematisches Institut der Universität
Im Neuenheimer Feld 288
69120 Heidelberg
Germany

December 9, 1994

Abstract

We describe a new application of algebraic coding theory to universal hashing and authentication without secrecy. This permits to make use of the hitherto sharpest weapon of coding theory, the construction of codes from algebraic curves. We show in particular how codes derived from Artin-Schreier curves, Hermitian curves and Suzuki curves yield classes of universal hash functions which are substantially better than those known before.

1 Introduction

The concept of universal classes of hash functions was introduced by Carter and Wegman in [6]. It has found numerous applications. We mention cryptography, complexity theory, search algorithms and associative memory (see the Introduction in [16]). Most important are the following classes (see [17]).

Definition 1 *Let $\epsilon > 0$. A multiset Σ of b functions from a k-set C to a v-set E is $\epsilon-$*almost strongly universal$_2$ *(short: ASU_2) if*

1. *for every $u \in C$ and $x \in E$ the number of elements of Σ mapping $u \mapsto x$ is b/v,*

2. *for every pair $u_1, u_2 \in C, u_1 \neq u_2$, and every pair $x_1, x_2 \in E$ the number of elements of Σ affording the operation $u_1 \mapsto x_1, u_2 \mapsto x_2$ is $\leq \epsilon \cdot b/v$.*

Definition 2 *Let $\epsilon > 0$. A multiset Σ of b functions from a k-set C to a v-set E is $\epsilon-$*almost universal$_2$ *(short: AU_2) if for every pair $u_1, u_2 \in C, u_1 \neq u_2$ the number $\delta(u_1, u_2)$ of elements $f \in \Sigma$ such that $f(u_1) = f(u_2)$ satisfies*

$$\delta(u_1, u_2) \leq \epsilon \cdot b.$$

We use the language of *arrays,* just as in [3]. Thus a (C, E)-array Σ consists of a set C of columns, $|C| = k$, a set E of entries, $|E| = v$, and is a multiset of cardinality b of mappings $C \longrightarrow E$. The elements of Σ are written as rows of the array.

The notion of ASU_2 is clearly a generalization of orthogonal arrays of strength 2. In fact, assume equality always holds in condition 2. of Definition 1. Then the number λ of elements of Σ affording the operation $u_1 \mapsto x_1, u_2 \mapsto x_2$ is $\lambda = \epsilon \cdot b/v$. It is clear that $b = v^2\lambda$ in this case. Thus $\epsilon = 1/v$ and Σ is an orthogonal array of strength 2. This description of ASU_2-classes as generalizations of orthogonal arrays of strength 2 has been observed in earlier work. It seems however to have escaped attention that AU_2-classes admit an even neater description: the columns of an AU_2-class of hash functions form the words of an error-correcting v-ary code. This is an equivalent description of AU_2-classes:

Lemma 1 *Let $\epsilon > 0, |C| = k, |E| = v, \Sigma$ a (C, E)-array of n functions from C to E. Then the following are equivalent:*

- *Σ is an $\epsilon - AU_2$ class of hash functions.*
- *The columns of Σ form the words of a v-ary code of length n with minimum distance d, where*

$$1 - \frac{d}{n} \leq \epsilon.$$

The proof is trivial, but the fact is important. The theory of AU_2-classes is nothing but the theory of error-correcting codes. This shows that the machinery of coding theory may be used to produce classes of hash functions. Of central importance is the composition-construction of Stinson's (see [17]).

Definition 3 *Let* Σ_1 *be a* $(C, E_1)-$ *array with* b_1 *and* Σ_2 *an* $(E_1, E)-$*array with* b_2 *rows. The* composition $\Sigma = \Sigma_2 \circ \Sigma_1$ *is the* $(C, E)-$*array with* $b_1 \cdot b_2$ *rows whose elements are all the compositions* $f_2 \circ f_1$*, where*
$f_1 \in \Sigma_1, f_2 \in \Sigma_2$.

