Joint Conference of 1996 International Computer Symposium
December 19~21, Kaohsiung, Taiwan, R.O.C.

Block Codes for Collusion-Nonpermissible Secret Sharing Systems*

Phen-Lan Lin(1) and James G. Dunham(2)

(1)Department of Computer Science and Information management
Providence University
Shalu, Taichung, Taiwan
lan@simon.pu.edu.tw

(2)Department of Electrical Engineering
Southern Methodist University
Dallas, Texas, U.S.A.

Abstract

The (n, k) block codes are translated into (l, p, r, n) secret sharing schemes in which collusion between players are not permitted. The performance of both linear block codes and random block codes as (l, p, r, n) SSSs in terms of the relationship between reconstructability and privacy; and the trade-off between privacy and resiliency is compared to the performance bounds derived from a secret sharing model GS^3. It is demonstrated that linear codes, when viewed as (l, p, r, n) SSSs, can at best achieve one-half of the performance bounds asymptotically. It is also demonstrated that the random codes do not have privacy even though they can achieve the capacity of reconstructability and resiliency with high probability for sufficiently large n. Lastly, a hashed random code (HRC) is presented and shown to achieve the performance bounds asymptotically.

Secret Sharing Reconstructability Privacy Resiliency Collusion-nonpermissible (l, p, r, n) SSS secret sharing model Hashed random code HRC

1. Introduction

In cryptographic and large distributed systems, when a group of people share a common secret key, it is highly desirable to have robust key management such that a maximum level of privacy can be achieved while allowing resiliency and preserving reconstructability. Several schemes, including collusion-permissible and nonpermissible, have been devised to achieve certain level of these requirements [1, 2, 3, 4, 5, 6, 7].

In this paper, the (n, k) block codes are translated into (l, p, r, n) secret sharing schemes [7] in which collusion between players are not permitted. Based on the well established error-detecting, error-correcting, and erasures-recovering capability derived in coding theory, the relationship among reconstructability and resiliency of an (l, p, r, n) SSS and the code rate of both linear and random block codes are established. Privacy, which is the key requirement that distinguishes the (l, p, r, n) SSS from the coding problem, is also established. The performance of both codes as (l, p, r, n) SSSs in terms of the relationship between reconstructability and privacy; and the trade-off between privacy and resiliency is then compared to the performance bounds, which were established from a model GS^3 [7]. We show that linear codes, when viewed as (l, p, r, n) SSSs, can not achieve the capacity bounds; We also show that an (n, k) random code, when viewed as an (l, p, r, n) SSS, can actually achieve the capacity of reconstructability and resiliency with high probability for sufficiently large n and alphabet. But the drawback is that it does not have privacy because the level of secrecy decreases as shares are being released.

We then devise a hashed random code (HRC) which is an (n, k) random code with the index being hashed such that the space of the new secret becomes k-fold smaller than that of the original secret. We show that such a code can asymptotically be perfect and can achieve the capacity of reconstructability and resiliency with high probability. Lastly, some computer evaluations of the privacy of HRC are given for the non-asymptotic case.

2. Preliminary

- An (l, p, r, n) secret sharing scheme [7], abbreviated as (l, p, r, n) SSS, divides the secret S into n pieces of information called 'shares' $(s_0, s_1, \ldots, s_{n-1})$ in such a way that the following properties hold:

(i) Knowledge of any l or more shares make S easily computable, and l is called 'reconstructability';

(ii) Knowledge of any $p - 1$ or fewer correct shares leaves S completely undetermined in a sense that all possible values of S are equally likely, and p is called 'privacy'; and

(iii) No set of r or fewer incorrect shares can affect

*This work was partly supported by National Science Council, Republic of China, under Grant NSC86-2213-E-126-008

the correctness of S, and r is called 'resiliency'.

- The performance bounds of the (l, p, r, n) SSS is

$$\begin{cases} l & \geq k+t \quad \geq p+t \\ p+r & \leq n-\rho \end{cases} \qquad (1)$$

for sufficiently large n and N, where t is the number of errors, ρ is the number of missing pieces which have occurred, and N is the size of the alphabet.

3. Linear Codes as (l,p,r,n) SSSs

McEliece and Sarwate had related Shamir's (k, n) threshold scheme [1] to a special case of Reed-Solomon code [3]. In this paper, we relate the (l, p, r, n) SSS to a general (n, k) linear code. The results are stated in the theorems below.

Theorem 3.1: Viewing an (n, k) linear code as an (l, p, r, n) SSS, we have for reconstructability l, privacy p, and resiliency r that $l \geq k + 2t \geq p + 2t$, and $p + 2r \leq n - \rho$, where t is the number of errors and ρ is the number of missing shares.

Proof: Choose an i from $0, \ldots, k-1$ and let u_i of the message word $\mathbf{U}$ be the secret S, $u_{j, j=0,\ldots,k-1; j\neq i}$ be any arbitrary elements in a Galois Field GF(q), and $v_{l, l=0,\ldots,n-1}$ of the code word $\mathbf{V}$ be the shares given to the n participants, then we can view an (n, k) linear code as an (l, p, r, n) SSS. Since the secret S can be reconstructed once the code word $\mathbf{V}$ is recovered, finding l is equivalent to decoding a linear code. From [8, pp. 125], any received pattern of t errors and ρ erasures can be decoded provided that $d_{min} \geq 2t+1+\rho$. But $d_{min} \leq n-k+1$ for an (n, k) code over GF(q), so $2t+1+\rho \leq d_{min} \leq n-k+1$, and hence $n-\rho \geq k+2t$. Since $n - \rho$ is the number of pieces in the received pattern, thus the minimum number of shares required to reconstruct an (n, k) linear code (and hence the secret) in a t-error environment is $\geq k+2t$, and so we have

$$l \geq k + 2t. \qquad (2)$$

Also, we have $t \leq \frac{1}{2}(n-\rho-k)$, so the maximun number of errors that the code can tolerate is $\frac{1}{2}(n-\rho-k)$ and hence

