Universal Hash Functions from Exponential Sums over Finite Fields and Galois Rings *

Tor Helleseth[1] and Thomas Johansson[2]

[1] Department of Informatics, University of Bergen,
HIB, N-5020 Bergen, Norway
[2] Department of Information Theory, Lund University,
Box 118, S-221 00 Lund, Sweden

Abstract. In this paper new families of strongly universal hash functions, or equivalently, authentication codes, are proposed. Their parameters are derived from bounds on exponential sums over finite fields and Galois rings. This is the first time hash families based upon such exponential sums have been considered. Their performance improves the previously best known constructions and they can be made general in their choice of parameters. Furthermore, the constructions are suitable both for hardware and software implementations. The latter is an aspect that is significant and has been considered in several recent papers.

Keywords. Universal hash functions, authentication codes, exponential sums, Galois rings.

1 Introduction

Universal hashing is a concept that was introduced by Wegman and Carter [19] in 1979. Since then, many results in theoretical computer science use different kinds of universal hashing.

One of the more interesting topics in universal hashing is named *strongly universal hashing*. Other names used are two-point based sampling, or pairwise independent random variables [18]. There is a large amount of applications of this topic in computer science. In cryptography we find applications in for example interactive proof systems. However, the most widely known application in cryptography is the construction of unconditionally secure authentication codes. We will return to the equivalence between strongly universal hash functions and authentication codes in Section 2.

The applications of exponential sums in coding theory have proved to be many, including bounds on the minimum distance and covering radius [6] of

* This research was partly done during a visit by the authors to the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, 1996. The first author was supported in part by The Norwegian Research Council under grant numbers 107542/410 and 107623/420. The second author was supported in part by NUTEK under grant number P5892-1.

N. Koblitz (Ed.): Advances in Cryptology - CRYPTO '96, LNCS 1109, pp. 31-44, 1996.

codes, as well as applications to sequence designs. Recently, a new direction in coding theory has been to apply the Gray map to codes that are linear over $\mathbb{Z}_4$ to obtain binary nonlinear codes better than comparable binary linear codes. The distance properties of these codes as well as the correlation properties of sequences obtained from $\mathbb{Z}_4$-linear codes depend on exponential sums over Galois rings.

It is well known that coding theory and universal hashing are closely related, and our aim is to explore how exponential sums over finite fields and Galois rings can be used to construct families of strongly universal hash functions. The results are positive and we obtain constructions that improve the previously best known constructions. These are not the only positive aspects. We also recognize that the constructions are simple to implement both in software and hardware. Such implementation aspects have recently been considered important, and there are several papers focusing on this topic [8], [9], and [14].

This paper is organized as follows. In Section 2 the basic definitions in authentication theory and in universal hashing are given, as well as the connection between them. Section 3 introduces exponential sums over finite fields, and in Section 4 we construct hash families over finite fields. In Section 5 we introduce exponential sums over Galois rings, and in Section 6 we construct hash families over Galois rings. We end with some concluding remarks.

2 Authentication codes and universal hash functions

Authentication theory as originally described by Simmons [15], [16], see also [4], considers the problem of two trusting parties, who want to send information from the transmitter to the receiver in the presence of an adversary. The adversary may introduce false messages to the receiver or replace a legal message with a false one. To protect against these threats, the sender and the receiver share a secret key. The key is then used in an authentication code (A-code).

A *systematic* (or Cartesian) A-code is a code where the information to be transmitted appears in plaintext in the transmitted message. Such a code is a triple $(\mathcal{S}, \mathcal{E}, \mathcal{Z})$ of finite sets and a map $f : \mathcal{S} \times \mathcal{E} \to \mathcal{Z}$. Here $\mathcal{S}$ is the set of source states, i.e., the information that is to be transmitted, $\mathcal{E}$ is the set of keys, and $\mathcal{Z}$ is the tag alphabet. When the transmitter wants to send the information $s \in \mathcal{S}$ using his secret key $e \in \mathcal{E}$, he transmits the message $m = (s, z)$, where $z = f(s, e)$, and $m \in \mathcal{M} = \mathcal{S} \times \mathcal{Z}$. When the receiver receives a message $m' = (s', z')$, he checks the authenticity by calculating whether $z' = f(s', e)$ or not. If equality holds, the message m is called valid. The adversary has two different attacks to choose between. He might introduce a false message $m = (s, z)$, and hence impersonating the transmitter, called the *impersonation attack*. He can also choose to observe a transmitted message $m = (s, z)$, and then replace this message with another message $m' = (s', z')$, where $s' \neq s$. This is called the *substitution attack*. The probability of success for the adversary when trying either of the two attacks, denoted by P_I and P_S respectively, are formally defined by $P_I = \max_{s,z} P(m = (s, z) \text{ valid})$ and

$P_S = \max_{s,z} \max_{s' \neq s,z'} P(m' = (s', z') \text{ valid} | m = (s, z) \text{ observed})$. We assume that the keys are uniformly distributed. Then these probabilities can be written as

$$P_I = \max_{s,z} \frac{|\{e \in \mathcal{E} : z = f(s,e)\}|}{|\{e \in \mathcal{E}\}|}, \tag{1}$$

$$P_S = \max_{s,z} \max_{s' \neq s,z'} \frac{|\{e \in \mathcal{E} : z = f(s,e), z' = f(s',e)\}|}{|\{e \in \mathcal{E} : z = f(s,e)\}|}. \tag{2}$$

For a review of different bounds and constructions of A-codes, we refer to [7], which gives a good account of the recent developments in the area.