In the new notation Stinson's composition-method for ASU_2-classes looks as follows:

Theorem 1 (Stinson) *If* Σ_2 *is an* $\epsilon_2 - ASU_2$ *class with* k_1 *columns and* v *entries and* Σ_1 *is an array whose columns form the words of a* k_1-*ary code of length* n *with* k *code words and minimum distance* d *such that* $\frac{d}{n} \geq 1 - \epsilon_1$*, then the composition* $\Sigma = \Sigma_2 \circ \Sigma_1$ *is an* $\epsilon - ASU_2$ *with* k *columns and* v *entries, where*

$$\epsilon = \epsilon_1 + (1 - \epsilon_1)\epsilon_2 < \epsilon_1 + \epsilon_2.$$

The composition-construction needs two types of ingredients. For the ASU_2-classes we may use orthogonal arrays. We make use of a family constructed in [1],p.363:

Theorem 2 *Let* q *be a prime-power and let* m, n *be natural numbers,* $m \geq n$*. Then there is an*

$$OA_{q^{m-n}}(2, q^m, q^n).$$

In [17] the construction of this family is simplified. The present author extended the construction to general t ([2]).
Consider the second ingredient, the error-correcting code. We fix the ground-field $\mathbb{F}_q$ and the relative minimum distance d/n of a q-ary code. In fact the minimum distance has to be extremely large, as $\epsilon = 1 - \frac{d}{n}$ plays the role of a probability. For a fixed number M of code-words we ask for the minimum length of such a code.

Definition 4 *Let natural numbers* $v, M,$ *and the real number* $\epsilon, 0 < \epsilon < 1$ *be given. Define* $m(\epsilon, v, M)$ *as the minimum length* n *of a* v*-ary code with* M *codewords and minimum distance* d *satisfying*

$$d/n \geq 1 - \epsilon.$$

This is a somewhat unusual question in coding theory. Equally unusual is the fact that we are only interested in q-ary codes with relatively large q. Binary codes are not interesting at all in our context. Here is the corresponding notion for the ASU_2-classes:

Definition 5 *Let natural numbers v, M, and the real number $\epsilon, 0 < \epsilon < 1$ be given. Define $m_A(\epsilon, v, M)$ as the minimum number of functions of an $\epsilon - ASU_2$ class of hash functions from an M-set to a v-set.*

With this terminology Stinson-composition looks as follows:

Theorem 3 (Stinson-composition)

$$m_A(\epsilon_1 + \epsilon_2, v, k) \leq m_A(\epsilon_1 + (1 - \epsilon_1)\epsilon_2, v, k) \leq m(\epsilon_1, k_1, k) \cdot m_A(\epsilon_2, v, k_1).$$

2 The use of geometric codes

In order to make Stinson-composition efficient in the construction of ASU_2-classes of hash-functions, bounds on the function $m(\epsilon, q, q^k)$ are needed. The monotonicity-properties of the function m follow from the definition:

$$\text{If } \epsilon_1 < \epsilon_2, \text{then } m(\epsilon_1, v, M) \geq m(\epsilon_2, v, M).$$

$$\text{If } v_1 < v_2, \text{then } m(\epsilon, v_1, M) \geq m(\epsilon, v_2, M).$$

$$\text{If } M_1 < M_2, \text{then } m(\epsilon, v, M_1) \leq m(\epsilon, v, M_2).$$

Here are some more basic properties of the function m :

$$m(\epsilon, v^i, M^i) \leq m(\epsilon, v, M).$$

This is Stinson's **Cartesian product-** construction (see [17]). In the case of linear codes we may describe this as **extension of constants**. **Concatenation** of codes yields

$$m(\epsilon_1 + \epsilon_2, v, M) \leq m(\epsilon_2, v, M_1) \cdot m(\epsilon_1, M_1, M).$$

In [13] the relation between ASU_2-classes and AU_2-classes of hash functions is studied. We note the following main results in our present notation:
If q is a prime-power, then

- $m(\epsilon, q, q(q-1)k+q) \leq m_A(\epsilon, q, k)$
- $m_A(\epsilon, q, (q-1)k+1) \leq q \cdot m_A(\epsilon, q, k)$

The 2-dimensional Reed-Solomon code shows $m(\frac{1}{q}, q, q^2) \leq q$. Using the cartesian product-construction and concatenation of codes recursively one obtains

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

for every prime-power q and every $i \geq 1$. This is a construction of Stinson's, expressed in different words. It is optimal in case $i = 1$, as remarked by Stinson. We want to show how more sophisticated classes of linear codes, in particular of codes defined on algebraic curves, may be used to improve this bound considerably, for all $i > 1$. It is natural in our context to use the machinery of geometric codes in the following form:

Theorem 4 (Canonical construction) *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 (equivalently: prime divisors of degree 1) of K. Consider the divisors $D = P_1 + P_2 \ldots + P_n, G = uP_0$. Let $u_1 = 0, u_2, \ldots, u_l, \ldots$ be the pole-orders of P_0 in ascending order. Consider the code*

$$\mathcal{C}_l = \mathcal{C}(D, u_l P_0)$$

of functions which are everywhere holomorphic except for a pole of degree $\leq u_l$ at P_0, evaluated at $P_1, \ldots P_n$ (this is the L-construction of [18]). Assume $u_l < n$. Then $\mathcal{C}_l$ has dimension l and the following hold:

- *$\mathcal{C}_l$ has minimum distance $\geq n - u_l$. Hence*

$$m(\frac{u_l}{n}, q, q^l) \leq n.$$

- *If $u_l - 1$ is a Weierstraß gap, then $\mathcal{C}_l$ has minimum distance $\geq n - u_{l-1}$. Hence*

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

The last statement is the main result of a recent paper of A. Garcia et.al. ([8]). It is sometimes useful in strengthening the results of the canonical construction.

We need curves with many rational points and at least one rational Weierstraß point whose gaps are as large as possible.

In [15] and [14] 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 [14]: 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

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

Let $\sum_{i=1}^r a_i q^{r-i}(q+1)^{i-1}$ be a pole-order of $P_0, a_i \geq 0$. Define for the moment the *weight* of this expression to be $w = \sum_{i=1}^r a_i$. It is easily seen that the weight w of a pole-order is uniquely determined if $w < \frac{q}{r-1}$. It is also easily seen that under this assumption the representation of each pole-order of weight w as a linear combination of the $q^{r-i}(q+1)^{i-1}$ is uniquely determined. The number of pole-orders of weight w is then the number of distributions of w undistinguishable objects in r cells. It is a basic combinatorial fact that this latter number is

$$\binom{w+r-1}{r-1}.$$

Assume $i - 1 < \frac{q}{r-1}$. Upon using the well-known combinatorial identity

$$\sum_{j=m}^{n} \binom{j}{m} = \binom{n+1}{m+1}$$

we see that the number of pole-orders of weight $< i$ is precisely

$$\sum_{w=0}^{i-1} \binom{w+r-1}{r-1} = \binom{i+r-1}{r}.$$

If $\binom{i+r-1}{r} \geq 2^i$, then $u_{2^i} < iq^{r-1}$, hence

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

Theorem 5 *Let q be a prime-power, i, r natural numbers such that*

$$q > (i-1)(r-1), \binom{i+r-1}{r} \geq 2^i.$$

Then

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

Here are some examples:

- *If $q > 8$, then*

$$m(\frac{5}{q}, q, q^{32}) \leq q^3.$$

- *If $q \geq 23$, then*

$$m(\frac{8}{q}, q, q^{256}) \leq q^4.$$

- *If $q \geq 49$, then*

$$m(\frac{12}{q}, q, q^{4096}) \leq q^5.$$

In fact we can get a precise asymptotic statement. Consider the binary entropy-function $H : (0,1) \longrightarrow \mathbb{R}_+$, where

$$H(x) = -x log(x) - (1-x) log(1-x).$$

Here the logarithm is binary. Observe $H(x) = H(1-x)$. The following inequality is well-known in information-theory, see [9]:

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

Hence asymptotically

$$\binom{i+r-1}{r} \sim 2^{(i+r-1)H(r/(i+r-1))}.$$

We fix i and ask for the smallest r satisfying $\binom{i+r-1}{r} \geq 2^i$. Asymptotically this is equivalent to

$$(i+r-1)H(\frac{r}{i+r-1}) \geq i$$

and to

$$H(\frac{r}{i+r-1}) \geq \frac{i}{i+r-1}.$$

Put $x = \frac{r}{i+r-1}$. Then $\frac{i}{i+r-1} \sim 1-x$. Let $q_0 > 0$ be the unique solution of the equation

$$H(q_0) = q_0.$$

We may choose r such that x is arbitrarily close to $1-q_0$. Observe $r = (i-1) \cdot \frac{x}{1-x}$. We conclude:

Theorem 6 *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. Denote by $L(n)$ the maximum cardinality of a family of pairs (A_i, B_i) of subsets of an n-set satisfying

$$A_i \cup B_i \not\subseteq A_j \cup B_j \ (i \neq j),$$

$$A_i \cap B_j = \emptyset \text{ if and only if i=j.}$$

Then $L(n) \sim 2^{q_0 \cdot n}$ asymptotically in n. This is proved in [10].