$$r = \frac{1}{2}(n - \rho - k) \qquad (3)$$

The generator matrix of an (n, k) linear code has rank k, so there exists at least one set of k equations, which are generated by the k valid pieces of the code word, to provide an unique solution to the k unknown message pieces u_i's, and hence break the secret S. Thus $p \leq k$. For an non-systematic linear code, v_i's can be expressed explicitly as $v_i = u_0 + u_1\alpha_i + u_2\alpha_i^2 + \ldots + u_{k-1}\alpha_i^{k-1}$, where $\alpha_i, i = 0, 1, \ldots, n-1$ are the n non-zero, distinct elements in $GF(q)$. Assume any $k-1$ of v_i's are available, then we have $k-1$ distinct points (x_i, y_i) in the 2-dimensional plane, where y_i is the share v_i and x_i is the identifying index α_i. Note that $v_i = i$ in Shamir's scheme. So by the interpolation of polynomial argument that Shamir had used [1], we have $p = k$. But an (n, k) linear code can be systematic, that is, $v_{n-k+i} = u_i$ for $0 \leq i < k$ and $v_j = u_0 p_{0,j} + \ldots + u_{k-1} p_{k-1,j}$ where $p_{ij} \in$ GF(q) for $0 \leq i \leq k-1, 0 \leq j < n-k$ [9]. Since the secret is one dimension, v_{n-k+i} can break the secret u_i for $0 \leq i < k$. Hence $p = 1$ for this case. Thus, for a general linear (n, k) code, we have

$$1 \leq p \leq k. \qquad (4)$$

Combining (2), (3), and (4), we have $l \geq k + 2t \geq p + 2t$, and $p + 2r \leq n - \rho$. //

Fig. 1 shows a relationship between *reconstructability* and *privacy* of an (n, k) linear code when viewed as an (l, p, r, n) SSS. The point line represents the bound for optimality and the shaded area is the achievable region for linear codes, while the bold line represents the performance bound of an (l, p, r, n) SSS.

Remark: 1. For an error-free (l, p, r, n) SSS, we have $t = 0$; thus, 'reconstructability' can be interpreted as 'erasures-recovering' capability in coding theory.
2. Shamir's (k, n) threshold scheme [1] is in fact an error-free (l, p, r, n) SSS, and its $(\lfloor \frac{n}{2}, \frac{n}{2} \rfloor)$-capacity sits right on the optimal bound for linear codes as shown in Fig. 1.

Fig. 2 depicts a trade-off between privacy and resiliency of an (n, k) linear code when viewed as an (l, p, r, n) SSS, given ρ missing shares. The shaded area is the achievable region for linear codes while Reed-Solomon codes are right on the optimal bound for linear codes, which is represented by the point line. The bold line represents the performance bound of an (l, p, r, n) SSS.

Remark: 1. For a full-participating (l, p, r, n) SSS, we have $\rho = 0$; thus, resiliency can be interpreted as 'error-correcting' capability in coding theory.
2. Ben-Or *et al.*'s scheme [5] is a full-participating (l, p, r, n) SSS, and its result of $(p = n/3, r = n/3)$ lies on the bound for linear codes when $\rho = 0$ as shown in Fig. 2.

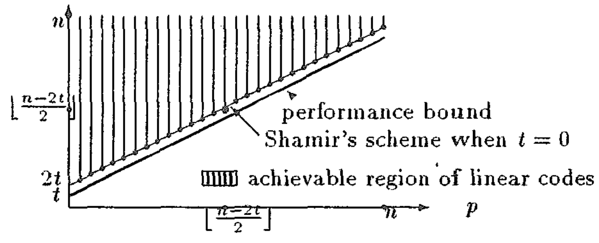

Fig. 1. Reconstructability l vs. privacy p for an (n, k) linear code when viewed as an (l, p, r, n) SSS, where t is the number of errors occurred.

4. Random Codes as (l,p,r,n) SSSs

4.1 Reconstructability of Random Codes

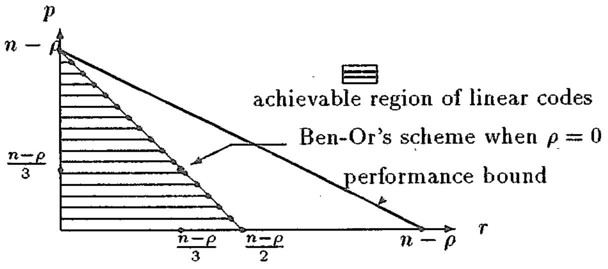

Fig. 2. Trade-off between privacy p and resiliency r for an (n, k) linear code when viewed as an (l, p, r, n) SSS, given ρ missing shares.

Adopting the concept of weak universal code from source coding [10], we define **a Weak Universal Channel Code** for a channel class Ω as a random code C such that for each channel in Ω, the average probability of decoding error for code C is arbitrary small with high probability. We start with the following notations.

- ω: An N-ary, symmetric, erasure, discrete, memoryless channel with error rate ϵ, and erasure rate δ as defined in section 2.
- Ω: A class of channels ω. For example, $\{\omega : \forall\, \epsilon, \delta, \gamma$ such that $0 \le \epsilon, \delta, \gamma \le 1$, and $\epsilon + \delta \le \gamma\}$ is a class Ω_γ for a fixed γ.
- $\mathcal{A}$: The set of input/output alphabet.
- X_m^n: The code word of length n chosen for secret message m.
- Y^n: The received sequence of length n at the output of channel ω.
- (e^{nR}, n): a codebook of size e^{nR} with each code word of length n.

Construct a random code C^* as described in [13, pp. 200] with a rate $R = (1-\gamma)\ln N$ in nats, and a uniform distribution $q(x) = u(x) = 1/N$. Thus, the probability of generating a code C^* is $Pr(C^*) = \prod_{m=1}^{e^{nR}} \mathrm{U}(X_m^n)$, where $\mathrm{U}(X_m^n) = (1/N)^n$ is a uniform distribution vector for code word X_m^n. Consider the following notation: [11, pp. 136–139]

- $\overline{P_{e,m}^n(\omega)}$: The average probability of decoding error over the ensemble of random block codes, given that message m is sent through a channel ω, when the maximum-likelihood decoding rule is used. Thus,

$$\begin{aligned}\overline{P_{e,m}^n(\omega)} &= \sum_{X_m^n}\sum_{Y^n} \mathrm{U}(X_m^n) P_\omega(Y^n|X_m^n) \\ &\quad \times P(\text{error}|m, X_m^n, Y^n).\end{aligned}$$

where $P(\text{error}|m, X_m^n, Y^n)$, which is either 0 or 1 valued, is the probability of decoding error conditioned on the selection of the particular sequence X_m^n as message m's code word and the reception of sequence Y^n, and $P_\omega(Y^n|X_m^n) = \prod_{i=1}^n P_\omega(y_i|x_{mi})$ is the transition probability of the received sequence Y^n given that the code word X_m^n is sent through the channel ω. The summations are for all possible input sequences X_m^n and output sequences Y^n respectively.