In universal hashing, we consider a hash family $\mathcal{G}$, which is a set $\mathcal{G}$ of $|\mathcal{G}|$ functions such that $g : X \to Y$ for each $g \in \mathcal{G}$. Interesting parameters for a hash family are $|\mathcal{G}|$, $|X|$, and $|Y|$. Two relevant definitions are the following.

Definition 1. A hash family $\mathcal{G}$ is called *ϵ-almost universal$_2$* if for any two distinct elements $x_1, x_2 \in X$, there are at most $\epsilon|\mathcal{G}|$ functions $g \in \mathcal{G}$ such that $g(x_1) = g(x_2)$. We use the abbreviation ϵ-AU$_2$ for the family.

Definition 2. A hash family $\mathcal{G}$ is called *ϵ-almost strongly universal$_2$* if

i) for any $x \in X$ and any $y \in Y$, there are exactly $|\mathcal{G}|/|Y|$ functions $g \in \mathcal{G}$ such that $g(x) = y$.
ii) for any two distinct elements $x_1, x_2 \in X$, and for any two elements $y_1, y_2 \in Y$, there are at most $\epsilon|\mathcal{G}|/|Y|$ functions $g \in \mathcal{G}$ such that $g(x_1) = y_1$, and $g(x_2) = y_2$.

We here use the abbreviation ϵ-ASU$_2$.

For a more thorough treatment of universal hashing, we refer to [17], where these concepts are derived further. We will instead consider the known equivalences between strongly universal hashing and authentication codes.

Lemma 3 [1],[19],[17].

i) If there exists a q-ary code with codeword length n, cardinality M, and minimum Hamming distance d, then there exists an ϵ-AU$_2$ family of hash functions where $\epsilon = 1 - d/n$, $|\mathcal{G}| = n$, $|X| = M$, and $|Y| = q$. Conversely, if there exists an ϵ-AU$_2$ family of hash functions, then there exists a code with parameters as above.

ii) If there exists an A-code with parameters $|\mathcal{S}|$, $|\mathcal{E}|$, $P_I = 1/|\mathcal{Z}|$, and P_S, then there exists an ϵ-ASU$_2$ family of hash functions where $\epsilon = P_S$, $|\mathcal{G}| = |\mathcal{E}|$, $|X| = \mathcal{S}$, and $|Y| = |\mathcal{Z}|$. Conversely, if there exists an ϵ-ASU$_2$ family of hash functions, then there exists an A-code with parameters as above.

We review the equivalence ii) above. Each key $e \in \mathcal{E}$ in the A-code corresponds to a unique function g_e in $\mathcal{G}$, and $\mathcal{S} = X$. The tag z in the authentication code is then obtained as

$$z = g_e(s).$$

The significance of ϵ-AU$_2$ families in strongly universal hashing lies in the fact that they are very useful when constructing strongly universal hash families. This is due to the following result by Stinson.

Lemma 4 [17]. *Let $\mathcal{G}_1$ be ϵ_1-AU_2 from X_1 to Y_1 and let $\mathcal{G}_2$ be ϵ_2-ASU_2 from Y_1 to Y_2. Then $\mathcal{G} = \{g_2(g_1(x)) : g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$ is ϵ-ASU_2 with $\epsilon = \epsilon_1 + \epsilon_2$.*

Virtually all constructions of ϵ-ASU$_2$ families of hash functions for large $|X|$ use this composition construction. The constructions giving best performance [1] uses Reed-Solomon codes as the ϵ-AU$_2$ family in the above composition construction. Our aim in this paper is to derive even better ϵ-ASU$_2$ families of hash functions in a direct way, without using this composition construction. The idea of using exponential sums allows this to be done in a neat way.

3 Exponential sums over finite fields

Exponential sums have been an important tool in number theory for solving problems involving integers. Such sums can be considered in the framework of finite fields and turn out to be useful in various applications. For more details, see [11].

Let $Tr_{q^m/q}(\alpha)$ be the trace function from $\mathbb{F}_{q^m}$ to $\mathbb{F}_q$ defined by

$$Tr_{q^m/q}(\alpha) = \alpha + \alpha^q + \cdots + \alpha^{q^{m-1}}.$$

Furthermore, let $q = p^e$, where p is the characteristic of $\mathbb{F}_q$, and let $Tr_{q/p}(\alpha)$ be the trace function from $\mathbb{F}_q$ to $\mathbb{F}_p$. Note that $Tr_{q^m/p}(\alpha) = Tr_{q/p}(Tr_{q^m/q}(\alpha))$. Let ω be a complex primitive p-th root of unity. We will consider exponential sums of the form

$$\sum_{\alpha \in \mathbb{F}_{q^m}} \omega^{Tr_{q^m/p}(f(\alpha))},$$

where $f(x) \in \mathbb{F}_{q^m}[x]$. An important result on exponential sums of the above kind is the *Weil-Carlitz-Uchiyama bound* [3]. This is also the result that is the basis for our constructions.

Theorem 5 Weil-Carlitz-Uchiyama bound. *Let $f(x) = \sum_{i=1}^{D} f_i x^i \in \mathbb{F}_{q^m}[x]$, $q = p^e$, be a polynomial of degree D that is not expressible in the form $f(x) = g(x)^p - g(x) + \theta$ for any $g(x) \in \mathbb{F}_{q^m}[x]$, $\theta \in \mathbb{F}_{q^m}$. Then*

$$\left| \sum_{\alpha \in \mathbb{F}_{q^m}} \omega^{Tr_{q^m/p}(f(\alpha))} \right| \leq (D-1)\sqrt{q^m},$$

where ω is a complex primitive p-th root of unity.