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

$$u_1 = 0, u_2 = q, u_3 = q+1, u_4 = 2q, \ldots u_8 = 3q+1, \ldots.$$

By choosing $l = 4$ and $l = 8$ in the canonical construction we get

$$m(\frac{2}{q^2}, q^2, q^8) \leq q^3,$$

$$m(\frac{3}{q^2} + \frac{1}{q^3}, q^2, q^{16}) \leq q^3$$

for every prime-power q.

Theorem 7 *Let $q = p^{2f}$ be a quadratic prime-power. Then*

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

$$m(\frac{3}{q} + \frac{1}{q^{3/2}}, q, q^8) \leq q^{3/2}.$$

Let us consider the distribution of pole-orders in greater detail. The $w+1$ integers between $w \cdot q$ and $w \cdot (q+1)$ are pole-orders. Let us call w the **weight** of such a pole-order. If $w \leq q$, then a pole-order of weight w doesn't have any smaller weight. The number of pole-orders of weight $\leq w$ is then $1 + 2 + \ldots + (w+1) = \binom{w+2}{2}$.

Lemma 2 *Let (u_l) be the pole-orders of the Hermitian curve over $\mathbb{F}_{q^2}$, in ascending order. If $w \leq q$, then*

$$u_{\binom{w+1}{2}+1} = w \cdot q$$

$$u_{\binom{w+2}{2}} = w \cdot (q+1).$$

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 [12]. 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, genus $q_0(q-1)$ and a Weierstraßpoint whose semigroup of pole-orders is

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

As before define the weight of the expression $a_1q + a_2(q+q_0) + a_3(q+2q_0) + a_4(q+2q_0+1)$ to be $w = \sum_i a_i$. Assume $w \leq q_0$. Then a pole-order of weight w cannot have any smaller weight. Assume it can be written as a linear combination like above in more than one way:

$$a_1q + a_2(q+q_0) + a_3(q+2q_0) + a_4(q+2q_0+1) =$$

$$= a'_1q + a'_2(q+q_0) + a'_3(q+2q_0) + a'_4(q+2q_0+1),$$

where $\sum a_i = \sum a'_i = w \leq q_0$. Put $x_i = a_i - a'_i$. Then

$$x_1q + x_2(q+q_0) + x_3(q+2q_0) + x_4(q+2q_0+1) = 0$$

and the sum of the coefficients is zero, hence

$$x_2q_0 + 2x_3q_0 + x_4(2q_0+1) = 0.$$

It follows $q_0 \mid x_4$. We claim $x_4 = 0$.
Otherwise $x_4 = \pm q_0$, and this is possible only if $w = q_0$. We may assume without restriction $x_4 = q_0$, hence $a_4 = q_0, a_3 = a_2 = a_1 = 0, a'_4 = 0$. Factoring out q_0 in the above equation yields

$$x_2 + 2x_3 + 2q_0 + 1 = 0.$$

As $x_2 = -a'_2, x_3 = -a'_3$ this shows $a'_2 + 2a'_3 = 2q_0 + 1$. Using the fact that $a'_1 + a'_2 + a'_3 = q_0$, we get by subtraction $a'_3 - a'_1 = q_0 + 1$, a contradiction.
We have proved $x_4 = 0$. It follows

$$x_2 + 2x_3 = 0.$$

We conclude that the only ambiguity in the representation of pole-orders of weight $w \leq q_0$ is the following: Coefficients (a_1, a_2, a_3, a_4) and $(a_1 + x, a_2 - 2x, a_3 + x, a_4)$ represent the same pole-order ($x \in \mathbb{Z}$). The representation becomes unique by choosing $a_2 \in \{0, 1\}$. The same combinatorial fact used above shows that the number of pole-orders of weight $w \leq q_0$ is $\binom{w+2}{2} + \binom{w+1}{2} = (w+1)^2$. Let $s_i = 1^2 + 2^2 + \ldots + i^2$. It is well-known that $s_i = (2i+1)(i+1)i/6$. It follows that the number of pole-orders of weight $\leq w$ is s_{w+1}, if $w \leq q_0$.