- $\overline{P_e^n(\omega)}$: The average probability of decoding error over the ensemble of random block codes, then over all the messages that are sent through a channel ω, when the maximum-likelihood decoding rule is used. Thus,

$$\overline{P_e^n(\omega)} = \sum_{m=1}^{e^{nR}} Pr(m)\overline{P_{e,m}^n(\omega)}.$$

- $P_e^n(\omega, C^*)$: The average probability of decoding error for code C^* over all messages that are sent through a channel ω.

Furthermore, let $f(\omega)$ be the density function of $\omega \in \Omega_\gamma$, and assume that all channels in Ω_γ are uniformly distributed. Thus

$$f(\omega) = \begin{cases} \dfrac{2}{\gamma^2}, & 0 \le \epsilon + \delta \le \gamma, 0 \le \epsilon, \delta, \gamma \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 4.1.1: The random coding exponent for channel $\omega \in \Omega_\gamma$ is

$$\begin{aligned} E_r^\omega(R) = \max_\rho - \ln \Big([N^{\frac{-\rho}{1+\rho}}(1-\epsilon-\delta)^{\frac{1}{1+\rho}} \\ +(\epsilon)^{\frac{1}{1+\rho}}]^{1+\rho} + \delta - \rho R, \text{ for sufficiently large } N.\end{aligned}$$

Sketch of Proof: From [11, pp. 138], the random coding exponent for channel ω is defined as

$$E_r^\omega(R) = \max_\rho\{-\ln \sum_{y\in\mathcal{A}\cup\{\mathcal{E}\}} [\sum_{x\in\mathcal{A}} u(x)P_\omega(y|x)^{\frac{1}{1+\rho}}]^{1+\rho} - \rho R\}.$$

Substituting $P_\omega(y|x)$ by eq (1) in [7] and we have

$$\begin{aligned} E_r^\omega(R) &= \max_\rho - \ln \Big(\textstyle\sum_{y\in\mathcal{A}}[\sum_{x\in\mathcal{A}} u(x)P_\omega(y|x)^{\frac{1}{1+\rho}}]^{1+\rho} \\ &\quad +[\textstyle\sum_{x\in\mathcal{A}} u(x)P_\omega(\mathcal{E}|x)^{\frac{1}{1+\rho}}]^{1+\rho}\Big) - \rho R \\ &= \max_\rho - \ln \Big([N^{\frac{-\rho}{1+\rho}}(1-\epsilon-\delta)^{\frac{1}{1+\rho}} + (\epsilon)^{\frac{1}{1+\rho}}]^{1+\rho} \\ &\quad +\delta\Big) - \rho R,\end{aligned}$$

for $0 \le \rho \le 1$ and N becomes sufficiently large. //

Lemma 4.1.2: The ensemble of the random codes that perform well for a given channel $\omega \in \Omega_\gamma$ also perform asymptotically well for all channels in the same class.

Proof: Gallager's Noisy Channel Coding Theorem [11, pp. 143] states that the random coding exponent of a discrete, memoryless channel is non-negative for all R with $0 \leq R < C$, where C is the channel capacity. Furthermore, from (??) we have that the capacity C of channel ω is $(1-\epsilon-\delta)\log N$ for sufficiently large N, and that such capacity is achieved by a uniform input distribution $u(x) = 1/N\ \forall x \in \mathcal{A}$. Since $\epsilon+\delta \leq \gamma\ \forall\omega \in \Omega_\gamma$, so we have $C \geq (1-\gamma)\log N$. Thus, $\forall\omega \in \Omega_\gamma$ and $0 \leq R < (1-\gamma)\log N$, the random coding exponent $E_r^\omega(R)$ in Lemma 4.1.1 is non-negative. Let $g_n(\omega) = \exp\{-nE_r^\omega(R)\}\ \forall n \in \mathcal{N}, \omega \in \Omega_\gamma$, and we have

$$\begin{cases} (i) & g_1(\omega) \geq g_2(\omega) \geq \ldots \geq g_n(\omega) \geq 0 \\ (ii) & \lim_{n\to\infty} g_n(\omega) = 0. \end{cases}$$

Applying the Monotone Convergence Theorem MCT [12, pp. 211], we get

$$\lim_{n\to\infty} E[g_n(\omega)] = E[\lim_{n\to\infty} g_n(\omega)] = 0.$$

From [11, pp. 138], we get

$$\begin{aligned} \overline{P_{e,m}^n(\omega)} &\leq e^{nR\rho}\{\sum_{y\in\mathcal{A}\cup\{\mathcal{E}\}}[\sum_{x\in\mathcal{A}} u(x)P_\omega(y|x)^{\frac{1}{1+\rho}}]^{1+\rho}\}^n \\ &= \exp\{-nE_r^\omega(R)\} = g_n(\omega). \end{aligned}$$

So for an arbitrary set of message probabilities $Pr(m)$, we have

$$\overline{P_e^n(\omega)} = \sum_{m=1}^{e^{nR}} Pr(m)\overline{P_{e,m}^n(\omega)} \leq \sum_{m=1}^{e^{nR}} Pr(m)g_n(\omega) = g_n(\omega).$$

Thus $0 \leq \lim_{n\to\infty}\overline{P_e^n(\omega)} \leq \lim_{n\to\infty} g_n(\omega) = 0$, $\forall\omega \in \Omega_\gamma$ which implies that $\lim_{n\to\infty}\overline{P_e^n(\omega)} = 0$, $\forall\omega \in \Omega_\gamma$. Thus

$$\int_{\Omega_\gamma}\lim_{n\to\infty}\overline{P_e^n(\omega)}f(\omega)d\omega = 0.$$

Applying the Dominated Convergence Theorem DCT [12, pp. 72], we get

$$\lim_{n\to\infty}\int_{\Omega_\gamma}\overline{P_e^n(\omega)}f(\omega)d\omega = 0, \qquad (5)$$

which implies that the ensemble of the random codes that perform well for a given channel $\omega \in \Omega_\gamma$ will also perform asymptotically well for all channels in the same class. //

Theorem 4.1: Viewing an (n,k) random code as an (l,p,r,n) SSS, we have for reconstructability l that $l \geq k+t$ with high probability for sufficiently large n and N, where N is the size of alphabet and t is the number of errors which have occurred.