4 Universal hash functions from exponential sums over finite fields

We start by considering a straightforward construction of strongly universal hash functions from exponential sums. We first need to prove a lemma.

Lemma 6. *Let* $f(x) = \sum_{i=1}^{D} f_i x^i \in \mathbb{F}_{q^m}[x]$ *be a polynomial of degree* D *that is not expressible in the form* $f(x) = g(x)^p - g(x) + \theta$ *for any* $g(x) \in \mathbb{F}_{q^m}[x]$, $\theta \in \mathbb{F}_{q^m}$. *Let*

$$N_\alpha(f) = |\{x \in \mathbb{F}_{q^m} : Tr_{q^m/q}(f(x)) = \alpha\}|.$$

Then

$$|N_\alpha(f) - q^{m-1}| \leq (D-1)\sqrt{q^m}.$$

Proof. We will calculate $N_\alpha(f)$ from an exponential sum. We have

$$qN_\alpha(f) = \sum_{x \in \mathbb{F}_{q^m}} \sum_{y \in \mathbb{F}_q} \omega^{Tr_{q/p}(y(Tr_{q^m/q}(f(x)) - \alpha))}.$$

This follows since the inner sum is q when $Tr_{q^m/q}(f(x)) = \alpha$ and 0 otherwise. Changing the order of summation and observing that in the case $y = 0$ the right hand side contributes q^m, leads to

$$qN_\alpha(f) - q^m = \sum_{y \in \mathbb{F}_q \setminus \{0\}} \omega^{-Tr_{q/p}(y\alpha)} \sum_{x \in \mathbb{F}_{q^m}} \omega^{Tr_{q^m/p}(yf(x))}.$$

From the Weil-Carlitz-Uchiyama bound, Theorem 5, it follows that

$$\begin{aligned} |N_\alpha(f) - q^{m-1}| &\leq \frac{1}{q} \sum_{y \in \mathbb{F}_q \setminus \{0\}} \left| \sum_{x \in \mathbb{F}_{q^m}} \omega^{Tr_{q^m/p}(yf(x))} \right| \\ &\leq (D-1)\sqrt{q^m}. \end{aligned}$$

Consider the set $\mathcal{F}_D$ of polynomials of degree $D \leq \sqrt{q^m}$, defined by

$$\mathcal{F}_D = \{f(x) : f(x) = f_1 x + f_2 x^2 + \cdots + f_D x^D \in \mathbb{F}_{q^m}[x], f_i = 0 \text{ whenever } p|i\}.$$

The condition $f_i = 0$ if $p|i$ for all $f \in \mathcal{F}_D$ guarantees that f is not expressible in the form $f(x) = g(x)^p - g(x) + \theta$ for any $g(x) \in \mathbb{F}_{q^m}[x]$, $\theta \in \mathbb{F}_{q^m}$, and hence Lemma 6 can be applied. Since $f(x)$ can contain all terms $f_i x^i$ where $f_i \in \mathbb{F}_{q^m}$, $1 \leq i \leq D$ and $p \not| i$, it follows that $|\mathcal{F}_D| = q^{m(D - \lfloor D/p \rfloor)}$. The construction of ϵ-ASU$_2$ hash families is described by the following theorem.

Theorem 7. *Let the functions in* $\mathcal{G}$ *map from* $X = \mathcal{F}_D$ *to* $Y = \mathbb{F}_q$, *let* $f \in \mathcal{F}_D = X$ *and define*

$$g_{\alpha,\beta}(f) = \beta + Tr_{q^m/q}(f(\alpha)).$$

Then the family

$$\mathcal{G} = \{g_{\alpha,\beta}(f) : \alpha \in \mathbb{F}_{q^m}, \beta \in \mathbb{F}_q\}$$

is an ϵ*-ASU*$_2$ *family of hash functions where*

$$|\mathcal{G}| = q^{m+1}, \ |X| = q^{m(D - \lfloor D/p \rfloor)}, \ |Y| = q, \ \epsilon = \frac{1}{q} + \frac{D-1}{\sqrt{q^m}}.$$

Proof. We verify property i) of Definition 2. For any $x \in X$ and $y \in Y$ we have

$$|\{g \in \mathcal{G} : y = g(x)\}| = |\{(\alpha, \beta) : y = \beta + Tr_{q^m/q}(f(\alpha)), \alpha \in \mathbb{F}_{q^m}, \beta \in \mathbb{F}_q\}| = q^m,$$

since for each $\alpha \in \mathbb{F}_{q^m}$ there is exactly one $\beta \in \mathbb{F}_q$ such that $y = \beta + Tr_{q^m/q}(f(\alpha))$. Secondly, we calculate ϵ as

$$\begin{aligned}
\epsilon &= \max_{x \neq x', y, y'} \frac{|\{g \in \mathcal{G} : y = g(x), y' = g(x')\}|}{|\mathcal{G}|/|Y|} && (3)\\
&= \max_{f \neq f', y, y'} \frac{|\{(\alpha, \beta) : y = \beta + Tr_{q^m/q}(f(\alpha)), y' = \beta + Tr_{q^m/q}(f'(\alpha))\}|}{q^m} && (4)\\
&= \max_{f \neq 0, y} \frac{|\{\alpha : y = Tr_{q^m/q}(f(\alpha))\}|}{q^m} && (5)\\
&= \max_{f \neq 0, y} \frac{N_y(f)}{q^m}\\
&\leq \frac{q^{m-1} + (D-1)\sqrt{q^m}}{q^m} = \frac{1}{q} + \frac{D-1}{\sqrt{q^m}},
\end{aligned}$$

where the inequality follows from Lemma 6, since it is valid for any nonzero $f \in \mathcal{F}_D$ and $y \in \mathbb{F}_q$. Furthermore $|X| = |\mathcal{F}_D| = q^{m(D - \lfloor D/p \rfloor)}$ and $|Y| = q$.

