J. Cryptology (1990) 2: 23–49

Journal of Cryptology

The Combinatorics of Authentication and Secrecy Codes[1]

D. R. Stinson
Department of Computer Science, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Abstract. This paper is a study of the combinatorics of unconditionally secure secrecy and authentication codes, under the assumption that each encoding rule is to be used for the transmission of some number L of successive messages. We obtain bounds on the number of encoding rules required in order to obtain maximum levels of security. Some constructions are also given for codes which have the minimum number of encoding rules. These constructions use various types of combinatorial designs.

Key words. Authentication code, Secrecy code, Combinatorial design.

1. Authentication and Secrecy

This paper is a study of the combinatorics of secrecy and authentication codes. We are interested in the *unconditional*, or *theoretical*, security provided by such codes. That is, we assume that any opponents have unlimited computational resources. The theory of unconditional secrecy is due to Shannon [16]. More recently, Simmons has developed an analogous theory of unconditional authentication [17], [19], [20], [21].

By the *combinatorics* of codes, we are referring to two aspects. First, the bounds on the security of the codes and on the minimum sizes of codes attaining specified levels of security are combinatorial in nature and/or are proved by combinatorial (i.e., counting) arguments. Second, the constructions for "good" codes which meet the various bounds with equality make essential use of combinatorial designs. This will become evident in the rest of the paper. For a general refernce on design theory, we mention [1].

We use the model of a secrecy system developed by Shannon in [16], updated to include authentication, as described in [11]. In this model, there are three participants: a *transmitter*, a *receiver*, and an *opponent*. The transmitter wants to communicate some information to the receiver using a public communications channel. The *source state* (or *plaintext*) is encrypted to obtain the *message* (*ciphertext*), which

[1] Date received: December 5, 1988. Date revised: September 6, 1989.

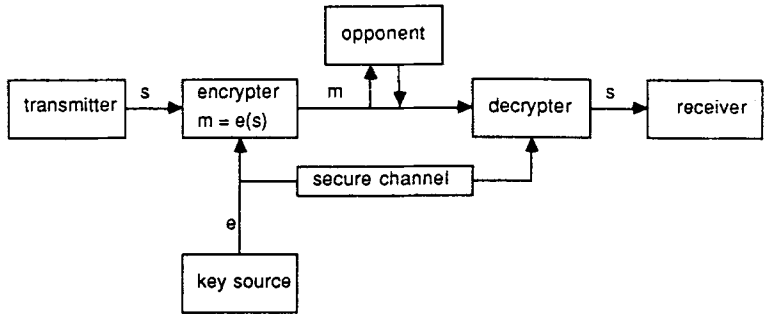

Fig. 1. Shannon's model of a general secrecy system.

is sent through the channel. An *encoding rule* (or *key*) e defines the message $e(s)$ to be sent to communicate any source state s. Each encoding rule will be a one-to-one function from the source space to the message space. We assume the transmitter has a key source from which he obtains a key. Prior to any messages being sent, this key is communicated to the receiver by means of a secure channel. Figure 1 shows our model, taken from [11].

We use the following notation. Let $\mathcal{S}$ be a set of k source states, let $\mathcal{M}$ be a set of v messages, and let $\mathcal{E}$ be a set of b encoding rules. As stated above, each encoding rule is a one-to-one function from $\mathcal{S}$ to $\mathcal{M}$. It is useful to think of a code as being represented by a $b \times k$ matrix, where the rows are indexed by encoding rules, the columns are indexed by source states, and the entry in row e and column s is $e(s)$. We call this matrix the *encoding matrix*. For any encoding rule $e \in \mathcal{E}$, define $M(e) = \{e(s): s \in \mathcal{S}\}$, i.e., the set of *valid* messages under encoding rule e. For an encoding rule e, and a set of messages $M \subseteq M(e)$, define $f_e(M) = S$ if $\{e(s): s \in S\} = M$, that is, $f_e(M)$ is set of source states which are encrypted to the set of messages M under encoding rule e.

Assume that the same key is used to encrypt up to L consecutive source states, where L is some fixed positive integer. We make the following simplifying assumptions, which are not strictly necessary, but which avoid some difficulties in the mathematical analysis. First, we assume that the L source states that occur are all distinct. Second, we ignore the order in which the messages are sent through the channel, and the order in which the corresponding source states occur. Hence, we refer only to *sets* of L messages or source states occurring (as opposed to sequences). Finally, for any $i \leq L$, we assume that there is some probability distribution on the set of i subsets of source states, so that any set of i source states has a nonzero probability of occurring. Given a set of i source states S, we denote by $p(S)$ the probability that the source states in S are the i source states that occur.

We should note that other researchers have considered the order in which messages and source states occur (this is the model used in [4], [11], [12], and [15]). In such a model, we would speak of sequences of messages observed in the channel corresponding to sequences of source states. Of course, for $L = 1$, the two models are equivalent, but for $L > 1$, there are important differences.

First, we consider the property of secrecy. Assume an opponent observes i ($\leq L$)

distinct messages being sent over the communication channel using the same encoding rule. Although he knows that the same encoding rule is being used to transmit the i messages, he does not know what that encoding rule is. Our goal is that the opponent be unable to determine any information regarding the i source states from the i messages he has observed. This concept is made precise as follows. We say that a code has *perfect L-fold secrecy* if, for every $L' \leq L$, for every set M' of L' messages observed in the channel, and for every set S_1 of L' source states, we have the $p(S_1|M_1) = p(S_1)$. That is, the conditional probability distribution on the L' source states after observing a set of L' messages in the channel is the same as the *a priori* probability distribution on the L' source states.

Example 1.1. A code having $k = 4$ source states, $v = 4$ messages, and $b = 4$ encoding rules, and which achieves perfect onefold secrecy. Use each encoding rule with probability 1/4.

	s_1	s_2	s_3	s_4
e_1	1	2	3	4
e_2	2	1	4	3
e_3	3	4	1	2
e_4	4	3	2	1

An important consideration in the construction of a code is the number of encoding rules. For, the encoding rule is information that must be communicated using a secure channel. If there are b encoding rules, then $\log_2 b$ bits of key must be communicated. Hence, it is clear that we want to minimize b. In general, b may depend on v, k, and the level of secrecy required. Having proved a lower bound on b as a function of these other parameters, we would want to find constructions for codes where the number of encoding rules meets, or is close to, the lower bounds. These are the main objectives in this paper.

The following is a lower bound required on the number of encoding rules required in a code having perfect L-fold secrecy. We prove this bound in Section 2.

Theorem 2.1. *If a code achieves perfect L-fold secrecy, then*

$$b \geq \binom{k}{L}.$$

The above theorem is a straightforward generalization of the well-known result of Shannon [16] that a code achieving perfect onefold secrecy must have at least as many keys as source states. Let us say that a code achieving L-fold secrecy is *optimal* if

$$b = \binom{k}{L}.$$

As an example, we mention the Vernam one-time pad [28], which is an optimal onefold secrecy code. In Section 2 we construct some optimal twofold and threefold

secrecy codes using a type of combinatorial design called a perpendicular array. Conversely, we show that optimal L-fold secrecy codes can be constructed only in this fashion.

Next, we extend the model of the secrecy system to include authentication in the same way as Massey did in [11]. As before, an opponent observes i distinct messages which are sent using the same encoding rule. However, the opponent now has the ability to introduce new messages into the channel and/or to modify existing messages. Assume the opponent places a message m' into the channel by either of these methods, where m' is distinct from the i messages already sent. His goal is to have m' accepted as authentic by the receiver. That is, if e is the encoding rule being used, then the opponent is hoping that $m' = e(s)$ for some source state s. In [11] Massey calls this a *spoofing attack* of order i. This problem was first studied in [6]. The special case $i = 0$ and $i = 1$ were analyzed by Simmons in [17], [19], and [20]. The case $i = 0$ is called the *impersonation* game, and the case $i = 1$ is called the *substitution* game. More recently, several other researchers have studied these cases; see, for example, [2], [11], [5], and [25]. Less is known about the cases $i \geq 2$; some results can be found in [5], [15], [26], and [27].

Given the probability distributions on the source states described above, the receiver and transmitter will choose a probability distribution for $\mathscr{E}$, called an *encoding strategy*. It is assumed that the opponent knows the encoding strategy being used. Once the transmitter/receiver have chosen encoding strategies, we can calculate, for each $i \geq 0$, a probability denoted Pd_i, which is the probability that the opponent can deceive the transmitter/receiver with a spoofing attack of order i. The following lower bound on Pd_i can be proved.

Theorem [11, p. 12]. *In an authentication code with k source states and v messages, $Pd_i \geq (k - i)/(v - i)$.*

Following Massey [11], we say that the authentication code is *L-fold secure against spoofing if* $Pd_i = (k - i)/(v - i)$ for $0 \leq i \leq L$. We refer to this bound as the *combinatorial* bound, since it does not take into account the probability distributions on the source states and encoding rules. Other lower bounds on Pd_i can be proved, which depend on the entropies of these probability distributions. Given a particular code, these entropy bounds measure how "efficiently" information is being sent through the channel. We do not discuss these bounds in this paper, instead we refer to [2], [20], and [25].