Lemma 3 *Let (u_l) be the pole-orders of the Suzuki curve over $\mathbb{F}_q, q = 2^{2f+1}$, in ascending order,*

$$s_i = 1^2 + 2^2 + \ldots + i^2 = (2i+1)(i+1)i/6.$$

If $w \leq q_0 = 2^f$, then

$$u_{s_{w-1}+1} = w \cdot q$$
$$u_{s_w} = w \cdot (q + 2q_0 + 1).$$

We record some low-dimensional cases, which yield improvements on Stinson's construction:

Theorem 8 *Let $q = 2^{2f+1} \geq 128$. Then*

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

In reality the situation is even better. For example there is not much of a difference between probabilities $\frac{3}{q^2}$ and $\frac{3}{q^2} + \frac{1}{q^3}$. For practical purposes the second statement of Theorem 7 should therefore be interpreted as

$$m(\approx \frac{3}{q}, q, q^8) \leq q^{3/2} \text{ (q a square prime-power).}$$

The same situation occurs regularly when using the canonical construction. The Artin/Schreier-series yields examples:

$$m(\approx \frac{6}{q}, q, q^{64}) \leq q^3$$

$$m(\approx \frac{9}{q}, q, q^{512}) \leq q^4.$$

Here are some more applications of the strengthened form of the canonical construction:

Theorem 9 • *Let q be a quadratic prime-power. Then*

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

$$m(\frac{4}{q}+\frac{4}{q^{3/2}},q,q^{16})\leq q^{3/2}.$$

• *Let $q=2^{2f+1}\geq 128$. Then*

$$m(\frac{3}{q},q,q^{16})\leq q^2,$$

$$m(\frac{4}{q},q,q^{32})\leq q^2.$$

Proof: In the case of the Hermite-curve defined over $\mathbb{F}_{q^2}$ we have $u_3 = q+1, u_4 = 2q, u_{15} = 4q+4, u_{16} = 5q$. Apply the foregoing Theorem and change notation ($q^2 \mapsto q$). This yields the first statement.
In the case of the Suzuki-curves we have $u_{15} = 3q, u_{16} = 3q+q_0, u_{31} = 4q, u_{32} = 4q+q_0$. ■

3 The use of Deligne-Lusztig codes for the construction of universal hash classes

The starting point in the Deligne-Lusztig theory of ordinary representations of finite groups of Lie type is the study of a certain variety, the Deligne-Lusztig variety, associated to a connected reductive algebraic group G defined over a finite field. If G has Lie rank 1, then the variety is a projective curve. Thus there are algebraic curves associated with the 2-dimensional linear groups, the 3-dimensional unitary groups, the Suzuki groups and the Ree groups (see [5, 7, 11]). The corresponding codes obtained via the L-construction (see Theorem 4) will be called **Deligne-Lusztig codes.** The case of the groups $PGL_2(q)$ leads to the Reed-Solomon codes. Thus the RS-codes are special cases of Deligne-Lusztig codes. The Hermitian codes (associated with the groups $PGU_3(q^2)$) and Suzuki codes (associated with ${}^2B_2(q), q = 2^{2f+1}$) have been considered in the preceding section. In this

section we will make use of these codes to construct ASU_2-classes of hash functions with a small number of keys (hash functions). It is natural to conjecture that also the Ree curves will yield good codes and good classes of hash functions. We have not yet been able to verify this as the Weierstraß-points and their pole orders seem to be unknown.
We use Stinson-composition (Theorem 1) with the class of orthogonal arrays of Theorem 2 in the role of Σ_2. In our notation this yields

$$m_A(\frac{1}{q^m}, q^m, q^n) \leq q^{n+m} \ (n \geq m, q \text{ a prime-power}).$$