This family of strongly universal hash functions results in the following A-codes.

Corollary 8. *Let $\mathcal{S} = \mathcal{F}_D$, $\mathcal{E} = \{(\alpha, \beta) : \alpha \in \mathbb{F}_{q^m}, \beta \in \mathbb{F}_q\}$, and let the tag z be generated as*

$$z = \beta + Tr_{q^m/q}(f(\alpha)).$$

Then the parameters for the A-code are

$$|\mathcal{S}| = q^{m(D - \lfloor D/p \rfloor)}, \ |\mathcal{E}| = q^{m+1}, \ |\mathcal{Z}| = q,$$

and

$$P_I = \frac{1}{q}, \ P_S = \frac{1}{q} + \frac{D-1}{\sqrt{q^m}}.$$

We can verify the good performance of this construction by comparing with the parameters of the previously best known constructions [1]. For example, consider the parameters $q = 2^{20}$, $P_S = 2^{-19}$, and $m = 3$. By choosing q to be a large prime around 2^{20} we get $\log|\mathcal{S}| = 20 \cdot 3 \cdot \sqrt{2^{20}} = 60 \cdot 2^{10}$. In the construction in [1], we would for the same parameters get a number of source state bits which is $30 \cdot 2^{10}$. Or the other way around, that construction requires 82 key bits to authenticate $\log|\mathcal{S}| = 60 \cdot 2^{10}$ source bits, while our new construction only requires 80 key bits.

One might argue that the above construction gives only A-codes for a few values of the key size, namely $q^4, q^5, q^6, \ldots$, etc.. We will now show that by combining several strongly universal hash families for a smaller alphabet (q), we can consider any value of the number of key bits. This generalizes the previous considerations. First we prove a result analogue to Lemma 6.

Lemma 9. *Let $f_1, f_2 \in \mathcal{F}_D$ such that $f_1(x) \neq \alpha f_2(x)$, for all $\alpha \in \mathbb{F}_q$. Let*

$$N_{\alpha_1,\alpha_2}(f_1, f_2) = |\{x \in \mathbb{F}_{q^m} : Tr_{q^m/q}(f_1(x)) = \alpha_1, Tr_{q^m/q}(f_2(x)) = \alpha_2\}|.$$

Then

$$|N_{\alpha_1,\alpha_2}(f_1, f_2) - q^{m-2}| \leq (D-1)\sqrt{q^m}.$$

Proof. We calculate $N_{\alpha_1,\alpha_2}(f_1, f_2)$ from an exponential sum. Similar to the proof of Lemma 6, we obtain

$$\begin{aligned} q^2 N_{\alpha_1,\alpha_2}(f_1, f_2) &= \sum_{x \in \mathbb{F}_{q^m}} \sum_{y_1,y_2 \in \mathbb{F}_q} \omega^{Tr_{q/p}(y_1(Tr_{q^m/q}(f_1(x)) - \alpha_1))} \omega^{Tr_{q/p}(y_2(Tr_{q^m/q}(f_2(x)) - \alpha_2))} \\ &= \sum_{y_1,y_2 \in \mathbb{F}_q} \omega^{-Tr_{q/p}(y_1\alpha_1 + y_2\alpha_2)} \sum_{x \in \mathbb{F}_{q^m}} \omega^{Tr_{q^m/p}(y_1 f_1(x) + y_2 f_2(x))}. \end{aligned}$$

Isolating the case $y_1 = y_2 = 0$, which contributes q^m to the sum in the right hand side, and using the Weil-Carlitz-Uchiyama bound, Theorem 5, it then follows that

$$|N_{\alpha_1,\alpha_2}(f_1, f_2) - q^{m-2}| \leq (D-1)\sqrt{q^m}.$$

Theorem 10. *Let $\gamma \in \mathbb{F}_{q^m}$ be a primitive element, and let the functions in $\mathcal{G}$ map from $X = \mathcal{F}_D$ to $Y = \mathbb{F}_q^2$. Let $f \in \mathcal{F}_D = X$ and define*

$$g_{\alpha,\beta_1,\beta_2}(f) = (\beta_1 + Tr_{q^m/q}(f(\alpha)), \beta_2 + Tr_{q^m/q}(f(\gamma\alpha))).$$

Then the family

$$\mathcal{G} = \{g_{\alpha,\beta_1,\beta_2}(f) : \alpha \in \mathbb{F}_{q^m}, \beta_1, \beta_2 \in \mathbb{F}_q\}$$

is an ϵ-ASU_2 family of hash functions where

$$|\mathcal{G}| = q^{m+2}, \ |X| = q^{m(D-\lfloor D/p \rfloor)}, \ |Y| = q^2, \ \epsilon = \frac{1}{q^2} + \frac{D-1}{\sqrt{q^m}}.$$

Proof. Property i) in Definition 2 is easily verified. With the same steps as in (3)-(5) we get

$$\epsilon = \max_{f \neq 0, y} \frac{N_{y_1,y_2}(f(x), f(\gamma x))}{q^m} \leq \frac{q^{m-2} + (D-1)\sqrt{q^m}}{q^m} = \frac{1}{q^2} + \frac{D-1}{\sqrt{q^m}},$$

where the inequality follows from Lemma 9, since $f(x)$ and $f(\gamma x)$ both belong to $\mathcal{F}_D$ and $f(x) \neq \alpha f(\gamma x)$, for all $\alpha \in \mathbb{F}_q$, for any choice of $f(x) \neq 0$. Also, $|X| = q^{m(D-\lfloor D/p \rfloor)}$ and $|Y| = q^2$.