Proof: Step 1. Show the existence of a weak universal channel code for class Ω_γ. In other words, we want to show that there exists a codebook $\mathcal{C}^*$ that $\lim_{n\to\infty}\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega = 0$.

From [13, pp. 200], $\overline{P_e^n(\omega)} = \sum_{\mathcal{C}^*} Pr(\mathcal{C}^*)P_e^n(\omega,\mathcal{C}^*)$ Substitute this into (5), we have

$$\begin{aligned} 0 &= \lim_{n\to\infty}\int_{\Omega_\gamma}\sum_{\mathcal{C}^*} Pr(\mathcal{C}^*)P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega \\ &= \lim_{n\to\infty}\sum_{\mathcal{C}^*} Pr(\mathcal{C}^*)\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega \\ &= \lim_{n\to\infty} E_{\mathcal{C}^*}\left[\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega\right]. \qquad (6) \end{aligned}$$

Since $P_e^n(\omega,\mathcal{C}^*) \geq 0$, thus $\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega \geq 0$. Also, (6) implies that $E_{\mathcal{C}^*}\left[\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega\right]$ can be arbitrary small for sufficiently large n. So if we let $E_{\mathcal{C}^*}[\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega] \leq (\epsilon/2)$ for any arbitrary small $\epsilon > 0$, and apply Markov's inequality to $\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega$, then we have [13, pp. 57]

$$P\left(\left[\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega\right] \geq \epsilon\right) \leq \frac{\epsilon/2}{\epsilon} = \frac{1}{2},$$

which implies

$$P\left(\left[\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega\right] \leq \epsilon\right) \geq \frac{1}{2},$$

for any arbitrary small $\epsilon > 0$. Thus, at least half of $\mathcal{C}^*$'s have $\lim_{n\to\infty}\int_{\Omega_\gamma} P_e^n(\omega,\mathcal{C}^*)f(\omega)d\omega = 0$.

Step 2. Relate the weak universal channel code $\mathcal{C}^*$ to an (l,p,r,n) SSS and show that, with high probability, $l \geq k+t$ for sufficiently large n and N, where t is the number of errors, and k is the rate of code $\mathcal{C}^*$.

In *Step 1*, we have constructed a code $\mathcal{C}^*$ with code rate $R = (1-\gamma)\ln N$. But the rate can also be expressed in terms of k as $R = \frac{k}{n}\ln N$ if $e^{nR} = N^k$. Hence, $\frac{k}{n}\ln N = (1-\gamma)\ln N$, for sufficiently large N. Thus, $k = n(1-\gamma) \leq n(1-\epsilon-\delta)\ \forall\omega \in \Omega_\gamma$. Let ρ and t be the number of erasures and errors in the received word Y^n, respectively. Then, we have $t = n\epsilon$ and $\rho = n\delta$ for sufficiently large n, based on the Chernoff bound [11, pp. 127]. Thus, $k \leq n-t-\rho$, $\forall(t,\rho)$, where $t+\rho \leq n\gamma$. This implies that, with high probability of correct decoding, the minimum number of correct elements in Y^n is greater than or equal to k for any combination of errors and erasures as long as their sum is less than or equal to $n\gamma$. Since we choose the secret m to be the index of such a code word, and we assume that the codebook is available, hence, for sufficiently large n and N, any $k+t$ or more shares suffice to recover the secret where t is the number of erroneous participating shares.//

4.2 Privacy of Random Codes

Theorem 4.2: An (n,k) random code does not have privacy when viewed as an (l,p,r,n) SSS.

Proof: Denote

- S: A random variable representing the secret which takes on a value from $\{1,\ldots,M\}$. In other words, S is the index of the code word in a codebook.
- X_j: A random variable representing the jth given share and takes on a value in the alphabet $\mathcal{A} = \{a_0, a_1, \ldots, a_{N-1}\}$ of size N.
- C^*: An (n,k) random code, or a codebook (M,n).
- C_j^*: The jth column of a codebook C^*.
- $H(S|X)$: The conditional entropy of S given a released share X.

Following Shannon's argument on the computation of the key equivocation of random cipher [14], we have

$$H(S|X_j = a_i \text{ and } C_j^* \text{ has } q\ a_i's) = \log q, \qquad (7)$$

because the code vectors in each row are equiprobable. Also because of the symmetry of the codes,

$$H(S|X_j = a_i) = H(S|X_1 = a_i). \qquad (8)$$

The elements are chosen independently, so

$$P(C_1^* \text{ has } q\ a_i's) = \binom{M}{q}\left(\frac{1}{N}\right)^q\left(1-\frac{1}{N}\right)^{M-q}, \qquad (9)$$

and

$$P(X_1 = a_i | C_1^* \text{ has } q\ a_i's) = \frac{q}{M}. \qquad (10)$$

Also

$$\begin{aligned} &P(X_1 = a_i \text{ and } C_1^* \text{ has } q\, a_i's) \\ &\quad = P(X_1 = a_i | C_1^* \text{ has } q\, a_i's) \times P(C_1^* \text{ has } q\, a_i's). \end{aligned} \qquad (11)$$