Define an (L_S, L_A)-*code* to be a code which achieves perfect L_S-fold secrecy and is L_A-fold secure against spoofing. If we want a code to have secrecy and authentication, then we would most likely require that L_S be close to L_A. The cases $L_S = L_A$ and $L_S = L_A + 1$ are both very natural special cases to study. These $(L_S, L_S - 1)$-codes are probably most natural to consider when the opponent can introduce new messages into the channel but cannot modify existing messages. For, under these assumptions, the receiver would ignore any messages received after he has received L_S messages in the channel. If the opponent can also modify existing messages, then it is of interest to study (L_S, L_S)-codes. In these two cases, we have lower bounds on the number of encoding rules required in any such code, as follows.

Theorem [26, Theorem 2]. *If a code achieves perfect L_S-fold secrecy and is $(L_S - 1)$-fold secure against spoofing, then*

$$b \geq \binom{v}{L_S}.$$

Theorem 4.1. *If a code achieves perfect L_S-fold secrecy and is L_S-fold secure against spoofing for any source probability distribution, then*

$$b \geq \binom{v}{L_S} \cdot \frac{v - L_S}{k - L_S}.$$

An (L_S, L_A)-code, where $L_S = L_A$ or $L_S = L_A + 1$, is *optimal* if the number of encoding rules meets the appropriate lower bound with equality. Constructions have been given for infinite classes of optimal (1, 0)-codes in [5], for optimal (2, 1)-codes in [26], and for nearly optimal (3, 2)-codes in [27]. These constructions are reviewed in Section 3, and some constructions are also given for optimal and near-optimal (L_S, L_S)-codes in Section 4.

Example 1.2. An optimal (2, 1)-code having $k = 5$ source states, $v = 5$ messages, and $b = 10$ encoding rules. Use each encoding rule with probability 1/10.

	s_1	s_2	s_3	s_4	s_5
e_1	1	2	3	4	5
e_2	2	3	4	5	1
e_3	3	4	5	1	2
e_4	4	5	1	2	3
e_5	5	1	2	3	4
e_6	1	3	5	2	4
e_7	2	4	1	3	5
e_8	3	5	2	4	1
e_9	4	1	3	5	2
e_{10}	5	2	4	1	3

The next topic we address in this paper is authentication *without* secrecy. We note that there are applications where we require a code that provides authentication, but secrecy cannot be tolerated. For example, this situation arose in the authentication of data to verify compliance with a nuclear weapons test ban treaty [18].

Hence, we define a code to have *perfect disclosure* if any message m observed in the channel determines a unique source state s. (Sometimes the term Cartesian is used to describe this situation.) In terms of probability distributions, we require that $p(s|m) = 1$ and $p(s|m') = 0$ if $m' \neq m$.

It is easy to see that a code with perfect disclosure cannot be even onefold secure against spoofing. We therefore prove in Section 5 the following lower bound on Pd_i for codes with perfect disclosure.

Theorem 5.1. *If a code has perfect disclosure, then* $Pd_0 \geq k/v$. *Moreover, if* $Pd_0 = k/v$, *then* $Pd_L \geq k/v$ *for any* $L \geq 0$.

We then prove the following bound on the number of encoding rules.

Theorem 5.2. *If a code has perfect disclosure, and* $Pd_i = k/v$ *for* $0 \leq i \leq L-1$, *then* $b \geq (v/k)^{L+1}$.

Then we give constructions for codes that meet these bounds and do so with the minimum number of encoding rules.

Example 1.3. A perfect disclosure code having $k = 4$ source states, $v = 12$ messages, and $b = 9$ encoding rules, for which $Pd_0 = Pd_1 = 1/3$. Use each encoding rule with probability 1/9.

	s_1	s_2	s_3	s_4
e_1	1	4	7	10
e_2	1	5	8	11
e_3	1	6	9	12
e_4	2	4	8	12
e_5	2	5	9	10
e_6	2	6	7	11
e_7	3	4	9	11
e_8	3	5	7	12
e_9	3	6	8	10

The bounds we give in this paper are all combinatorial, in the sense that they are independent of the various probability distributions involved. Moreover, most of the codes we construct in this paper attain the desired level of secrecy and/or security against spoofing for an arbitrary source probability distribution. (Unless otherwise stated, any code in this paper will have this property). This is clearly a very desirable property, since we might not even know the source probability distribution, for example.

With regard to secrecy codes, we prove in Section 2 that a code having perfect L-fold secrecy for some *fixed* source probability distribution will also achieve perfect L-fold secrecy for *any* source probability distribution. For codes providing secrecy and authentication, the situation is more complex. In fact, it is easier to design codes if the source states are known to be equiprobable. For example, suppose we consider a code that achieves perfect onefold secrecy and is onefold secure against spoofing. If this is to be true for an arbitrary source probability distribution, then at least $v(v-1)/(k-1)$ encoding rules are required, by Theorem 4.1. However, if the source states are known to be equiprobable, then we can achieve the same security from a code having only $v(v-1)/(k(k-1))$ encoding rules, provided a suitable design exists (Theorem 6.4). Constructions for authentication/secrecy codes for equiprobable source distributions are presented in Section 6.

We give an example to illustrate the effect of source probability distribution on the deception probabilities of an authentication code.

Example 1.4. A code having $k = 2$ source states, $v = 3$ messages, and $b = 3$ encoding rules, each used with probability 1/3. Assume the source probability distribution is $p(s_1) = \delta$, $p(s_2) = 1 - \delta$, where $\delta \geq 1/2$. Then $Pd_0 = 2/3$ for any value of δ. However, Pd_i depends on δ, as follows. If 1 or 3 is observed in the channel, then the opponent succeeds with probability 1/2. However, if 2 is the message in the channel, then the opponent can deceive the transmitter/receiver with probability δ ($\geq 1/2$) by substituting message 3. The probability of observing 1, 2, and 3 are respectively $2\delta/3$, 1/3, and $2(1 - \delta)/3$. Hence, $Pd_1 = (1 + \delta)/3$, which exceeds 1/2 (unless $\delta = 1/2$).

	s_1	s_2
e_1	1	2
e_2	1	3
e_3	2	3

Next, let us mention the attributes of the codes we study in this paper in relation to the taxonomy of authentication schemes Simmons has given in [23]. In the terminology of Simmons's taxonomy, we are studying codes that are unconditionally secure, both with and without secrecy, but without arbitration. All the codes in this paper are unconditionally secure, but in a given application it may be sufficient to use codes which offer only *computational* security. Computational security is when the security is based on the assumed difficulty of solving some problem, e.g., RSA [14] is based on the infeasibility of factoring large integers. Finally, we note that the codes described in this paper require that the transmitter and receiver trust each other, since either one can cheat the other in various ways. Simmons has constructed authentication codes in [22] which include an arbiter who can determine with high probability when one of the transmitter or receiver is cheating (see also [3]).

For descriptions of authentication and secrecy codes in relation to other aspects of cryptography, we refer to [11], [12], and [24].

2. Secrecy Codes and Perpendicular Arrays

The following theorem generalizes Shannon's basic result that a code which achieves perfect onefold secrecy must satisfy $b \geq k$.

Theorem 2.1. *If a code achieves perfect L-fold secrecy, then*

$$b \geq \binom{k}{L}.$$

Proof. Let e_0 be any encoding rule and let $M_1 \subseteq M(e_0)$, $|M_1| = L$. Let S_1 be any set of L source states. Assume there is no encoding rule e_1 such that $S_1 = f_{e_1}(M_1)$. Then $p(S_1|M_1) = 0 \neq p(S_1)$. Hence, we do not have perfect L-fold secrecy. Consequently, there are at least $\binom{k}{L}$ encoding rules e such that $M_1 \subseteq M(e)$. Therefore,

$$b \geq \binom{k}{L}. \qquad \square$$

We can construct codes which meet the above bound with equality using a type of combinatorial design known as a perpendicular array. A *perpendicular array* $\mathrm{PA}_\lambda(t, k, v)$ is a $\lambda \cdot \binom{v}{t} \times k$ array, A, of the symbols $\{1, \ldots, v\}$, which satisfies the following properties:

(i) Every row of A contains k distinct symbols.
(ii) For any t columns of A, and for any t distinct symbols, there are precisely λ rows r of A such that the t given symbols all occur in row r in the given t columns.

For $t \geq 2$, it is easy to see that the property (i) is implied by the other assumptions. For information on PAs see [7], [9], and [13].

In using PAs to construct secrecy codes, the following result is important.

Theorem 2.2 [9, Theorem 1.1]. *Assume* $0 \leq t' \leq t$ *and*

$$\binom{k}{t} \geq \binom{k}{t'}.$$

Then, a $\mathrm{PA}_\lambda(t, k, v)$ *is also a* $\mathrm{PA}_{\lambda(t')}(t', k, v)$, *where*

$$\lambda(t') = \lambda \cdot \binom{v-t'}{t-t'} \Big/ \binom{t}{t'}.$$

Hence,

$$\lambda \cdot \binom{v-t'}{t-t'} \equiv 0 \mod \binom{t}{t'}.$$

Proof. Let A be a $\mathrm{PA}_\lambda(t, k, v)$, and name the columns by $1, \ldots, k$. Let Y be any set of t' distinct symbols. For any set J' of t' columns, define $I(J')$ to be the number of rows of A in which the symbols in Y are all contained in the columns in J'. We obtain some linear equations in the $I(J')$ as follows. For any set J of t columns, we get an equation

$$\sum_{J' \subseteq J, |J'|=t'} I(J') = \lambda \cdot \binom{v-t'}{t-t'}.$$

In this way we get $\binom{k}{t}$ equations in $\binom{k}{t'}$ unknowns. If

$$\binom{k}{t} \geq \binom{k}{t'},$$

then the system has the unique solution

$$I(J') = \lambda \cdot \binom{v-t'}{t-t'} \Big/ \binom{t}{t'}$$

for every J'. Consequently, A is a $\mathrm{PA}_{\lambda(t')}(t', k, v)$, where $\lambda(t')$ is as above. □

We can now prove that secrecy codes can be obtained from PAs.