It is clear that the above results can be further generalized. We give the results and omit the proofs.

Lemma 11. *Let $f_1, f_2, \ldots, f_n \in \mathcal{F}_D$ be n $(\leq m)$ linearly independent polynomials over $\mathbb{F}_q$. Let*

$$N_{\alpha_1,\ldots,\alpha_n}(f_1,\ldots,f_n) = |\{x \in \mathbb{F}_{q^m} : Tr_{q^m/q}(f_1(x)) = \alpha_1, \ldots, Tr_{q^m/q}(f_n(x)) = \alpha_n\}|.$$

Then

$$|N_{\alpha_1,\ldots,\alpha_n}(f_1,\ldots,f_n) - q^{m-n}| \leq (D-1)\sqrt{q^m}.$$

Theorem 12. *Let the functions in $\mathcal{G}$ map from $X = \mathcal{F}_D$ to $Y = \mathbb{F}_q^n$. Let $\gamma \in \mathbb{F}_{q^m}$ be a primitive element, and let $f \in \mathcal{F}_D = X$. Define*

$$g_{\alpha,\underline{\beta}}(f) = (\beta_1 + Tr_{q^m/q}(f(\alpha)), \beta_2 + Tr_{q^m/q}(f(\gamma\alpha)), \ldots, \beta_n + Tr_{q^m/q}(f(\gamma^{n-1}\alpha))).$$

Then the family

$$\mathcal{G} = \{g_{\alpha,\underline{\beta}}(f) : \alpha \in \mathbb{F}_{q^m}, \beta_1, \ldots, \beta_n \in \mathbb{F}_q\}$$

is an ϵ-ASU_2 family of hash functions where

$$|\mathcal{G}| = q^{m+n},\ |X| = q^{m(D-\lfloor D/p \rfloor)},\ |Y| = q^n,\ \epsilon = \frac{1}{q^n} + \frac{D-1}{\sqrt{q^m}}.$$

Corollary 13. *Let $\mathcal{S} = \mathcal{F}_D$, $\mathcal{E} = \{(\alpha, \beta_1, \ldots, \beta_n) : \alpha \in \mathbb{F}_{q^m}, \beta_1, \ldots, \beta_n \in \mathbb{F}_q\}$, and let the tag z be generated as*

$$z = (\beta_1 + Tr_{q^m/q}(f(\alpha)), \beta_2 + Tr_{q^m/q}(f(\gamma\alpha)), \ldots, \beta_n + Tr_{q^m/q}(f(\gamma^{n-1}\alpha))).$$

Then the parameters for the A-code are

$$|\mathcal{S}| = q^{m(D-\lfloor D/p \rfloor)},\ |\mathcal{E}| = q^{m+n},\ |\mathcal{Z}| = q^n,$$

and

$$P_I = \frac{1}{q^n},\ P_S = \frac{1}{q^n} + \frac{D-1}{\sqrt{q^m}}.$$

If the last construction is used with $q = 2$, we can compare the parameters with [1] and find that they will be *exactly* the same. However, the proposed constructions have two advantages. Firstly, one may get a construction for any number of key bits in the authentication case, which is not possible in [1]. Secondly, by using Corollary 8 with $q = p$, where p is a large prime, one gets improvements compared with [1] which roughly is a doubling of the number of source bits that can be authenticated, or for a fixed number of source bits, a reduction of the number of key bits by roughly 2.

We will next show that using the theory of exponential sums over Galois rings we can also construct universal hash families.

5 Exponential sums over Galois rings

Some preliminaries on Galois rings are given below. For more details on Galois rings, the reader is referred to [12, 13, 2] and [5]. Let p be a fixed prime. For applications, the case $p = 2$ is most relevant. Let $e \geq 1$ be an integer and set $q = p^e$. Let $\mathbb{Z}_q$ denote the integers mod-q, and $\mathbb{F}_q$ the finite field with q elements.

Let $\mu : \mathbb{Z}_q \to \mathbb{Z}_p = \mathbb{F}_p$ be the mod-p reduction map. We extend μ to a map $\mathbb{Z}_q[x] \to \mathbb{Z}_p[x]$ in the natural way. A monic polynomial $g(x) \in \mathbb{Z}_q[x]$ is said to be a *monic basic irreducible* if $\mu(g(x)) = \overline{g}(x)$ is a monic irreducible polynomial in $\mathbb{Z}_p[x]$. A Galois ring $GR(q, m), m \geq 1$ of q^m elements is simply a Galois extension of $\mathbb{Z}_q$. We will write $R_{q^m} = GR(q, m)$ for short. Every such ring is isomorphic to the ring $\mathbb{Z}_q[x]/(g(x))$, where $g(x)$ is monic basic irreducible of degree m. R_{q^m} is a local ring having a unique maximal ideal $M_{q^m} = pR_{q^m}$. Clearly μ has a natural extension to R_{q^m} and therefore to $R_{q^m}[x]$, and $\mu(R_{q^m}) = R_{q^m}/M_{q^m} \cong F_{p^m}$.