Since $a_i's$ are equiprobable in codebook C^* and there are N a_i's, so we have

$$\begin{aligned} H(S|X) &= \sum_{a_i}\sum_{q} H(S|X_1 = a_i \text{ and } C_1^* \text{ has } q\ a_i's) \\ &\qquad \times P(X_1 = a_i \text{ and } C_1^* \text{ has } q\ a_i's) \\ &= \frac{N}{M}\sum_{q=1}^{M} q\binom{M}{q}\left(\frac{1}{N}\right)^q\left(1-\frac{1}{N}\right)^{M-q}\log q. \end{aligned} \qquad (12)$$

Next, consider the case of giving two shares $X_j = a_i, X_m = a_l$. Since only the correct released shares are considered, the pair (a_i, a_l) must be the jmth element of some code words in C^*. Again, because of the symmetry of the codes and the elements are chosen independently, so we have $P(C_{12}^* \text{ has } q\ (a_i,a_l)'s) = \binom{M}{q}(\frac{1}{N^2})^q(1-\frac{1}{N^2})^{M-q}$, and $H(S|X_1 = a_i, X_2 = a_l \text{ and } C_{12}^* \text{ has } q\ (a_i,a_l)'s) = \log q$. Also, $P(X_1 = a_i, X_2 = a_l | C_{12}^* \text{ has } q\ (a_i,a_l)'s) = \frac{q}{M}$. Since (a_i,a_l)'s are equiprobable in codebook C^* and there are N^2 possible pairs, so

$$\begin{aligned} &H(S|X_1, X_2) \\ &\quad = \sum_{a_i}\sum_{a_l}\sum_{q} H(S|X_1 = a_i, X_2 = a_l \\ &\qquad \text{and } C_{12}^* \text{ has } q\ (a_i,a_l)'s) \\ &\qquad \times P(X_1 = a_i, X_2 = a_l \text{ and } C_{12}^* \text{ has } q\ (a_i,a_l)'s) \\ &\quad = \frac{N^2}{M}\sum_{q=1}^{M} q\binom{M}{q}\left(\frac{1}{N^2}\right)^q\left(1-\frac{1}{N^2}\right)^{M-q}\log q. \end{aligned}$$

In general, the conditional entropy of the secret, when h shares are given, is

$$\begin{aligned} &H(S|X_{i_1}, \ldots, X_{i_h}) \\ &\quad = \sum_{a_{j_1}}\cdots\sum_{a_{j_h}}\sum_{q} H(S|X_{i_1}, \ldots, X_{i_h} = (a_{j_1}, \ldots, a_{j_h}) \\ &\qquad \text{and } C_{i_1 \ldots i_h}^* \text{ has } q\ (a_{j_1} \ldots a_{j_h})'s) \\ &\qquad \times P(X_{i_1}, \ldots, X_{i_h} = (a_{j_1}, \ldots, a_{j_h}) \\ &\qquad \text{and } C_{i_1 \ldots i_h}^* \text{ has } q\ (a_{j_1} \ldots a_{j_h})'s) \\ &\quad = \frac{N^h}{M}\sum_{q=1}^{M} q\binom{M}{q}\left(\frac{1}{N^h}\right)^q\left(1-\frac{1}{N^h}\right)^{M-q}\log q. \end{aligned}$$

Asymptotically, when M gets sufficiently large, we can evaluate the conditional entropy as Shannon did for the random cipher [14]. The variation of $\log q$ over the range where $\binom{M}{q}$ assumes large values will be small, so $\log q$ can be replaced by $\log \bar{q}$, where $\bar{q}$ is the expectation value of q. Thus

$$\begin{aligned} &H(S|X_{i_1}, \ldots, X_{i_h}) \\ &\quad = \frac{N^h}{M}\log \bar{q}\sum_{q=1}^{M} q\binom{M}{q}\left(\frac{1}{N^h}\right)^q\left(1-\frac{1}{N^h}\right)^{M-q} \\ &\quad = \log M - h\log N, \end{aligned}$$

as $\bar{q} = M \cdot \frac{1}{N^h}$, based on Chernoff bound [11, pp. 128]. Thus, for a random (M,n) code, where $M = N^k$, for $1 \le h \le k-1$, we have

$$H(S|X_{i_1}, \ldots, X_{i_h}) = (k-h)\log N < k\log N = H(S).//$$

Remark: Note that the information rate of a random code is much greater than 1. Hence, the above result coincides with a result derived by Karnin *et al.* [15] that the information rate for a perfect (l,p,r,n)

SSS must be ≤ 1. Fig. 3 illustrates how the entropy drops when 0 to $k+5$ shares are given for alphabet size $N = 2, 4, 8, 16, 64$ and secret size $M = N^k$ with $k = 10$. Observe that $H(S|X_{i_1}, \dots, X_{i_h})$ starts at $\log M$ or $k \log N$ at $h = 0$, decreases linearly with a slope of $-\log N$ to the neighborhood of k, then tapers off with a smaller slope.

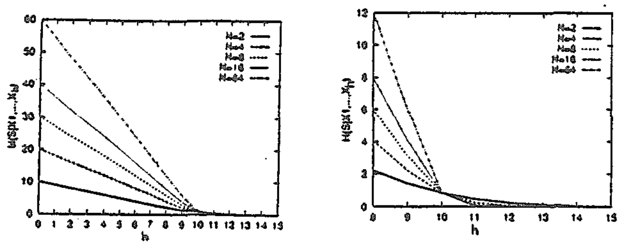

Fig. 3. The conditional entropy of secret given 0 to $k+5$ pieces when (n,k) random codes are used for $k = 10$, and various N.

5. Hahsed Random Codes as (l,p,r,n) SSSs

5.1 Hashed Random Code (HRC)

A hashed random (n,k) code (HRC) is a random code C as described in [13, pp. 199] combined with a hasher $\mathcal{G}$ which maps the index of a random code C from the space $\mathcal{S}_s = [0, \dots, N-1]^k$ onto the space $\mathcal{S}_{\bar{s}} = [0, \dots, N-1]$ such that the new index is uniformly distributed in space $[0, \dots, N-1]$. The hashed index is used as the secret of HRC, and is denoted by $\bar{S}$. Since the hasher $\mathcal{G}$ maps S from space $\mathcal{S}_s$ onto space $\mathcal{S}_{\bar{s}}$ uniformly, the random codebook $C = (M,n)$ is equally divided into N blocks. Each block of M/N code words is associated with a particular value $i = 0, 1, \dots, N-1$ that the secret $\bar{S}$ can take on. Because of the symmetry of the random codes, the hasher can be any mapping from $\mathcal{S}_s$ to $\mathcal{S}_{\bar{s}}$ as long as it equally divides the space into N subspaces.