Our aim is to get good upper bounds on $m_A(\frac{2}{q^m}, q^m, q^N)$, where N is large with respect to m. This amounts to authenticating a source of $N \cdot log_2 q$ bits using $m \cdot log_2 q$ authenticator bits, with deception probability $\epsilon \leq \frac{2}{q^m}$. The deception probability can never be smaller than the reciprocal of the number of authenticators. Moreover, in the case of equality the ASU_2-classes of hash functions is an orthogonal array of strength 2. In that case the number of keys is very large:

$$m_A(\frac{1}{q^m}, q^m, q^N) \geq q^N(q^m - 1) + 1 > q^N.$$

It was Wegman&Carter's crucial observation (see [19]) that the number of keys (of hash functions) can be dramatically reduced when ϵ is increased just a little bit. We consider the case $\epsilon = \frac{2}{q^m}$ when ϵ is the double of the theoretical minimum.
We use the canonical construction (Theorem 4) in its standard form. As we took care to determine the pole-orders of the Hermitian and Suzuki curves, it suffices to collect the data given in the preceding section. In the case of Reed-Solomon codes things are easier yet. The results are as follows:

Theorem 10 *Let q be a prime-power. Then the following holds:*

- $m(\frac{k-1}{q}, q, q^k) = q$ *for every* $k \geq 2$.
- *If* $w \leq q$, *then*
$$m(\frac{w}{q^2}, q^2, q^{2\{\binom{w+1}{2}+1\}}) \leq q^3$$

- *If* $q = 2^{2f+1}, q_0 = 2^f \geq w$, *then*

$$m(\frac{w}{q}, q, q^{1+(2w+1)(w+1)w/6}) \leq q^2.$$

Here the items correspond to Reed-Solomon, Hermitian and Suzuki codes. It follows from the Singleton bound that we have equality in the first case. We make use of Stinson-composition in the form of Theorem 3. Let $\mathbb{F}_{q^n}, n \geq m$ be the field over which the code is defined. As $\epsilon_2 = 1/q^m$ we have to take care that $\epsilon_1 \leq 1/q^m$. The result is as follows:

Theorem 11 *Let q be a prime-power, m, n natural numbers. Then the following hold:*

1. $$m_A(\frac{2}{q^m}, q^m, q^{n(1+q^{n-m})}) \leq q^{2n+m}$$

   ($m \leq n$).

2. $$m_A(\frac{2}{q^m}, q^m, q^{n\{1+q^{n-m}(q^{n-m}+1)/2\}}) \leq q^{5n/2+m}$$

   ($m \leq n \leq 2m, q^n$ a square).

3. $$m_A(\frac{1}{2^{m-1}}, 2^m, 2^{n\{1+(2^{n-m+1}+1)(2^{n-m}+1)2^{n-m}/6\}}) \leq 2^{3n+m}$$

   ($m \leq n \leq 2m-1, n$ odd).

Let us compare the three methods (Reed-Solomon=RS,Hermite=H, Suzuki=S) of Theorem 11 in the binary case. We are thus looking for bounds on $m_A(\frac{1}{2^{m-1}}, 2^m, 2^N)$. Denote by log the binary logarithm, put $L = log(N)$. When using the RS-method, we have to determine $n = n_{RS}$ from the equation $L = log(n(1+2^{n-m}))$. This is approximately equivalent with $n_{RS} = L + m - log(n_{RS})$. In the case of the H- and S-method we get $n_H \approx \frac{1}{2}L + m + \frac{1-log(n_H)}{2}$ and $n_S \approx \frac{1}{3}L + m + \frac{log3-log(n_S)}{3}$, respectively. Here $m \leq n_H \leq 2m, m \leq n_S \leq 2m-1$. The H-method needs less bits than the RS-method as soon as

$4n_{RS} > 5n_H$. By substituting the above values of n_{RS} and n_H and putting $n_{RS} = \alpha \cdot n_H$ we see that this is equivalent with

$$L > \frac{2}{3}m + log(n_H) + \frac{5}{3} + \frac{8}{3}log(\alpha).$$

Certainly $log(m) \leq log(n_H) \leq 1 + log(m)$. Furthermore we have a priori bounds on α : in fact we are assuming $\alpha \geq 5/4$, and by equating the two expressions for N we get $n_{RS}2^{n_{RS}-m} \approx n_H 2^{2(n_H-m)-1}$, equivalently $\alpha \approx 2^{(2-\alpha)n_H-m-1} \geq 5/4$. It follows $(2-\alpha)n_H - m - 1 \geq log5 > 2$, hence $1.25 \leq \alpha < 1.5$. We see that the H-method needs less key than the RS-method as soon as

$$L > \frac{2}{3}m + log(m) + c_m,$$

where c_m is a small constant, approximately $2.4 < c_m < 4.3$. The H-method is best as long as it is applicable, i.e. up to $L = 2m + log(m)$. From then on the S-method is better.