Theorem 2.3. *If there exists a* $\mathrm{PA}_\lambda(t, k, v)$, *where* $k \geq 2t - 1$, *then there is a code for* k *source states with* v *messages and* $\lambda \cdot \binom{v}{t}$ *encoding rules, which achieves perfect* t*-fold secrecy.*

Proof. Let A be a $\mathrm{PA}_\lambda(t, k, v)$. We construct an encoding rule from each row r of A: for each row $r = (x_1, \ldots, x_k)$, and for each source state s $(1 \leq s \leq k)$, define $e_r(s) = x_s$. Use each encoding rule with probability $1/\lambda \cdot \binom{v}{t}$.

It is necessary only to prove that we have perfect t'-fold secrecy for all $t' \leq t$. Since $k \geq 2t - 1$, we have

$$\binom{k}{t} \geq \binom{k}{t'};$$

hence A is a $\mathrm{PA}_{\lambda(t')}(t', k, v)$, by Theorem 2.2. Therefore, any set of t' messages corresponds equally often to every possible set of t' source states. It is now an easy computation that for every set S_1 of t' source states, and for every set M_1 of t' messages, we have $p(S_1|M_1) = p(S_1)$. For

$$\begin{aligned} p(S_1|M_1) &= \frac{p(M_1|S_1) \cdot p(S_1)}{p(M_1)} \qquad \text{(by Bayes' theorem)} \\ &= \frac{(\lambda(t')/b) \cdot p(S_1)}{\sum\limits_{\{e:\, M_1 \subseteq M(e)\}} p(e) \cdot p(f_e(M_1))} \\ &= \frac{(\lambda(t')/b) \cdot p(S_1)}{\sum\limits_{\{S \subseteq \mathscr{P}:\, |S|=t'\}} \; \sum\limits_{\{e:\, S = f_e(M_1)\}} (1/b) \cdot p(S)} \\ &= \frac{(\lambda(t')/b) \cdot p(S_1)}{(1/b) \cdot \lambda(t')} \\ &= p(S_1), \end{aligned}$$

as desired. This completes the proof. □

We note that the condition $k \geq 2t - 1$ in Theorem 2.3 is necessary, as shown by the following example.

Example 2.1. The following is a $\mathrm{PA}_1(1, 4, 4)$ and also a $\mathrm{PA}_1(3, 4, 4)$, but it is not a $\mathrm{PA}_\lambda(2, 4, 4)$ for any λ. Hence, it provides perfect onefold secrecy, but it does not

provide perfect twofold secrecy:

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

Observe that the secrecy code constructed from a PA (via Theorem 2.3) is optimal if and only if $\lambda = 1$ and $k = v$. In fact, the existence of an optimal L-fold secrecy code implies the existence of a $\mathrm{PA}_1(L, k, k)$ (in the case $L = 1$, this was noted by Shannon [16, p. 681]).

Theorem 2.4. *Assume there is an optimal L-fold secrecy code for k source states. Then there is a* $\mathrm{PA}_1(L, k, k)$.

Proof. Let e_0 be any encoding rule and let $M_1 \subseteq M(e_0)$, $|M_1| = L$. *Let* S_1 be any set of L source states. As in the proof of Theorem 2.1, there is at least one encoding rule e_1 such that $S_1 = f_{e_1}(M_1)$. In order for

$$b = \binom{k}{L},$$

there must be exactly one such encoding rule. Consequently, $M_1 \subseteq M(e)$ for all $\binom{k}{L}$ encoding rules. Now, there are $\binom{k}{L}$ different L-subsets of messages which are contained in $M(e_0)$. Each of these occurs in $\binom{k}{L}$ encoding rules. On the other hand, each of the $\binom{k}{L}$ encoding rules contains $\binom{k}{L}$ different L-subsets of messages. It follows that $M(e_0) = M(e)$ for every encoding rule e, and that the encoding matrix is a $\mathrm{PA}_1(L, k, k)$ on the symbols in $M(e_0)$. □

Hence, the arrays $\mathrm{PA}_1(t, v, v)$ are of interest. Such arrays are known to exist as follows.

Theorem 2.5. *For all integers $v \geq 1$, there is* $\mathrm{PA}_1(1, v, v)$. *Hence, for all $v \geq 1$, there is an optimal code for v source states having perfect onefold secrecy.*

Proof. Any Latin square of order v is a $\mathrm{PA}_1(1, v, v)$. □

Let us give a brief description of the Vernam one-time pad [28] in this setting. Assume the source space $\mathcal{S}$ consists of all binary n-tuples (so $k = 2^n$). Given any binary n-tuple w, define an encoding rule e_w by $e_w(s) = w + s$, where addition is componentwise addition modulo 2. Note that a message m is decoded by the same method: $s = w + m$. We have that $b = 2^n = k$. It is not difficult to see that the

encoding matrix is a Latin square of order 2^n, i.e., a $PA_1(1, 2^n, 2^n)$. Of course, this particular Latin square makes encoding and decoding very easy.

Theorem 2.6. *For any odd prime power $q \geq 3$, there is a $PA_1(2, q, q)$. Hence, for all such q, there is an optimal code for q source states having perfect twofold secrecy.*

Proof. This is a construction of Mullin *et al.* [13, Corollary 2.5]. Let g be a primitive element in the Galois field $GF(q)$. For $0 \leq i \leq (q-3)/2$ and for all $x \in GF(q)$, define a row

$$x \quad x + g^i \quad x + g^{i+1} \quad x + g^{i+2} \quad x + g^{i+3} \quad \cdots \quad x + g^{i+q-2}.$$

It is easy to check that the resulting array is a $PA_1(2, q, q)$. □

Some examples with $t \geq 3$ come from homogeneous permutation groups. A permutation group G is said to have *degree* n if it acts on a set, say S, of n symbols. Group G is defined to be *t-homogeneous* if, for all t-subsets $S_1, S_2 \subseteq S$, there are the same number of permutations $\pi \in G$ such that $(S_1)^\pi = S_2$. The number of such π must be $|G|/\binom{n}{t}$. It is clear that if we write down the permutations in a t-homogeneous group of degree n as the rows of an array, then we obtain a PA, as follows.

Theorem 2.7 [27, p. 10]. *Assume G is a t-homogeneous permutation group of degree n. Then there is a $PA_\lambda(t, n, n)$, where*

$$\lambda = |G|/\binom{n}{t}.$$

Theorem 2.8 [27, Lemma 3.4]. *There exists a $PA_1(3, v, v)$ for $v = 8$ and 32.*

Proof. The groups AGL(1, 8) and AΓL(1, 32) are 3-homogeneous (see, for example, [1]). Hence, they give rise to PAs with $\lambda = 1$. □

Example 2.2. A $PA_1(3, 8, 8)$. Develop the following rows modulo 7:

x	0	1	2	3	4	5	6
0	x	3	6	1	5	4	2
1	3	x	4	0	2	6	5
2	6	4	x	5	1	3	0
3	1	0	5	x	6	2	4
4	5	2	1	6	x	0	3
5	4	6	3	2	0	x	1
6	2	5	0	4	3	1	x

The above three theorems provide examples of optimal L-fold secrecy codes when $L = 1$, 2, and 3. It seems that no examples are known when $L \geq 4$. We have the following infinite class of $PA_3(3, v, v)$.

Theorem 2.9 [27, Theorem 3.5]. *There exists a* $PA_3(3, q+1, q+1)$ *for all prime powers* $q \equiv 3$ modulo 4. *Hence, for all such q, there is a code for* $q+1$ *source states with* $(q^3 - q)/2$ *encoding rules, having perfect threefold secrecy.*

Proof. The group PSL(2, q) is 3-homogeneous of degree $q+1$ if q is a prime power and $q \equiv 3$ modulo 4 (see [1]). Hence, it gives rise to PA with $\lambda = 3$. □

We also have two examples of $PA_4(4, v, v)$.

Theorem 2.10 [27, p. 12]. *There exists a* $PA_4(4, v, v)$ *for* $v = 9$ *and* 33.

Proof. The groups PGL(2, 8) and PΓL(2, 32) are both 4-homogeneous (see [1]), and yield the desired PAs. □

Finally, we prove that a code which achieves perfect L-fold secrecy for some particular source probability distribution will do so for an arbitrary source probability distribution.

Theorem 2.11. *Assume a code achieves perfect L-fold secrecy for a given source probability distribution* p_0. *Then the same code achieves perfect L-fold secrecy for an arbitrary source probability distribution* p_1.