5.2 Reconstructability and Privacy of HRC

Theorem 5.2.1: Viewing an HRC as an (l,p,r,n) SSS, we have for reconstructability l and privacy p that $l \geq k + t = p + t$, for sufficiently large N with high probability, where t is the number of errors which have occurred.

Proof: We have that $l \geq k + t$ with high probability as an immediate result from Theorem 4.1, since an HRC is a random code in which any $k + t$ shares can reconstruct the code word with high probability. Once the code word is reconstructed, the index S is known and the secret $\bar{S}$ can be revealed by applying the hasher $\mathcal{G}$ to S.

To show that $p = k$ for sufficiently large N, we'll show that the conditional entropy of the secret remains unchanged when $1 \leq h \leq k-1$ shares are revealed. Use the same notations which were described in Section 4.2 except replacing the key S by $\bar{S}$. Assume there are q a_i's in the first column of codebook C and that these q a_i's are distributed by $[q_0, q_1, \dots, q_{N-1}]$ where q_i represents the number of a_i's which fall in the block associated with $\bar{S} = i$, $0 \leq q_i \leq q$, and $\sum_{i=0}^{N-1} q_i = q$. Denoting $\bar{\mathrm{q}}$ as the distribution $[q_0, q_1, \dots, q_{N-1}]$, then we have

$$P(\bar{S} = i | X_1 = a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) = \frac{q_i}{q},$$

$$\begin{aligned} H(\bar{S}|X_1) &= a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &= \sum_{i=0}^{N-1} -P(\bar{S} = i | X_1 = a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &\quad \times \log \left(P(\bar{S} = i | X_1 = a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \right) \\ &\equiv h\left(\frac{q_0}{q}, \frac{q_1}{q}, \dots, \frac{q_{N-1}}{q} \right) \end{aligned}$$

Since

$$\begin{aligned} P(X_1 &= a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &= P(X_1 = a_1 | C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &\quad \times P(C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &= P(X_1 = a_1 | C_1 \text{ has } q\, a_1' s) \times P(\bar{\mathrm{q}} | C_1 \text{ has } q\, a_1' s) \\ &\quad \times P(C_1 \text{ has } q\, a_1' s) \end{aligned}$$

because $\bar{\mathrm{q}}$ will not affect the computation of $P(X_1 = a_1 | C_1 \text{ has } q a_1' s)$. Also,

$$P(\bar{\mathrm{q}} | C_1 \text{ has } q\, a_1' s) = \binom{q}{q_0 \cdots q_{N-1}} \prod_{i=0}^{N-1} (p_i)^{q_i}, \quad (13)$$

where p_i is the probability that symbol a_1 falls in block i, and hence is $1/N$ based on the symmetry of the codes. Combining (13) with (9) and (10) yields

$$\begin{aligned} P(X_1 &= a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &= \sum_{q=1}^{M} \frac{q}{M} \binom{M}{q} (\frac{1}{N})^q (1 - \frac{1}{N})^{M-q} \binom{q}{q_0 \cdots q_{N-1}} \\ &\quad \times \textstyle\prod_{i=0}^{N-1} (\frac{1}{N})^{q_i} \end{aligned} \quad (14)$$

Thus, the conditional entropy of secret $\bar{S}$ given X_1 is

$$\begin{aligned} H(\bar{S}&|X_1) \\ &= \sum_{a_i} \sum_{q} \sum_{\bar{\mathrm{q}}} H(\bar{S}|X_1 = a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &\quad \times P(X_1 = a_1, C_1 \text{ has } q\, a_1' s \text{ with } \bar{\mathrm{q}}) \\ &= \frac{N}{M} \sum_{q=1}^{M} q \binom{M}{q} \left(\frac{1}{N}\right)^q \left(1 - \frac{1}{N}\right)^{M-q} \\ &\quad \times \sum_{\bar{\mathrm{q}}} h\left(\frac{q_0}{q}, \dots, \frac{q_{N-1}}{q}\right) \binom{q}{q_0 \cdots q_{N-1}} \prod_{i=0}^{N-1} \left(\frac{1}{N}\right)^{q_i} \end{aligned} \quad (15)$$

Similarly, for two released shares a_1 and a_2, we have $P(\bar{S} = i | X_1 = a_1, X_2 = a_2,$
C_{12} has $q\,(a_1, a_2)'s$ with $\tilde{\mathbf{q}}) = \frac{q_i}{q}$,

$$
\begin{aligned}
&H(\bar{S}|X_1 = a_1, X_2 = a_2, C_{12} \text{ has } q\,(a_1,a_2)'s \text{ with } \tilde{\mathbf{q}}) \\
&\quad = \sum_{i=0}^{N-1} -P(\bar{S} = i | X_1 = a_1, a_2, \\
&\qquad C_{12} \text{ has } q\,(a_1,a_2)'s \text{ with } \tilde{\mathbf{q}}) \\
&\qquad \times \log\left(P(\bar{S} = i | X_1 = a_1, X_2 = a_2,\right. \\
&\qquad C_{12} \text{ has } q\,(a_1,a_2)'s \text{ with } \tilde{\mathbf{q}} \\
&\quad = h\left(\frac{q_0}{q}, \frac{q_1}{q}, \ldots, \frac{q_{N-1}}{q}\right)
\end{aligned}
$$