What happens for even larger sources, when the S-method is no longer applicable in the form of Theorem 11, i.e. when $L > 3m + log(m)$? If we use the Artin-Schreier curves $K_q^{(r)}$ as introduced before Theorem 5 in the same spirit as Deligne-Lusztig curves were used in Theorem 11, then let us speak of the AS_r-method,$r \geq 3$ (the case $r = 2$ is in fact not interesting). Moreover we can use geometric codes for arbitrarily large dimensions. Instead of the canonical construction of Theorem 4 we use then the equation $u_l = l + 1 + g$, where g is the genus of the curve in question. It turns out that this variant is never best possible when applied to the Hermite curves or the AS_r-curves. In the case of Suzuki curves, let us call this the S_+-method. The following is the result of the (trivial) analysis when we compare all these methods in case $L > 3m + log(m)$. Here we omit the summand $log(m)$ as well as some small additive constant:

- If $3m \leq L \leq 14m$ (approximately), then the S_+-method is best.
- For $14m \leq L \leq 20m$ (approximately), method AS_4 is best.
- For larger L method $AS_r(r > 4)$ will be best among the constructions considered here for approximately $(r-1)r \cdot m \leq L \leq r(r+1) \cdot m$.

We close with some explicit examples, which are applications of Theorem 11. Here $m = 20, q = 2$:

Example 1 *1.* $m_A(2^{-19}, 2^{20}, 2^{2^{10}}) \leq 2^{72}$.

2. $m_A(2^{-19}, 2^{20}, 2^{2^{35}}) \leq 2^{110}$.

3. $m_A(2^{-19}, 2^{20}, 2^{2^{50}}) \leq 2^{125}$.

In the first case ($L = 10$), the RS-method is best. In fact $n_{RS} = 26, n_H = 22, n_S = 23$. The H-method needs $\frac{5}{2} \cdot 22 + 20 = 75$ bits, the S-method needs $3 \cdot 23 + 20 = 89$ bits.
In the second case ($L = 35$) we have $n_H = 36$, and the H-method is best. The RS- and S-method need 120 and 113 bits, respectively.
In case $L = 50$, the H-method is no longer applicable. We have $n_S = 35$. It follows that the S-method needs $3 \cdot 35 + 20 = 125$ bits, which is better than the RS-method ($n_{RS} = 64$, needs 148 bits).
We restate the last inequality in a more familiar language: It is possible to authenticate a $2^{50}-$bit source with 20 authenticator bits and 125 bits of key such that the probabilities of success of an impersonation- or substitution-attack are bounded by 2^{-19}.
Finally let us consider the practicality of the construction. For simplicity we concentrate on the case of Example 1,1. The source has size $2^{2^{10}}$. The use of Stinson composition implies that our family of hash functions is the composition of a family A of functions : $2^{2^{10}} \longrightarrow 2^{26}$ und a family B of functions : $2^{26} \longrightarrow 2^{20}$. For B we took the 2^{46} rows of the orthogonal array $OA_{64}(2, 2^{26}, 2^{20})$ of Theorem 2. More explicitly the elements of B are pairs (β, γ), where $\beta \in I\!\!F_{2^{26}}, \gamma \in I\!\!F_{2^{20}}$. The entry in row (β, γ), and column $x \in I\!\!F_{2^{26}}$ of the orthogonal array is $\Phi(\beta \cdot x) + \gamma$, where Φ is a fixed $I\!\!F_2-$linear mapping : $I\!\!F_{2^{26}} \longrightarrow I\!\!F_{2^{20}}$.
For A we took the coordinate functions of the 40-dimensional RS-code over $I\!\!F_{2^{26}}$. Thus each element of A is an element $\alpha \in I\!\!F_{2^{26}}$, and each source state is a polynomial $p(X)$ of degree ≤ 39 with coefficients in $I\!\!F_{2^{26}}$. The hash functions are triples (α, β, γ), where $\alpha, \beta \in I\!\!F_{2^{26}}, \gamma \in I\!\!F_{2^{20}}$. Thus $26+26+20 = 72$ random bits are needed to choose a hash function uniformly at random. The corresponding hashed value is

$$\Phi(p(\alpha) \cdot \beta) + \gamma \in I\!\!F_{2^{20}}.$$

We need field arithmetic in $I\!\!F_{2^{26}}$, but only addition in $I\!\!F_{2^{20}}$. The field $I\!\!F_{2^{26}}$ should be seen as an extension of $I\!\!F_2$. The mapping Φ may be chosen as projection onto the first 20 coordinates, say.