As a multiplicative group, the units $R^*_{q^m}$ in R_{q^m} contain a cyclic group of order $p^m - 1$. Let $\beta \in R^*_{q^m}$ be a generator of this cyclic group. Let $\mathcal{T}_m = \{0, 1, \beta, \cdots, \beta^{p^m-2}\}$. It can be shown that every element $z \in R_{q^m}$ has the p-adic expansion

$$z = z_0 + pz_1 + p^2 z_2 + \cdots + p^{e-1} z_{e-1},\ z_i \in \mathcal{T}_m.$$

The ring R_{q^m} is an extension ring of $\mathbb{Z}_q = R_q$ having a cyclic Galois group of order m generated by the Frobenius automorphism σ given by

$$\sigma(z) = z_0^p + pz_1^p + p^2 z_2^p + \cdots + p^{e-1} z_{e-1}^p,$$

where $z = z_0 + pz_1 + p^2 z_2 + \cdots + p^{e-1} z_{e-1},\ z_i \in \mathcal{T}_m$. Given $x \in R_{q^m}$, we define the *trace* $T_{q^m/q} : R_{q^m} \to R_q$ via

$$T_{q^m/q}(x) = \sum_{i=0}^{m-1} \sigma^i(x).$$

We next present some recent results on exponential sums over Galois rings. Let f be a polynomial $f(x) = \sum_{i=0}^{d} f_i x^i$ in $R_{q^m}[x]$ of degree d. Let

$$f(x) = F_0(x) + pF_1(x) + \cdots + p^{e-1} F_{e-1}(x),\ F_j(x) \in \mathcal{T}_m[x],\ 0 \leq j \leq e-1$$

be the p-adic expansion of $f(x)$. Such an expansion can be derived from a p-adic expansion of the coefficients of $f(x)$. Let d_i be the degree of $F_i(x)$. We will assume that it is not possible to express $f(x)$ in the form

$$f(x) = \sigma(g(x)) - g(x) + \theta \pmod{q}$$

for any $g(x) \in R_{q^m}[x]$, $\theta \in R_{q^m}$. Here σ is the Frobenius automorphism and $\sigma(\sum_i g_i x^i) = \sum_i \sigma(g_i) x^{pi}$. We will say that f is *non-degenerate* when f satisfies this condition. We define

$$D_f = \max\{d_o p^{e-1}, d_1 p^{e-2}, \ldots, d_{e-1}\},$$

and will refer to D_f as the *weighted degree* of $f(x)$.

The following generalization of the Weil-Carlitz-Uchiyama bound was proved by Kumar, Helleseth and Calderbank.

Theorem 14 [10]. *Let $f(x) \in R_{q^m}[x]$, $q = p^e$, be non-degenerate and let D_f be the weighted degree of $f(x)$. Then*

$$\left| \sum_{x \in \mathcal{T}_m} \omega^{T_{q^m/q}(f(x))} \right| \leq (D_f - 1)\sqrt{p^m}$$

where ω is a complex primitive q-th root of unity.

6 Universal hash functions from exponential sums over Galois rings

We can construct strongly universal hash functions from exponential sums over Galois rings in a manner similar to what we did in Section 4.

Lemma 15. *Let $f(x) = \sum_{i=1}^{d} f_i x^i \in R_{q^m}[x]$, $q = p^e$, be a non-degenerate polynomial of weighted degree D_f such that $f(x) \neq 0 \pmod{p}$. Let*

$$N_\alpha(f) = |\{x \in \mathcal{T}_m : T_{q^m/q}(f(x)) = \alpha\}|.$$

Then

$$|N_\alpha(f) - p^{m-e}| \leq (D_f - 1)\sqrt{p^m}.$$

Proof. We calculate $N_\alpha(f)$ from an exponential sum over a Galois ring. We have

$$qN_\alpha(f) = \sum_{x \in \mathcal{T}_m} \sum_{y \in \mathbb{Z}_q} \omega^{y(T_{q^m/q}(f(x)) - \alpha)},$$

since the inner sum is q when $T_{q^m/q}(f(x)) = \alpha$ and 0 otherwise. Observing that in the case $y = 0$ the right hand side contributes p^m, leads to

$$qN_\alpha(f) - p^m = \sum_{y \in \mathbb{Z}_q \setminus \{0\}} \omega^{-y\alpha} \sum_{x \in \mathcal{T}_m} \omega^{T_{q^m/q}(yf(x))}.$$

From Theorem 14, it follows that

$$|N_\alpha(f) - p^{m-e}| \leq (D_f - 1)\sqrt{p^m}.$$

We next introduce a generalization of the Gray-map. The Gray-map is used in coding theoretic applications for constructing binary codes and sequences. The main property for these maps is explained in the following lemma.

Lemma 16. *For $i \in \mathbb{Z}_p$, $x \in \mathbb{Z}_{p^2}$, let $x = x_1 p + x_0, x_0, x_1 \in \mathbb{Z}_p$ and define $\phi_i : \mathbb{Z}_{p^2} \to \mathbb{Z}_p$ by*

$$\phi_i(x) = x_1 + ix_0.$$

Then for any $x, y \in \mathbb{Z}_{p^2}$ and $z \in \mathbb{Z}_p$

$$|\{i \in \mathbb{Z}_p : \phi_i(x) - \phi_i(y) = z\}| = |\{i \in \mathbb{Z}_p : \phi_i(x - y) = z\}|.$$

Proof. Let $x = x_1 p + x_0$ and $y = y_1 p + y_0$, where $x_1, x_0, y_1, y_0 \in \mathbb{Z}_p$. Now $\phi_i(x)$ is defined by $\phi_i(x) = x_1 + ix_0$. Then $\phi_i(x) - \phi_i(y) = (x_1 - y_1) + i(x_0 - y_0)$. Now it is easy to see that if $x_0 \neq y_0$ each value of $\mathbb{Z}_p$ will appear once both for $\phi_i(x) - \phi_i(y)$ and for $\phi_i(x - y)$, when i runs through $\mathbb{Z}_p$. Furthermore, if $x_0 = y_0$ then $\phi_i(x) - \phi_i(y) = \phi_i(x - y) = x_1 - y_1$.