Since

$$
\begin{aligned}
&P(X_1 = a_1, X_2 = a_2, C_{12} \text{ has } q\,(a_1,a_2)'s \text{ with } \tilde{\mathbf{q}}) \\
&\quad = P(X_1 = a_1, X_2 = a_2 | C_{12} \text{ has } q\,(a_1,a_2)'s) \\
&\qquad \times P(\tilde{\mathbf{q}} | C_{12} \text{ has } q\,(a_1,a_2)'s) \\
&\qquad \times P(C_{12} \text{ has } q\,(a_1,a_2)'s)
\end{aligned}
$$

because $\tilde{\mathbf{q}}$ will not affect the computation of $P(X_1 = a_1, X_2 = a_2 | C_{12}$ has $q\,(a_1, a_2)'s)$. Thus, we have

$$
\begin{aligned}
&P(X_1 = a_1, X_2 = a_2, C_{12} \text{ has } q\,(a_1,a_2)'s \text{ with } \tilde{\mathbf{q}}) \\
&\quad = \sum_{q=1}^{M} \frac{q}{M} \binom{M}{q} (\frac{1}{N^2})^q (1 - \frac{1}{N^2})^{M-q} \binom{q}{q_0 \cdots q_{N-1}} \\
&\qquad \times \textstyle\prod_{i=0}^{N-1} (p_i)^{q_i} ,
\end{aligned}
\tag{16}
$$

where p_i is the probability that a pair of symbols (a_1, a_2) fall in the block i, and hence is $1/N$ due to the symmetry of the codes. So,

$$
\begin{aligned}
&H(\bar{S}|X_1, X_2) \\
&\quad = \textstyle\sum_{a_i} \sum_{a_l} \sum_{q} \sum_{\tilde{\mathbf{q}}} H(\bar{S} | X_1 = a_i, X_2 = a_l, \\
&\qquad C_{12} \text{ has } q\,(a_i,a_l)'s \text{ with } \tilde{\mathbf{q}}) \\
&\qquad \times P(X_1 = a_i, X_2 = a_l, C_{12} \text{ has } q\,(a_i,a_l)'s \text{ with } \tilde{\mathbf{q}}) \\
&\quad = \tfrac{N^2}{M} \textstyle\sum_{q=1}^{M} q \dbinom{M}{q} (\tfrac{1}{N^2})^q \left(1 - \tfrac{1}{N^2}\right)^{M-q} \\
&\qquad \times \textstyle\sum_{\tilde{\mathbf{q}}} h\left(\tfrac{q_0}{q}, \ldots, \tfrac{q_{N-1}}{q}\right) \dbinom{q}{q_0 \cdots q_{N-1}} \textstyle\prod_{i=0}^{N-1} (\tfrac{1}{N})^{q_i}
\end{aligned}
\tag{17}
$$

In general, for $1 \leq h \leq k - 1$, we have

$$
\begin{aligned}
&H(\bar{S}|X_{i_1}, \ldots, X_{i_h}) \\
&\quad = \sum_{a_{j_1}} \cdots \sum_{a_{j_h}} \sum_{q} \sum_{\tilde{\mathbf{q}}} \\
&\qquad H(\bar{S}|X_{i_1}, \ldots, X_{i_h} = (a_{j_1}, \ldots, a_{j_h})'s, \\
&\qquad C_{i_1 \ldots i_h} \text{ has } q\,(a_{j_1}, \ldots, a_{j_h})'s \text{ with } \tilde{\mathbf{q}}) \\
&\qquad \times P(X_{i_1}, \ldots, X_{i_h} = (a_{j_1}, \ldots, a_{j_h})'s, \\
&\qquad C_{i_1 \ldots i_h} \text{ has } q\,(a_{j_1}, \ldots, a_{j_h})'s \text{ with } \tilde{\mathbf{q}}) \\
&\quad = \frac{N^h}{M} \sum_{q=1}^{M} q \binom{M}{q} \left(\frac{1}{N^h}\right)^q \left(1 - \frac{1}{N^h}\right)^{M-q} \\
&\qquad \sum_{\tilde{\mathbf{q}}} h\left(\frac{q_0}{q}, \ldots, \frac{q_{N-1}}{q}\right) \binom{q}{q_0 \cdots q_{N-1}} \prod_{i=0}^{N-1} \left(\frac{1}{N}\right)^{q_i}
\end{aligned}
\tag{18}
$$

Asymptotically, when N gets sufficiently large and so does M and q, and we have that $q_0 = q_1 = \ldots = q_{N-1} = q/N$, . Thus $h\left(\frac{q_0}{q}, \ldots, \frac{q_{N-1}}{q}\right) = (\frac{1}{N}, \ldots, \frac{1}{N}) = \log N$. By the Multinomial Theorem, $\sum_{\tilde{\mathbf{q}}} \binom{q}{q_0 \cdots q_{N-1}} \prod_{i=0}^{N-1} (\frac{1}{N})^{q_i} = \left(\sum_{i=0}^{N-1} \frac{1}{N}\right)^q = 1.$ Thus, for $1 \leq h \leq k - 1$, (18) becomes

$$
\begin{aligned}
&H(\bar{S}|X_{i_1}, \ldots, X_{i_h}) \\
&\quad = \tfrac{N^h}{M} \left[\textstyle\sum_{q=1}^{M} q \dbinom{M}{q} (\tfrac{1}{N^h})^q \left(1 - \tfrac{1}{N^h}\right)^{M-q}\right] \log N \\
&\quad = \log N,
\end{aligned}
$$

and we obtain that

$$
H(\bar{S}|X_{i_1} \ldots X_{i_h}) = \begin{cases} \log N = H(\bar{S}), & 1 \leq h \leq k-1 \\ 0, & h = k. \end{cases}
\tag{19}
$$

Thus, we have $p = k$. //

Theorem 5.2.2: Viewing an HRC as an (l, p, r, n) SSS, its privacy p and resiliency r are related by $p + r = n - \rho$, on an average for sufficiently large n and N, where N is the size of the alphabet.