Proof. Let p denote the probability distribution on the encoding rules. The condition for perfect L-fold secrecy (with respect to probability distribution p_0) is that, for every $L' \leq L$, for every set M_1 of L' messages observed in the channel, and for every set S_1 of L' source states, we have that $p_0(S_1 | M_1) = p_0(S_1)$. By Bayes' theorem, this is equivalent to $p_0(M_1) = p_0(M_1 | S_1)$, or that

$$\sum_{\{e: M_1 \subseteq M(e)\}} p(e) \cdot p_0(f_e(M_1)) = \sum_{\{e: S_1 = f_e(M_1)\}} p(e). \qquad (*)$$

We want to prove an analogous equality with respect to probability distribution p_1. We compute the following:

$$\begin{aligned}
\sum_{\{e: M_1 \subseteq M(e)\}} p(e) \cdot p_0(f_e(M_1)) &= \sum_{\{S \subseteq \mathscr{P}: |S| = L'\}} p_1(S) \sum_{\{e: S = f_e(M_1)\}} p(e) \\
&= \sum_{\{S \subseteq \mathscr{P}: |S| = L'\}} p_1(S) \sum_{\{e: M_1 \subseteq M(e)\}} p(e) \cdot p_0(f_e(M_1)) \quad (\text{by}(*)) \\
&= 1 \cdot \sum_{\{e: S_1 = f_e(M_1)\}} p(e) \quad (\text{by}(*))
\end{aligned}$$

as desired. This completes the proof. □

3. Codes Providing Authentication and Secrecy: $L_A = L_S - 1$

In this section we investigate the existence of $(t, t-1)$-codes (i.e., $L_A = L_S - 1$). First, let us prove the combinatorial lower bound on Pd_i.

Theorem 3.1 [11, p. 12]. *In an authentication code with k source states and v messages, $Pd_i \geq (k-i)/(v-i)$.*

Proof. Let x_j $(1 \leq j \leq i+1)$ be distinct messages $(i \geq 0)$. Assume an opponent observes the i messages x_j $(1 \leq j \leq i)$ in the channel, and then sends x_{i+1}. Denote the probability that the message x_{i+1} is accepted by the receiver as authentic by payoff(x_{i+1}). Then

$$\text{payoff}(x_{i+1}) = \frac{\sum_{\{e: x_1, x_2, \ldots, x_{i+1} \in M(e)\}} p(e) \cdot p(\{s_1, \ldots, s_i\} = \{f_e(x_1), \ldots, f_e(x_i)\})}{\sum_{\{e: x_1, x_2, \ldots, x_i \in M(e)\}} p(e) \cdot p(\{s_1, \ldots, s_i\} = \{f_e(x_1), \ldots, f_e(x_i)\})}.$$

It follows that

$$\sum_{x_{i+1} \notin \{x_1, \ldots, x_i\}} \text{payoff}(x_{i+1}) = k - i.$$

Hence, there must be some x_{i+1} such that payoff$(x_{i+1}) \geq (k-i)/(v-i)$. For every set of i messages $\{x_j: 1 \leq j \leq i\}$, determine such an x_{i+1}. This defines a substitution strategy in which the transmitter/receiver can be deceived with probability at least $(k-i)/(v-i)$. □

Next, we prove a lower bound on the number of encoding rules required in an $(L_S, L_S - 1)$-code.

Theorem 3.2 [26, Theorem 2]. *If a code achieves perfect L_S-fold secrecy and is $(L_S - 1)$-fold secure against spoofing, then*

$$b \geq \binom{v}{L_S}.$$

Proof. Let M_1 be a set of $i \leq L_S - 1$ messages which are valid under a particular encoding rule. Let x be any message not in M_1. Assume there is no encoding rule under which all messages in $M_1 \cup \{x\}$ are valid. Then a slight modification of the proof of Theorem 3.1 shows that we would have $Pd_i > (k-i)/(v-i)$, a contradiction. Hence, it follows that every L_S-subset of messages is valid under at least one encoding rule. Now, the code has perfect L_S-fold secrecy. Hence, the proof of Theorem 2.1 states that if an L_S-subset of messages is valid under some encoding rule, then it must be valid under at least $\binom{k}{L_S}$ encoding rules (corresponding to every possible set of L_S source states).

We now count pairs of the form (e, M_1), where $e \in \mathscr{E}$, $|M_1| = L_S$, and $M_1 \subseteq M(e)$. If we choose e first, and then M_1, we see that the number of such pairs is exactly $b \cdot \binom{k}{L_S}$. On the other hand, suppose we choose M_1, and then e. Set M_1 can be chosen in $\binom{v}{L_S}$ ways. Then, for each M_1, there are at least $\binom{k}{L_S}$ choices for e. Hence,

there are at least $\binom{v}{L_S}\binom{k}{L_S}$ pairs (e, M_1). Therefore,

$$b\cdot\binom{k}{L_S} \geq \binom{v}{L_S}\binom{k}{L_S} \quad \text{or} \quad b \geq \binom{v}{L_S}. \qquad \square$$

We saw in the last section that perpendicular arrays $\mathrm{PA}_\lambda(t, k, v)$ yielded codes having t-fold secrecy. If we employ a PA that enjoys an extra property, the code constructed from a $\mathrm{PA}_\lambda(t, k, v)$ will also be $(t-1)$-fold secure against spoofing, and hence will give rise to a $(t, t-1)$-code. This motivates the following definition. A $\mathrm{PA}_\lambda(t, k, v)$, A, is said to be an *authentication PA* (and is denoted $\mathrm{APA}_\lambda(t, k, v)$) if the following property holds:

> for any $t' \leq t-1$, and for any $t'+1$ distinct symbols $x_i\,(1 \leq i \leq t'+1)$, we have that among all the rows of A which contain all the symbols $x_i\,(1 \leq i \leq t'+1)$, the t' symbols $x_i\,(1 \leq i \leq t')$ occur in all possible subsets of t' columns equally often.

It can be shown (see Theorem 2.3 of [27]) that an $\mathrm{APA}_\lambda(t, k, v)$ is also an $\mathrm{APA}_{\lambda(t')}(t', k, v)$ for all $t' \leq t$. Hence, by Theorem 2.2, a necessary condition for the existence of an $\mathrm{APA}_\lambda(t, k, v)$ is that

$$\lambda(t'+1)\cdot\binom{k}{t'+1} \equiv 0 \quad \text{modulo} \binom{k}{t'}$$

for all t', $0 \leq t' \leq t-1$. We also observe that if $v \geq 2t-1$, then a $\mathrm{PA}_\lambda(t, v, v)$ is an APA.

The following result was proved by Stinson and Teirlinck in Theorem 2.4 of [27].

Theorem 3.3. *If there exists an* $\mathrm{APA}_\lambda(t, k, v)$, *then there is a* $(t, t-1)$*-code for* k *source states with* v *messages and* $\lambda\cdot\binom{v}{t}$ *encoding rules. The code is optimal if and only if* $\lambda = 1$.

Proof. Let A be an $\mathrm{APA}_\lambda(t, k, v)$. Construct the code as in Theorem 2.2. We need only verify that $Pd_i = (k-i)/(v-i)$, $0 \leq i \leq t-1$. Let $x_i\,(1 \leq i \leq t'+1)$ be distinct messages $(0 \leq t' \leq t-1)$. Assume an opponent observes the t' messages $x_i\,(1 \leq i \leq t')$ in the channel, and then sends $x_{t'+1}$. His chance of successful deception is calculated to be

$$\frac{\displaystyle\sum_{\{e:\, x_1, x_2, \ldots, x_{t'+1} \in M(e)\}} p(e)\cdot p(\{s_1, \ldots, s_{t'}\} = \{f_e(x_1), \ldots, f_e(x_{t'})\})}{\displaystyle\sum_{\{e:\, x_1, x_2, \ldots, x_{t'} \in M(e)\}} p(e)\cdot p(\{s_1, \ldots, s_{t'}\} = \{f_e(x_1), \ldots, f_e(x_{t'})\})}$$

$$= \frac{\displaystyle\sum_{\{e:\, x_1, x_2, \ldots, x_{t'+1} \in M(e)\}} p(\{s_1, \ldots, s_{t'}\} = \{f_e(x_1), \ldots, f_e(x_{t'})\})}{\displaystyle\sum_{\{e:\, x_1, x_2, \ldots, x_{t'} \in M(e)\}} p(\{s_1, \ldots, s_{t'}\} = \{f_e(x_1), \ldots, f_e(x_{t'})\})} \quad \text{(since } p(e) \text{ is constant).}$$

Now, the numerator of this expression is equal to

$$\lambda(t'+1)\cdot\binom{k}{t'+1}\Big/\binom{k}{t'},$$

since the PA is an APA. Also, the denominator is equal to $\lambda(t')$. Hence, the probability of deception is

$$\frac{\lambda(t'+1)\cdot\binom{k}{t'+1}}{\lambda(t')\cdot\binom{k}{t'}} = \frac{(t'+1)\cdot\binom{k}{t'+1}}{(v-t')\cdot\binom{k}{t'}} \left(\text{since } \frac{\lambda(t'+1)}{\lambda(t')} = \frac{t'+1}{v-t'} \text{ by Theorem 2.2}\right)$$

$$= \frac{k-t'}{v-t'}.$$

Hence, $Pd_{t'} = (k-t')/(v-t')$, as desired. □

Let us now consider the existence of APAs. In Theorem 6.2 of [5], some constructions are given for optimal (1, 0)-codes using generalized quadrangles. We observe now that an optimal (1, 0)-code for k source states with v messages exists whenever $k \leq v$.