Conclusion

We have seen that the theory of almost universal (AU_2-) classes of hash functions is equivalent with the theory of error-correcting codes. The mechanism of geometric codes imposes precise conditions on algebraic curves to yield AU_2- classes of hash functions which can be efficiently used as one of the two ingredients in Stinson's composition-construction for almost strongly universal (ASU_2-) classes of hash functions. Most importantly we saw that Deligne-Lusztig curves and certain Artin-Schreier curves allow the construction of ASU_2- classes of hash functions which use much less key space than the methods which had hitherto been used.
Part of the material of this paper was used in a joint publication with T.Johansson, G.Kabatianskii and B.Smeets [4] for CRYPTO 93. In particular the method of Theorem 11 is inspired by the use my coauthors in [4] make of Reed-Solomon codes.

References

[1] T.Beth,D.Jungnickel,H.Lenz: *Design Theory,* Bibliographisches Institut, Zürich 1985.

[2] J.Bierbrauer: *Construction of orthogonal arrays,* to appear in *Journal of Statistical Planning and Inference.*

[3] J.Bierbrauer,Y.Edel: *Theory of perpendicular arrays,* to appear in *Journal of Combinatorial Designs.*

[4] J.Bierbrauer,T.Johansson,G.Kabatianskii,B.Smeets: *On families of hash functions via geometric codes and concatenation,* Advances in *Cryptology-CRYPTO 93, Lecture Notes in Computer Science 773 (1994),331-342.*

[5] R.W.Carter: *Finite groups of Lie type,* Wiley 1985.

[6] J.L.Carter,M.N.Wegman: *Universal Classes of Hash Functions, J.Computer and System Sci. 18(1979), 143-154.*

[7] P.Deligne,G.Lusztig: *Representations of reductive groups over finite fields, Annals of Mathematics 103(1976),103-161.*

[8] A.Garcia,S.J.Kim,R.F.Lax: *Consecutive Weierstrass gaps and minimum distance of Goppa codes,* Journal of Pure and Applied Algebra *84(1993),199-207.*

[9] L.Gargano,J.Körner,U.Vaccaro: *Sperner capacities,* Graphs and Combinatorics *9(1993),31-46.*

[10] L.Gargano,J.Körner,U.Vaccaro: *Capacities: from information theory to extremal set theory,* to appear in *Journal of Combinatorial Theory A.*

[11] J.P.Hansen: Group codes on Deligne-Lusztig varieties, *Coding Theory and Algebraic Geometry, Proceedings Luminy 1991, Springer LNM 1518,1992.*

[12] J.P.Hansen,H.Stichtenoth: *Group Codes on Certain algebraic curves with many rational points,* AAECC 1(1990),67-77.

[13] T. Johansson, G. Kabatianskii, B. Smeets, *On the relation between A-codes and codes correcting independent errors,* Proceedings *Eurocrypt'93,1-11.*

[14] B.H.Matzat: *Kanonische Codes auf einigen Überdeckungskurven, Manuscripta Mathematica 77(1992),321-335.*

[15] R.Pellikaan,B.Z.Shen, and G.J.M. van Wee: *Which linear codes are algebraic-geometric? IEEE Trans.Inform.Theory 37(1991),583-602.*

[16] D.R.Stinson: *Combinatorial techniques for universal hashing,* to appear in *Journal of Computer and Systems Science.*

[17] D.R.Stinson: *Universal Hashing and Authentication Codes,* to appear in *IEEE Transactions on Information Theory.* A preliminary version appeared in the *Proceedings of CRYPTO 91, Lecture Notes in Computer Science 576 (1992), 74-85.*

[18] M.A.Tsfasman,S.G.Vladut : *Algebraic-Geometric codes,* Kluwer Dordrecht/Boston/London 1991.

[19] M.N.Wegman,J.L.Carter: *New hash functions and their use in authentication and set equality, J.*Computer and System Sci. *22(1981),265-279.*