Proof: The rate of a random code is $(k/n) \log N$ since we let $M = N^k$, and it can achieve the capacity $C = (1 - \epsilon - \delta) \log N$ on an average for sufficiently large N with arbitrary small decoding error [11, 13, 16]. Since an HRC is a random code with key randomization, $\frac{k}{n} \log N = (1 - \epsilon - \delta) \log N$, and so we have $k_c = n - n\epsilon - n\delta$, where k_c is the capacity rate. Hence, $k \leq n - t - \rho$, and so $t \leq n - \rho - k$ for sufficiently large n, based on Chernoff bound [11, pp. 122] Thus, we have $r = n - \rho - k$. From Theorem 5.2.1 we have that $p = k$ for sufficiently large N. Thus $p + r = n - \rho$, for sufficiently large n and N. //

5.3 Computer Evaluation of HRC

Theorem 5.2.1 and 5.2.2 addresses the asymptotic case of an HRC as an (l, p, r, n) SSS. However, making N infinity is unrealistic in practice. Next, We give some computer evaluations of the conditional entropy of key $\bar{S}$ using (18). The solid line in Fig. 4 is for $N = 2$ and $k = 10$. The dash line represents the asymptotic case and the point line is the conditional entropy of random variable S which is computed using (12). Observe that, the conditional entropy of $\bar{S}$ stays almost at $\log N$ for 0 to $k - 4$ released shares, decreases exponentially for the next 3 released shares to $0.6 \log N$ at the $k -$ 1th released share, then tapers to 0 at the $k +$ 3th released share. Since $N = 2$ is the worst possible case, such a result indicates that the asymptotic analysis will kick in relatively quickly. Fig. 5 shows a comparison between the computer evaluations of $H(\bar{S}|X_1, \ldots, X_{k+5})$ for $N = 2, k = 5$ and $N = 4, k = 5$. The solid line represents the case of $N = 4$ and the point line represents the case of $N = 2$

respectivey. Notice that the conditional entropy drops much sharply at $k-1$ and tapers to 0 quicker for $N=4$ than $N=2$. This suggests that, an HRC has reconstructability $l=k+1$, and privacy $p=k-2$, for reasonably large N.

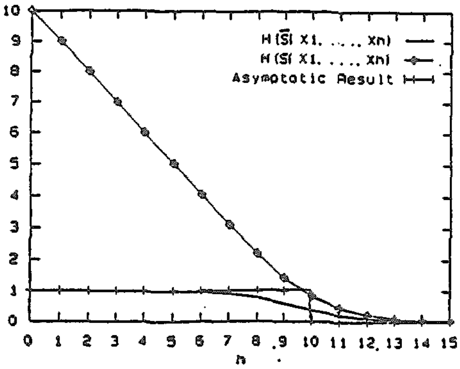

Fig. 4. The conditional entropy of the key $\bar{S}$ and S *given 0 to* $k+5$ *shares when an HRC and random* codes are used for $N=2$, $k=10$.

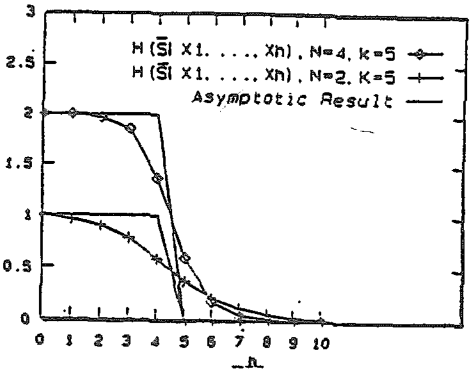

Fig. 5. The comparison between the conditional entropy of the key $\bar{S}$ given 0 to $k+5$ shares when an (n,k) HRC are used for $N=4, k=5$ and $N=2$, $k=5$.

6. Conclusions

In this paper, a general (n,k) code is translated into an (l,p,r,n) SSS. The relationship among reconstructability and resiliency of an (l,p,r,n) SSS and the code rate of both linear and random block codes are established. Privacy is also established. We show that linear codes, when viewed as (l,p,r,n) SSSs, can not achieve the capacity bounds; We also show that *an* (n,k) *random code, when viewed as an* (l,p,r,n) SSS, do not have privacy even though it can actually achieve the capacity of reconstructability and resiliency with high probability for sufficiently large n and alphabet. Lastly, we present a hashed random code (HRC) and show that such a code can asymptotically achieve the performance bounds with high probability.

References

[1] A. Shamir, "*How to Share a Secret,*" *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[2] G. R. Blakley, "*Safeguarding Cryptographic Keys,*" in *Proceedings of the AFIPS 1979 National Computer Conference*, pp. 313–317, 1979.

[3] R. McEliece and D. Sarwate, "*On Sharing Secrets and Reed-Solomon Codes,*" *Communications of the ACM*, vol. 24, no. 9, pp. 583–584, 1981.

[4] W. Q. K. Kurosawa, S. Obana, "*t-Cheater Identifiable (k,n) Threshold Secret Sharing Schemes,*" in *Advances in Cryptology-CRYPTO'95*, pp. 410–423, Springer-Verlag, 1995.

[5] M. Ben-Or, S. Goldwasser, and A. Widgerson, "*Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation,*" in *Proceedings of the* 20^{th} *STOC, ACM*. pp. 1–10, 1988.

[6] T. Rabin and M. Ben-Or, "*Verifiable Secret Sharing and Multiparty Protocols with Honest Majority,*" in *Proceedings of the* 21^{th} *STOC, ACM*, pp. 73–85, 1989.

[7] P. L. Lin and J. G. Dunham, "A Secret Sharing Model: *GS^3*," *IEE Electronics Letters*, vol. 30, no. 25, pp. 2116–2118, 1994.

[8] R. E. Blahut, *Algebraic Methods for Signal Processing and Communication Coding*. New York: Springer-Verlag, 1992.

[9] S. Lin and J. D.J. Costello, *Error Control Coding: Fundamentals and Applications*. : Prentice-Hall, 1983.

[10] J. C. Kieffer, "*A Survey of the Theory of Source Coding with a Fidelity Criterion,*" *IEEE transactions on Information Theory*, vol. 39, no. 5, pp. 1473–1490, 1993.

[11] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[12] P. Billingsley, *Probability and Measure*. : Weiley-Interscienceems, 1986.

[13] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[14] C. Shannon, "*Communication theory of secrecy systems,*" in *Bell Syst. Tech. J.*, pp. 565–715, Oct. 1949.

[15] E. Karnin, J. Greene, and M. Hellman, "*On Secret Sharing Systems,*" *IEEE transactions on Information Theory*, vol. 29, no. 1, pp. 35–41, 1983.

[16] M. Mansuripur, *Introduction To Information Theory*. : Prentice-Hall, 1987.