Consider the set $\mathcal{R}_D$ of polynomials of weighted degree at most $D \leq \sqrt{p^m}$, defined by

$$\mathcal{R}_D = \{f(x) : f(x) = f_1 x + f_2 x^2 + \cdots + f_d x^d \in R_{q^m}[x], D_f \leq D, f_i = 0 \text{ whenever } p|i\}.$$

The condition $f_i = 0$ if $p|i$ for all $f \in \mathcal{R}_D$ guarantees that f is non-degenerate. From the definitions of $\mathcal{R}_D$ and D_f it follows that $|\mathcal{R}_D| = p^{m(D - \lfloor D/p^e \rfloor)}$. The corresponding construction is the following.

Theorem 17. *Let the functions in $\mathcal{G}$ map from $X = \mathcal{R}_D$ to $Y = \mathbb{Z}_p$. Let $f \in \mathcal{R}_D = X$, $q = p^2$, and define*

$$g_{\alpha,\beta,i}(f) = \beta + \phi_i(T_{q^m/q}(f(\alpha))).$$

Then the family

$$\mathcal{G} = \{g_{\alpha,\beta,i}(f) : \alpha \in \mathcal{T}_m, \beta, i \in \mathbb{Z}_p\}$$

is an ϵ-ASU_2 family of hash functions where

$$|\mathcal{G}| = p^{m+2},\ |X| = p^{m(D - \lfloor D/p^2 \rfloor)},\ |Y| = p,\ \epsilon = \frac{1}{p} + \frac{D-1}{\sqrt{p^m}} + \min(\frac{1}{p^2}, \frac{D-1}{\sqrt{p^m}}).$$

Proof. Property i) in Definition 2 is easily verified. We calculate ϵ by

$$\begin{aligned} \epsilon &= \max_{f \neq f', y, y'} \frac{|\{(\alpha, \beta, i) : \beta + \phi_i(T(f(\alpha))) = y, \beta + \phi_i(T(f'(\alpha))) = y'\}|}{p^{m+1}} \\ &= \max_{f \neq 0, y} \frac{|\{(\alpha, i) : \phi_i(T(f(\alpha))) = y\}|}{p^{m+1}}. \end{aligned} \tag{6}$$

where (6) follows from Lemma 16. We consider two cases.

Case I. $f = 0 \pmod p$. Then $f = pf_1$ for some polynomial $f_1 \neq 0 \pmod p$ and $\phi_i(T(f(\alpha)))$ will take the same value for all i, so

$$\begin{aligned} \epsilon &= \max_{f \neq 0, y} \frac{|\{(\alpha, i) : \phi_i(T(f(\alpha))) = y\}|}{p^{m+1}} \\ &= \max_{f \neq 0, y} \frac{|\{\alpha : T(pf_1(\alpha)) = y\}|}{p^m} \\ &\leq \frac{p^{m-1} + (D-1)\sqrt{p^m}}{p^m} \\ &= \frac{1}{p} + \frac{D-1}{\sqrt{p^m}}, \end{aligned}$$

i.e., as in the finite field case.

Case II. $f \neq 0 \pmod p$. Now $\phi_i(T(f(\alpha)))$ will be uniformly distributed with i when $T(f(\alpha)) \neq 0 \pmod p$, and take the same value for all i when $T(f(\alpha)) = 0 \pmod p$.

Hence

$$\begin{aligned}\epsilon &= \max_{f\neq 0,y} \frac{|\{(\alpha,i) : \phi_i(T(f(\alpha))) = y\}|}{p^{m+1}} \\ &= \max_{f\neq 0,y} \frac{|\{\alpha : T(f(\alpha)) \neq 0 \pmod p\}| + p|\{\alpha : T(f(\alpha)) = py\}|}{p^{m+1}} \\ &= \frac{p^m - |\{\alpha : T(f(\alpha)) = 0 \pmod p\}| + p|\{\alpha : T(f(\alpha)) = py\}|}{p^{m+1}}.\end{aligned}$$

We use Lemma 15 and get

$$\begin{aligned}\epsilon &\leq \frac{p^m + p(p^{m-2} + (D-1)\sqrt{p^m})}{p^{m+1}} \\ &\leq \frac{1}{p} + \frac{1}{p^2} + \frac{D-1}{\sqrt{p^m}}.\end{aligned}$$

We note that when $p^{m-2} > (D-1)\sqrt{p^m}$, we get a better estimate of ϵ by

$$\epsilon \leq \frac{1}{p} + 2\frac{D-1}{\sqrt{p^m}},$$

since $|\{\alpha : T(f(\alpha)) = 0 \pmod p\}| \geq p(p^{m-2} - (D-1)\sqrt{p^m})$.

A closer study reveals that this ϵ-ASU$_2$ hash family has weaker performance than the hash families constructed in Section 4. However, for $p = 2$ it is possible to use sharper bounds than described here to get improvements on ϵ in Theorem 17. For authentication codes, p should typically be large and hence such improvements are of interest only for other applications.