Theorem 3.4. *For all $v \geq k \geq 1$, there exists an* $\mathrm{APA}_1(1, k, v)$. *Hence an optimal* (1, 0)*-code for k source states with v messages exists.*

Proof. Let the first row of the PA be $1\ 2\ 3\ \cdots\ k$. The obtain $v-1$ further rows by developing modulo v. □

The following theorem summarizes known results concerning $\mathrm{APA}_1(2, k, v)$ when $k = 3$ or 5 (see [26] and [27]).

Theorem 3.5. *There exists an* $\mathrm{APA}_1(2, 3, v)$ *if and only if $v \geq 7$ is odd. There exists an* $\mathrm{APA}_1(2, 5, v)$ *if $v \equiv 1$ or 5* modulo 10, $v \geq 11$, $v \neq 15$ [10].

The following infinite class of APAs was constructed in [7].

Theorem 3.6. *There exists an* $\mathrm{APA}_1(2, k, q)$ *if k is odd and $q \equiv 1$* modulo $2k$ *is a prime power. Hence, there exists an optimal* (2, 1)*-code with k source states and q messages for all such k and q.*

Proof. Let ω be a primitive element in the finite field $\mathrm{GF}(q)$, and let $\alpha = \omega^{(q-1)/k}$. For each $i = 1, \ldots, (q-1)/2k$, for each $j = 0, \ldots, k-1$, and for each $\beta \in \mathrm{GF}(q)$, define a row

$$\beta + \omega^i\alpha^j \quad \beta + \omega^i\alpha^{1+j} \quad \beta + \omega^i\alpha^{2+j} \quad \cdots \quad \beta + \omega^i\alpha^{k-1+j}.$$

The resulting array is an $\mathrm{APA}_1(2, k, q)$. □

Note that Example 1.2 is obtained from the above construction.

We remarked earlier that a $PA_\lambda(t, v, v)$ is an APA if $v \geq 2t - 1$. Hence, we have the following results from the PAs constructed in Theorems 2.8–2.10.

Theorem 3.7. *There exists an* $APA_1(3, v, v)$ *for* $v = 8$ *and* 32. *There exists an* $APA_3(3, q + 1, q + 1)$ *for all prime powers* $q \equiv 3$ modulo 4, $q \geq 7$. *There exists an* $APA_4(4, v, v)$ *for* $v = 9$ and 33.

Of course, codes where the number of messages equals the number of source states ($v = k$) are of no practical use for authentication, since the probability of deception is 1. We build codes with more messages than source states by means of a recursive construction using t-designs. A *t-design* $S_\lambda(t, k, v)$ is a set of k-subsets (called *blocks*) of a v-set, such that every t-subset occurs in exactly λ blocks.

Theorem 3.8 [27, Theorem 3.2]. *Assume there is a t-design* $S_{\lambda'}(t, k, v)$ *and an* $APA_\lambda(t, k, k)$. *Then there is an* $APA_{\lambda\cdot\lambda'}(t, k, v)$.

Proof. For each block in the $S_{\lambda'}(t, k, v)$, construct an $APA_\lambda(t, k, k)$. The union of all these $APA_\lambda(t, k, k)$ is an $APA_{\lambda\cdot\lambda'}(t, k, v)$. □

We use a class of 3-designs known as inversive geometries in our recursive construction. An *inversive geometry* is a 3-design $S_1(3, q + 1, q^d + 1)$, which exists for all prime powers q and for all $d \geq 1$ (see [29] or [1]). Hence, we obtain

Theorem 3.9 [27, Theorem 3.5]. *For any prime power* $q \equiv 3$ modulo 4 ($q \geq 7$), *and for any* $d \geq 1$, *there exists an* $APA_3(3, q + 1, q^d + 1)$. *Hence, a* (3, 2)*-code with* $q + 1$ *source states,* $q^d + 1$ *messages, and* $(q^{3d} - q^d)/2$ *encoding rules exists.*

Theorem 3.9 allows us to construct a (3, 2)-code for as many source states as desired (by taking q large enough), and incorporating any desired level of authentication security. For the resulting code has Pd_i approximately equal to $1/q^{d-1}$ ($i = 0, 1, 2$), which can be made arbitrarily small by taking d large enough.

We also obtain the following.

Theorem 3.10 [27, Theorem 3.6]. *For any* $d \geq 1$, *there exists an* $APA_1(3, 8, 7^d + 1)$ *and an* $APA_1(3, 32, 31^d + 1)$, *and hence an optimal* (3, 2)*-code with* $q + 1$ *source states and* $q^d + 1$ *messages, for* $q = 7$ *or* 31.

Finally, we prove the following partial converse to Theorem 3.3.

Theorem 3.11. *If there exists a code for k source states with v messages and* $\binom{v}{t}$ *encoding rules which achieves perfect* L_S*-fold secrecy and is* $(L_S - 1)$*-fold secure against spoofing for an arbitrary source probability distribution, and* $k \geq 2t - 1$, *then there is an* $APA_1(t, k, v)$.

Proof. From the proof of Theorem 3.2, we can see that every set of t messages corresponds to every set of t source states *exactly* once, since

$$b = \binom{v}{t}.$$

Hence, the encoding matrix is a $\mathrm{PA}_1(t, k, v)$. Since $k \geq 2t - 1$, the $\mathrm{PA}_1(t, k, v)$ is also a $\mathrm{PA}_{\lambda(t')}(t', k, v)$, for $1 \leq t' \leq t$.

We must show that the PA is an APA. Since the encoding matrix is a $\mathrm{PA}_1(t, k, v)$, every encoding rule must be used with equal probability in order to attain perfect secrecy. Now, let x_i $(1 \leq i \leq t' + 1)$ be distinct messages $(0 \leq t' \leq t - 1)$. Assume an opponent observes the t' messages x_i $(1 \leq i \leq t')$ in the channel, and then sends $x_{t'+1}$. As in Theorem 3.3, his chance of successful deception is calculated to be

$$\frac{\sum_{\{e: x_1, x_2, \ldots, x_{t'+1} \in M(e)\}} p(\{s_1, \ldots, s_{t'}\} = \{f_e(x_1), \ldots, f_e(x_{t'})\})}{\sum_{\{e: x_1, x_2, \ldots, x_{t'} \in M(e)\}} p(\{s_1, \ldots, s_{t'}\} = \{f_e(x_1), \ldots, f_e(x_{t'})\})},$$

and the denominator is equal to $\lambda(t')$ since the array is a $\mathrm{PA}_{\lambda(t')}(t', k, v)$.

Since this probability of deception is $(k - t')/(v - t')$, the numerator is determined, i.e.,

$$\sum_{\{e: x_1, x_2, \ldots, x_{t'+1} \in M(e)\}} p(\{s_1, \ldots, s_{t'}\} = \{f_e(x_1), \ldots, f_e(x_{t'})\}) = \lambda(t') \cdot (k - t')/(v - t').$$

This must be true for every source probability distribution. Hence, it follows that in the encoding rules in the above summation, the t' symbols x_i $(1 \leq i \leq t')$ occur in all possible subsets of t' columns equally often. Thus, the PA is an APA. □

4. Codes Providing Authentication and Secrecy: $L_A = L_S$

Next, we turn our attention to (t, t)-codes (i.e., $L_A = L_S$). First, we prove a lower bound on the number of encoding rules when it is required that the code attain the desired levels of security for an arbitrary source probability distribution.

Theorem 4.1. *If a code achieves perfect L_S-fold secrecy and is L_S-fold secure against spoofing for an arbitrary source probability distribution, then*

$$b \geq \binom{v}{L_S} \cdot \frac{v - L_S}{k - L_S}.$$

Proof. Using an argument similar to the proof of Theorem 3.2, we see that every set of $L_S + 1$ messages is valid under at least one encoding rule, and every set of L_S messages encodes every possible set of L_S source states. Let M_1 be a set of L_S messages which is valid under some encoding rule e_0, and denote $S_1 = f_{e_0}(M_1)$. Assume the source probability distribution is such that $p(S_1) = 1 - \varepsilon$, for some ε close to 0. Define $E' = \{e: \{e(s): s \in S_1\} = M_1\}$, In order that the code be L-fold secure against spoofing for the given source probability distribution, it must be the

case that, for every message $x \notin M_1$, there is an encoding rule $e \in E'$ under which x is valid. Hence, $|E'| \geq (v - L_S)/(k - L_S)$.