7 Conclusions

Construction	Key size (bits)							
	70	72	74	76	78	80	90	100
[1]	$25 \cdot 2^5$	$26 \cdot 2^6$	$27 \cdot 2^7$	$28 \cdot 2^8$	$29 \cdot 2^9$	$30 \cdot 2^{10}$	$35 \cdot 2^{15}$	$40 \cdot 2^{20}$
Cor. 8, q prime	-	-	-	-	-	$60 \cdot 2^{10}$	-	$80 \cdot 2^{20}$
Cor. 13, $q = 2$	$25 \cdot 2^5$	$26 \cdot 2^6$	$27 \cdot 2^7$	$28 \cdot 2^8$	$29 \cdot 2^9$	$30 \cdot 2^{10}$	$35 \cdot 2^{15}$	$40 \cdot 2^{20}$

Table 1. Table of the number of source bits in A-codes for different key sizes in some different constructions, for $P_I = 2^{-20}$, and $P_S \leq 2^{-19}$.

In Table 1 we present parameters of the derived constructions when used as authentication codes. We conclude by a short discussion around the implementation. Implementation aspects have recently gained attention, and several papers

focus on this topic [8], [9], and [14]. The time consuming part in the proposed constructions is the evaluation of a polynomial of large degree over a finite field (or Galois ring). We need to do additions and multiplications. In hardware, multiplications in a finite field of characteristic 2 has a simple implementation. In software, we can choose a large prime p close to 2^w, e.g., $p = 2^w - 1$, and a multiplication mod-p can be done by one multiplication mod-2^w together with one addition. Using more sophisticated methods for the evaluation of a high degree polynomial will improve the performance.

Finally, we remark that the trace codes obtained in a natural way from the family $\mathcal{F}_D$ correspond to the dual of the extended BCH-codes. The families of polynomials $\mathcal{F}_D$ and more recently $\mathcal{R}_D$ [10] have been used as a basis for constructing families of sequences with very good correlation properties for CDMA applications. An interesting problem for future research would be to investigate whether other families of sequences or codes combined with the methods in this paper will lead to further improvements in the design of universal hash functions and authentication codes.

References

1. J. Bierbrauer, T. Johansson, G. Kabatianskii, and B. Smeets, On families of hash functions via geometric codes and concatenation, *Advances in Cryptology, Proceedings of CRYPTO 93*, D.R. Stinson, ed., *Lecture Notes in Computer Science*, 773 (1994), 331-342.
2. S. Boztas, R. Hammons, Jr., and P.V. Kumar, 4-phase sequences with near-optimum correlation properties, *IEEE Trans. Inform. Theory*, 38 (1992), 1101-1113.
3. L. Carlitz and S. Uchiyama, Bounds on exponential sums, *Duke Math. J.*, (1957), 37-41.
4. E.N. Gilbert, F.J. MacWilliams, and N.J.A. Sloane, Codes which detect deception, *Bell Syst. Tech. J.*, 53 (1974), 405-424.
5. A.R. Hammons Jr., P.V. Kumar, A.R. Calderbank, N.J.A. Sloane, and P. Solé, The $\mathbb{Z}_4$-linearity of Kerdock, Preparata, Goethals, and related codes, *IEEE Trans. Inform. Theory*, 40 (1994), 301-319.
6. T. Helleseth, On the covering radius of cyclic linear codes and arithmetic codes, *Discrete Appl. Math.*, 11 (1985), 157-173.
7. G. Kabatianskii, B. Smeets, and T. Johansson, On the cardinality of systematic authentication codes via error correcting codes, *IEEE Trans. Inform. Theory*, 42 (1996), 566-578.
8. H. Krawczyk, LFSR-based hashing and authentication, *Advances in Cryptology, Proceedings of CRYPTO 94*, Y. Desmedt, ed., *Lecture Notes in Computer Science*, 839 (1994), 129-139.
9. H. Krawczyk, New hash functions for message authentication, *Advances in Cryptology, Proceedings of EUROCRYPT 95*, L.C. Guillou and J.-J. Quisquater, eds., *Lecture Notes in Computer Science*, 921 (1995), 140-149.
10. P.V. Kumar, T. Helleseth, and A.R. Calderbank, An upper bound for Weil exponential sums over Galois rings and applications, *IEEE Trans. Inform. Theory*, 41 (1995), 456-468.

11. R. Lidl and H. Niederreiter, Finite-fields, volume 20 of Encyclopedia of mathematics and its applications, Addison-Wesley, Reading, MA, 1983.
12. B.R. MacDonald, Finite rings with identity, Marcel Dekker, New York, 1974.
13. A. Nechaev, The Kerdock code in a cyclic form, *Discrete Appl. Math.*, 1 (1991), 365-384.
14. P. Rogaway, Bucket hashing and its application to fast message authentication, *Advances in Cryptology, Proceedings of CRYPTO 95*, D. Coppersmith, ed., *Lecture Notes in Computer Science*, 963 (1995), 29-42.
15. G.J. Simmons, A game theory model of digital message authentication, *Congr. Numer.*, 34 (1992), 413-424.
16. G.J. Simmons, Authentication theory/coding theory, in *Advances in Cryptology, Proceedings of CRYPTO 84*, G.R. Blakley and D. Chaum, eds., *Lecture Notes in Computer Science*, 196 (1985), 411-431.
17. D.R. Stinson, Universal hashing and authentication codes, *Codes, Designs and Cryptography* 4 (1994), 337-346.
18. D.R. Stinson, On the connection between universal hashing, combinatorial designs and error-correcting codes, *http://bibd.unl.edu/~stinson/*.
19. M.N. Wegman and J.L. Carter, New hash functions and their use in authentication and set equality, *J. Computer and System Sciences*, 22 (1981), 265-279.