The stated bound on b now follows by counting triples of the form (e, S_1, M_1), where $e(S_1) = M_1$. If we pick e, and then S_1, then M_1 is determined uniquely. Hence, the number of such triples is exactly $b \cdot \binom{k}{L_S}$. On the other hand, if we pick M_1, then S_1, and then e, we see that the number of such triples is at least

$$\binom{v}{L_S} \cdot \binom{k}{L_S} \cdot \frac{v - L_S}{k - L_S}. \qquad \square$$

Remark. In Theorem 5.3 of [5] it was claimed that if an (L_S, L_A)-code exists, where $L_S \leq L_A + 1$, then

$$b \geq \binom{v}{L_A + 1} \cdot \binom{k}{L_S} \Big/ \binom{k}{L_A + 1}. \qquad (**)$$

However, the proof of this bound given in [5] is incorrect. We have three observations:

(1) In the case $L_S = L_A$, we have shown in Theorem 4.1 that the bound (**) *does* hold if the security is required for an arbitrary source probability distribution. However, the proof of (**) given in [5] does not make use of any assumption about the source probability distribution. In Section 6, we construct (1, 1)-codes for equiprobable source distributions where the number of encoding rules is *less* than the bound (**) by a factor of k. Hence, the extra assumption that the code be secure for an arbitrary source probability distribution is a necessary assumption in this case.

(2) In the case $L_S = L_A + 1$, the bound (**) reduces to the bound proved in Theorem 3.2. Here, it is unnecessary to make any assumptions regarding the source probability distribution.

(3) In the cases $L_S < L_A$, we do not know if the bound (**) is valid or not.

A (t, t)-code for an arbitrary source probability distribution is defined to be *optimal* if the number of encoding rules meets the bound of Theorem 4.1 with equality. We give a construction for (t, t)-codes for arbitrary source probability distributions, which generalizes a construction for (1, 1)-codes due to Stinson [25, Corollary 3.11]. The construction also uses t-designs.

Theorem 4.2. *If there is a* $\mathrm{PA}_\lambda(t, k, k)$ *and an* $S_{\lambda'}(t + 1, k, v)$, *where* $k \geq 2t - 1$, *then there is a* (t, t)*-code for* k *source states, for an arbitrary source probability distribution, having* v *messages and*

$$\frac{\lambda\lambda'(v - t)}{k - t} \cdot \binom{v}{t}$$

encoding rules.

Proof. For every block B of the $S_{\lambda'}(t + 1, k, v)$, construct a $\mathrm{PA}_\lambda(t, k, k)$ on the points in B. Any $S_{\lambda'}(t + 1, k, v)$ is also an $S_{\lambda''}(t, k, v)$, where $\lambda'' = (\lambda'(v - t))/(k - t)$.

Hence, this produces a $PA_{\lambda\cdot\lambda''}(t, k, v)$, and the resulting code has perfect t-fold secrecy, since $k \geq 2t - 1$. Also, the number of encoding rules is

$$\frac{\lambda\lambda'(v-t)}{k-t}\cdot\binom{v}{t}.$$

It remains to verify that the code is t-fold secure against spoofing. The code is $(t-1)$-fold secure against spoofing, by Theorems 3.3 and 3.8. So, to prove the code is t-fold secure against spoofing, let x_i $(1 \leq i \leq t+1)$ be distinct messages. Assume an opponent observes the t messages x_i $(1 \leq i \leq t)$ in the channel, and then sends x_{t+1}. In a similar fashion as Theorem 3.3, it can be proved that his chance of successful deception is $\lambda'/\lambda'' = (k-t)/(v-t)$. Hence, $Pd_t = (k-t)/(v-t)$, as desired. □

Note that the code constructed above is optimal if and only if $\lambda = \lambda' = 1$. We can apply this theorem to get infinite classes of optimal (1, 1)-codes for arbitrary source probability distributions.

Theorem 4.3 [25, Corollary 3.11]. *Assume there is an $S_1(2, k, v)$. Then there is an optimal (1, 1)-code for an arbitrary source probability distribution, with k source states and v messages.*

Proof. We have noted that a $PA_1(1, k, k)$ exists for all k. □

As an illustration, we can apply this result using projective geometries (see [1]). If we take the lines of the projective geometry of order q and dimension d as blocks, we obtain a design $S_1(2, q+1, (q^{d+1}-1)/(q-1))$. These are known to exist (for all $d \geq 1$) whenever q is a prime power. Hence, we have

Theorem 4.4. *For all prime powers q and for all $d \geq 2$, there is an optimal (1, 1)-code for an arbitrary source probability distribution, with $q+1$ source states and $(q^{d+1}-1)/(q-1)$ messages.*

We also obtain a class of (2, 2)-codes for an arbitrary source probability distribution, where the number of encoding rules is twice the optimal value. The construction makes use of orthogonal arrays, which we now define. An *orthogonal array* $OA(k, v)$ is a $v^2 \times k$ array, A, of the symbols $\{1, \ldots, v\}$, which satisfies the following property:

> for any t columns $c_1, \ldots, c_t$ of A, and for any t distinct symbols $x_1, \ldots, x_t$, there is a unique row r of A such that x_i occurs in column c_i of row r, for $1 \leq i \leq t$.

For any prime power q, it is well known that there is an $OA(q, q)$ with the property that it constains q different "constant" rows, each of which contains one symbol q times. (This array can be constructed from an affine plane of order q, which is a 2-design $S_1(2, q, q^2)$.) If these q rows are removed, then we have a $PA_2(2, q, q)$.

Theorem 4.5. *For any Mersenne prime q, there is a (2, 2)-code for an arbitrary source probability distribution, with $q+1$ source states, q^d+1 messages, and $(q^d+1)(q^d)(q^d-1)/(q-1)$ encoding rules.*

Proof. Since $q+1=2^n$, for some n, there is an $OA(q+1, q+1)$. An $S_1(3, q+1, q^d+1)$ is an inversive geometry. Apply Theorem 4.2. □

Also, we construct an optimal (2, 2)-code whenever a Fermat prime exists.

Theorem 4.6. *For any Fermat prime $q=2^n+1$, there is an optimal (2, 2)-code for an arbitrary source probability distribution, with q source states and $(q-1)^d+1$ messages.*

Proof. There is a $PA_1(2, q, q)$; and an $S_1(3, q, (q-1)^d+1)$ exists since $q-1=2^n$ is a prime power. Apply Theorem 4.2. □

Finally, we prove a weak converse to Theorem 4.2.

Theorem 4.7. *If there exists a code for k source states with v messages and*

$$\binom{v}{t}\cdot\frac{v-t}{k-t}$$

encoding rules which achieves perfect t-fold secrecy and is t-fold secure against spoofing for an arbitrary source probability distribution, then there is an $S_\lambda(t+1, k, v)$, where

$$\lambda=\binom{k}{t}.$$

Proof. Let M_1 be a set of t messages, and let $x\notin M_1$. Let S_1 be any set of t source states, and define $E'=\{e: \{e(s): s\in S_1\}=M_1\}$. From the proof of Theorem 4.1, in order to have

$$b=\binom{v}{t}\cdot\frac{v-t}{k-t},$$

it must be the case that $|E'|=(v-L_S)/(k-L_S)$. Also, x must occur in exactly one encoding rule in E'. It follows that the rows of the encoding matrix form a $(t+1)$-design $S_\lambda(t+1, k, v)$, where

$$\lambda=\binom{k}{t}.$$ □

We note that the existence of the code hypothesized in Theorem 4.7 need *not* imply the existence of a $(t+1)$-design $S_1(t+1, k, v)$. For an $APA_1(2, 3, v)$ gives rise to a (2, 1)-code, which is certainly a (1, 1)-code. The number of encoding rules, $\binom{v}{2}$, is indeed equal to

$$\binom{v}{1}\cdot\frac{v-1}{3-1}.$$

However, there exist $\mathrm{APA}_1(2, 3, v)$ for $v \equiv 5$ modulo 6, $v \geq 11$ (Theorem 3.5); whereas it can be shown by an elementary counting argument that an $S_1(2, 3, v)$ does not exist if $v \equiv 5$ modulo 6.

5. Codes Providing Authentication Without Secrecy

As mentioned in the introduction, a code with perfect disclosure cannot be even one-fold secure against spoofing. In general, we have the following lower bounds on Pd_i for such codes.

Theorem 5.1. *If a code has perfect disclosure, then $Pd_0 \geq k/v$. Moreover, if $Pd_0 = k/v$, then $Pd_L \geq k/v$ for any $L \geq 0$.*

Proof. By Theorem 3.1, $Pd_0 \geq k/v$. Assume that $Pd_0 = k/v$. In order that the code has perfect disclosure, there must be v/k possible messages encoding each possible source state. Then an argument similar to that used in the proof of Theorem 3.1 shows that $Pd_L \geq k/v$ for any $L \geq 0$. □

The following example illustrates that we can sometimes decrease the probability Pd_1 at the expense of increasing the probability Pd_0, at least for some source probability distributions.

Example 5.1. The following code has $v = 9$ messages, $k = 3$ source states, $b = 16$ encoding rules, and has perfect disclosure.

	s_1	s_2	s_3
e_1	1	2	6
e_2	1	3	6
e_3	1	4	6
e_4	1	5	6
e_5	1	2	7
e_6	1	3	7
e_7	1	4	7
e_8	1	5	7
e_9	1	2	8
e_{10}	1	3	8
e_{11}	1	4	8
e_{12}	1	5	8
e_{13}	1	2	9
e_{14}	1	3	9
e_{15}	1	4	9
e_{16}	1	5	9

Assume the source distribution is $p(s_1) = 98/100$, $p(s_2) = 1/100$, $p(s_3) = 1/100$, and each encoding rule is used with probability 1/16. Then $Pd_0 = 1$ $(> k/v)$, since message 1 is always accepted as authentic. However,

$$Pd_1 = \frac{98}{100} \cdot \frac{1}{4} + \frac{1}{100} \cdot 1 + \frac{1}{100} \cdot 1 = \frac{53}{200} \qquad (< k/v).$$

Theorem 5.2. *If a code has perfect disclosure, and* $Pd_i = k/v$ *for* $0 \leq i \leq L$, *then* $b \geq (v/k)^{L+1}$.

Proof. For $0 \leq i \leq L + 1$, we prove that every set of i messages corresponding to different source states is valid under at least one encoding rule. Since there are v/k messages corresponding to each encoding rule, this implies $b \geq (v/k)^{L+1}$. First, if $i = 0$, then every message must be valid under at least one encoding rule (otherwise, $Pd_0 > k/v$). As an induction hypothesis, assume that every set of i (≥ 0) messages corresponding to different source states is valid under at least one encoding rule. In order that $Pd_i = k/v$, the result must be true for every set of $i + 1$ messages corresponding to different source states. By induction, the result is true for sets of $L + 1$ messages. □

A perfect disclosure code for which $Pd_i = k/v$ for $0 \leq i \leq L$, and in which $b = (v/k)^{L+1}$, is said to be *optimal.* We can construct optimal perfect disclosure codes using transversal designs, which we now define. A *transversal design* $\mathrm{TD}_\lambda(t, k, n)$ is a triple $(X, \mathscr{G}, \mathscr{A})$, where X is a set of kn *points*, $\mathscr{G}$ is a partition of X into k *groups* of n points each, and $\mathscr{A}$ is a set of λn^t *blocks*, each of which meets each group in a point, such that every t-subset of points from distinct groups occurs in exactly λ blocks.

We have the following construction, which was first given in the case $t = 2$ and $\lambda = 1$ in [2]. The special case of this construction using $\mathrm{TD}_1(2, q + 1, q)$ was in fact the first construction given in the literature for authentication codes; see [6]. The verifications are routine, so we omit them.

Theorem 5.3. *If there exists a transversal design* $\mathrm{TD}_\lambda(t, k, n)$, *then there is a perfect disclosure code for k source states, having kn messages and* λn^t *blocks, and for which* $Pd_i = 1/n\ (= k/v)$ *for* $0 \leq i \leq t - 1$. *The code is optimal if and only if* $\lambda = 1$.

We also have the following (partial) converse to Theorem 5.3.

Theorem 5.4. *Assume there is an optimal perfect disclosure code for k source states, having v messages* and $(v/k)^t$ *encoding rules, and for which* $Pd_i = k/v$ *for* $0 \leq i \leq t - 1$. *Then there exists a transversal design* $\mathrm{TD}_1(t, k, n)$, *where* $n = v/k$.

Proof. As in the proof of Theorem 5.2, every set of t messages corresponding to t different source states occurs in at least one encoding rule. In order that $b = (v/k)^t$, "at least" must be "exactly." It is then easy to see that we have a $\mathrm{TD}_1(t, k, n)$. □

We now mention a family of transversal designs that are useful in the construction of Theorem 5.3.

Theorem 5.5 [8, Lemma 3.5]. *For any prime power q and for any* $t \leq q$, *there is a* $\mathrm{TD}_1(t, q + 1, q)$. *Hence, there is an optimal perfect disclosure code for* $q + 1$ *source states, having* $q^2 + q$ *messages and* q^t *encoding rules, and for which* $Pd_i = 1/q$ *for* $0 \leq i \leq t - 1$.

Proof. Let β be a primitive element of GF(q). For any t-tuple $(\alpha_0, \ldots, \alpha_{t-1})$ of elements of GF(q), define a row

$$\alpha_0 \quad \alpha_{t-1} \quad \sum_{j=0}^{t-1} \alpha_j \quad \sum_{j=0}^{t-1} \alpha_j \beta^j \quad \sum_{j=0}^{t-1} \alpha_j \beta^{2j} \quad \cdots \quad \sum_{j=0}^{t-1} \alpha_j \beta^{(q-2)j}.$$

The resulting array is a $\mathrm{TD}_1(t, q+1, q)$. □

The result gives us optimal codes where the same encoding rule can be used up to $q-1$ times. Contrast this with the families of codes in the previous sections where the only known infinite families of optimal (or near-optimal) codes allowed the same encoding rule to be used at most two or three times.

Finally, we give a construction for perfect disclosure codes from transversal designs having $k > n+1$.

Theorem 5.6 [25, Theorem 3.7]. *For any prime power q and for any $d \geq 2$, there is a $\mathrm{TD}_{q^{d-2}}(2, (q^d-1)/(q-1), q)$. Hence, there is a perfect disclosure code for $(q^d-1)/(q-1)$ source states, having $q(q^d-1)/(q-1)$ messages and q^d encoding rules, and for which $Pd_i = 1/q$ for $i = 0$ and 1.*

6. Authentication Codes for Equiprobable Source Distributions

In this section we give some constructions for (1, 1)-codes when the source states are independent and equiprobable. Prior to doing that, however, we consider codes providing authentication, with no secrecy assumptions. The following lower bound on the number of encoding rules applies to *any* authentication code, regardless of the source distribution. It was first proved by Massey [11] and Schobi [15].

Theorem 6.1. *If a code is L_A-fold secure against spoofing, then*

$$b \geq \binom{v}{L_A+1} \Big/ \binom{k}{L_A+1}.$$

Proof. As in the proof of Theorem 3.2, any set of L_A+1 messages is valid under at least one encoding rule. The bound follows. □

Note that this bound is less than the bound of Theorem 4.1 by a factor of $\binom{k}{L_A}$, and less than the bound of Theorem 3.2 by a factor of $\binom{k}{L_A+1}$.

As with other codes, the term optimal is used if the number of encoding rules meets the lower bound with equality. Optimal codes can sometimes be constructed if the source probability distribution is known to be equiprobable. We have the following construction, given by Schobi [15] and De Soete [5].

Theorem 6.2. *Assume there is a t-design $S_\lambda(t, k, v)$. Then there is an authentication code for k equiprobable source states, having v messages and $\lambda \cdot \binom{v}{t} \Big/ \binom{k}{t}$ encoding rules, that is $(t-1)$-fold secure against spoofing.*

Proof. Order each block of the t-design arbitrarily, and take these ordered blocks as encoding rules. First, we verify that $Pd_{t-1} = (k-t+1)/(v-t+1)$. Let x_i $(1 \leq i \leq t)$ be distinct messages, and assume an opponent observes the $t-1$ messages x_i $(1 \leq i \leq t-1)$ in the channel, and then sends x_t. His chance of successful deception is calculated to be

$$\frac{\sum_{\{e: x_1, x_2, \ldots, x_t \in M(e)\}} p(e) \cdot p(\{s_1, \ldots, s_{t-1}\} = \{f_e(x_1), \ldots, f_e(x_{t-1})\})}{\sum_{\{e: x_1, x_2, \ldots, x_{t-1} \in M(e)\}} p(e) \cdot p(\{s_1, \ldots, s_{t-1}\} = \{f_e(x_1), \ldots, f_e(x_{t-1})\})},$$

$$= \frac{|\{e: x_1, \ldots, x_t \in M(e)\}|}{|\{e: x_1, \ldots, x_{t-1} \in M(e)\}|}$$

(since $p(e)$ is constant and the source states are equiprobable)

$$= \frac{k-t+1}{v-t+1}.$$

Now, to see that $Pd_i = (k-i)/(v-i)$ for $0 \leq i \leq t-2$, note that, for any $t' \leq t$, an $S_\lambda(t, k, v)$ is also an $S_{\lambda'}(t', k, v)$ for some λ'. □

Conversely, we prove the following result.

Theorem 6.3. *Assume there is an authentication code for k equiprobable source states, having v messages and $\binom{v}{t}\Big/\binom{k}{t}$ encoding rules, that is $(t-1)$-fold secure against spoofing. Then there is a t-design $S_1(t, k, v)$.*

Proof. In order to attain equality in the bound of Theorem 6.1, every set of t messages must be valid under *exactly* one encoding rule. □

In general, a code constructed from a t-design using Theorem 6.2 will not provide perfect $(t-1)$-fold secrecy. However, in the case $t = 2$, we can also obtain perfect onefold secrecy with the same number of encoding rules, provided that $v - 1 \equiv 0$ modulo $k(k-1)$, as follows.

Theorem 6.4. *Assume there is an $S_1(2, k, v)$, where $v - 1 \equiv 0$ modulo $k(k-1)$. Then there is an optimal* (1, 1)-code for *k equiprobable source states, having v messages and $v(v-1)/(k(k-1))$ encoding rules.*

Proof. It is necessary to order every block of the $S_1(2, k, v)$, such that every point occurs in each possible position in exactly $(v-1)/(k(k-1))$ blocks. Clearly, a necessary condition for this to be possible is $v - 1 \equiv 0$ modulo $k(k-1)$, since every point occurs in exactly $(v-1)/(k-1)$ blocks. The condition is also sufficient, as follows. Let the design $S_1(2, k, v)$ have point set X and block set $\mathscr{A}$. Construct a bipartite graph, having bipartition $(X, \mathscr{A})$, where xA is an edge if and only if $x \in A$ $(x \in X, A \in \mathscr{A})$. Clearly, it suffices to find an edge-colouring of this graph using k

colours, so that each vertex $A \in \mathscr{A}$ is adjacent to one edge of each colour, and each vertex $x \in X$ is adjacent to $(v-1)/(k(k-1))$ edges of each colour. But this can be done by first "splitting" each vertex x into $(v-1)/(k(k-1))$ vertices, each having degree k, and then finding a proper edge-colouring of the resulting k-regular bipartite graph using k colours. Finally, use the edge-colouring to impose an ordering on each block, and then take these ordered blocks as the encoding rules, using each with equal probability. □

Hence, using projective geometries, we obtain the following.

Theorem 6.5. *For all prime powers q and for all even $d \geq 2$, there is an optimal $(1, 1)$-code for an equiprobable source probability distribution with $q+1$ source states, having $(q^{d+1}-1)/(q-1)$ messages.*

Proof. The projective geometry yields a design $S_1(2, q+1, (q^{d+1}-1)/(q-1))$. If d is even, then $v - 1 \equiv 0$ modulo $k(k-1)$. Apply Theorem 6.4. □

In the case $d = 2$, it is particularly convenient to construct the code. For, it is well known that the design $S_1(2, q+1, q^2+q+1)$ can be constructed by developing a single "base block" modulo q^2+q+1 (see, for example, Theorem 6.2 of [1]). Such a design is said to be *cyclic*. This automatically yields perfect secrecy without recourse to the technique used in the proof of Theorem 6.4.

Example 6.1. A $(1, 1)$-code for $k = 3$ equiprobable source states, having $v = 7$ messages, and $b = 7$ encoding rules, constructed from a cyclic $S_1(2, 3, 7)$. Use each encoding rule with probability 1/7.

	s_1	s_2	s_3
e_1	1	2	4
e_2	2	3	5
e_3	3	4	6
e_4	4	5	7
e_5	5	6	1
e_6	6	7	2
e_7	7	1	3

7. Summary

We summarize the main classes of codes constructed in this paper. Secrecy codes are studied in Section 2. The basic construction for t-fold secrecy codes uses a perpendicular array $\mathrm{PA}_\lambda(t, k, v)$, where $k \geq 2t-1$. The optimal situation is to have a $\mathrm{PA}_1(t, v, v)$ $(v \geq 2t-1)$. Unfortunately, examples of these are known only for $t = 1, 2$, and 3, and infinite classes are known to exist only for $t = 1$ and 2.

Perfect disclosure authentication codes are considered in Section 5. To obtain

Table 1. Some optimal and near-optimal classes of (L_S, L_A)-codes.

(L_S, L_A)	k	v	b	Optimal?	Authority
(1, 0)	Any integer	Any integer $\geq k$	v	Yes	Theorem 3.4
(1, 1)	$q+1$	q^2+q+1	$(q^2+q+1)(q+1)$	Yes	Theorem 4.4 ($d=2$)
(2, 1)	Any odd integer	$q \equiv 1 \bmod 2k$	$q(q-1)/2$	Yes	Theorem 3.6
(2, 2)	$q+1$ q a Mersenne prime	q^2+1	$(q^2+1)(q^2)(q+1)$	2 × optimal	Theorem 4.5 ($d=2$)
(2, 2)	q q a Fermat prime	q^2-2q+2	$(q^2-2q+2)(q)(q-1)^2/2$	Yes	Theorem 4.6 ($d=2$)
(3, 2)	$q+1$ $q \equiv 3$ modulo 4	q^2+1	$(q^6-q^2)/2$	3 × optimal	Theorem 3.9 ($d=2$)

such a code, we employ a transversal design $\mathrm{TD}_\lambda(t, k, v)$. This yields a code where $\mathrm{Pd}_i = k/v$ for $0 \leq i \leq t-1$. For authentication purposes, we would want k/v to be very small. The number of encoding rules is optimal if $\lambda = 1$. Fortunately, optimal codes can be constructed for any value of t (Theorem 5.5).

Next, in Table 1, we give a brief summary of the optimal and near-optimal classes of (L_S, L_A)-codes we constructed in Sections 3 and 4. These are all codes for an arbitrary source distribution. In Table 1, q denotes a prime power.

Finally, codes for equiprobable source distributions are studied in Section 6. Here, an optimal code which is $(t-1)$-fold secure against spoofing can be constructed from an $S_1(t, k, v)$. In the case of codes which are onefold secure against spoofing, the code obtained from an $S_1(2, k, v)$ can also provide perfect secrecy if $v - 1 \equiv 0$ modulo $k(k-1)$.

References

1. Th. Beth, D. Jungnickel, and H. Lenz, *Design Theory*, Bibliographisches Institut, Zurich, 1985.
2. E. F. Brickell, A few results in message authentication, *Congr. Numer.* **43** (1984), 141–154.
3. E. F. Brickell and D. R. Stinson, Authentication codes with multiple arbiters (Extended Abstract), in *Advances in Cryptology—EUROCRYPT* '88, Lecture Notes in Computer Science, vol. 330, Springer-Verlag, Berlin, 1988, pp. 51–55.
4. G. Castagnoli, Comments on Massey's concepts of perfect secrecy and perfect authenticity, *Alta Frequenza* **51** (1987), 227–228.
5. M. De Soete, Some constructions for authentication—secrecy codes, in *Advances in Cryptology—EUROCRYPT* '88, Lecture Notes in Computer Science, vol. **330**, Springer-Verlag, Berlin, 1988, pp. 57–75.
6. E. Gilbert, F. J. MacWilliams, and N. J. A. Sloane, Codes which detect deception, *Bell System Tech. J.* **53** (1974), 405–424.
7. A. Granville, A. Moisiadis, and R. Rees, Nested Steiner n-gon systems and perpendicular arrays, *J. Combin. Math. Combin. Comput.* **3** (1988), 163–167.
8. H. Hanani, A class of three-designs, *J. Combin. Theory Ser. A* **26** (1979), 1–19.
9. E. S. Kramer, D. L. Kreher, R. Rees, and D. R. Stinson, On perpendicular arrays with $t \geq 3$, *Ars Combin.*, to appear.
10. C. C. Lindner and D. R. Stinson, Steiner pentagon systems, *Discrete Math.* **52** (1984), 67–74.
11. J. L. Massey, Cryptography—a selective survey, in *Digital Communications*, 1986, pp. 3–21.

12. J. L. Massey, An introduction to contemporary cryptology, *Proc. IEEE* **76** (1988), 533–549.
13. R. C. Mullin, P. J. Schellenberg, G. H. J. van Rees, and S. A. Vanstone, On the construction of perpendicular arrays, *Utilitas Math.* **18** (1980), 141–160.
14. R. L. Rivest, A. Shamir, and L. M. Adelman, A method for obtaining digital signatures and public-key cryptosystems, *Comm. ACM* **21** (1978), 120–126.
15. P. Schobi, Perfect authentication systems for data sources with arbitrary statistics, presented at Eurocrypt '86.
16. C. E. Shannon, Communication theory of secrecy systems, *Bell System Tech. J.* **28** (1949), 656–715.
17. G. J. Simmons, A game theory model of digital message authentication, *Congr. Numer.* **34** (1982), 413–424.
18. G. J. Simmons, Verification of treaty compliance—revisited, in *Proc. IEEE* 1983 *Symposium on Security and Privacy*, Oakland, 1984, pp. 61–66.
19. G. J. Simmons, Message authentication: a game on hypergraphs, *Congr. Numer.* **45** (1984), 161–192.
20. G. J. Simmons, Authentication theory/coding theory, in *Advances in Cryptology*: *Proceedings of CRYPTO* 84, Lecture Notes in Computer Science, vol. 196, Springer-Verlag, Berlin, 1985, pp. 411–432.
21. G. J. Simmons, The practice of authentication, in *Advances in Cryptology*: *EUROCRYPT* 85, Lecture Notes in Computer Science, vol. 219, Springer-Verlag, Berlin, 1986, pp. 261–272.
22. G. J. Simmons, Authentication codes that permit arbitration, *Congr. Numer.* **59** (1987), 275–290.
23. G. J. Simmons, A natural taxonomy for digital information authentication schemes, in *Advances in Cryptology*: *CRYPTO* 87 *Proceedings*, Lecture Notes in Computer Science, vol. 293, Springer-Verlag, Berlin, 1988, pp. 269–288.
24. G. J. Simmons, A survey of information authentication, *Proc. IEEE* **76** (1988), 603–620.
25. D. R. Stinson, Some constructions and bounds for authentication codes, *J. Cryptology* **1** (1988), 37–51.
26. D. R. Stinson, A construction for authentication/secrecy codes from certain combinatorial designs. *J. Cryptology* **1** (1988), 119–127.
27. D. R. Stinson and L. Teirlinck, A construction for authentication/secrecy codes from 3-homogeneous permutation groups, *European J. Combin.*, to appear.
28. G. S. Vernam, Cipher printing telegraph systems for secret wire and radio telegraph communications, *J. Amer. Inst. Elec. Eng.* **45** (1926), 109–115.
29. E. Witt, Uber Steinersche Systeme, *Abh. Math. Sem. Univ. Hamburg* **12** (1938), 265–275.