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

A Unified Model for Collusion-Permissible and Nonpermissible Secret Sharing Systems*

Phen-Lan Lin

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

Abstract

A unified model US^2 which contains a pair of encoder-decoder and a channel is presented to emulate the secret sharing systems of various schemes. Depending on collusion is permitted or not in the particular system, the encoder-decoder pair and channel vary. The channel for collusion- nonpermissible US^2 is developed. Its capacity is derived and used to establish the bounds on the (l,p,r,n) secret sharing scheme. The channel and a possible coding scheme is investigated as well for US^2 which allows players collude.

Secret sharing scheme (t,n) threshold scheme (l,p,r,n) secret sharing scheme non-systematic linear code Reed-Solomon code t-private t-resilient collusion-permissible secret sharing model collusion-nonpermissible secret sharing model

1. Introduction

In cryptographic and large distributed systems, when a group of people share a common key, it is highly desirable to have robust key management such that a maximum level of secrecy (privacy) can be achieved while allowing tolerance of faults (resiliency) and non-participation (reconstructability). In reality, there is a trade-off among these essential requirements. Several schemes have been devised to achieve certain level of these requirements such as noncheating (k,n) threshold scheme by Shamir and Blakley which achieves $n/2$ privacy and $n/2$ reconstructability [14, 7], t-cheater identifiable (k,n) threshold scheme which has reconstructability of $k+2t$ by McEliece and Sarwate [17] or $k \geq 3t+1$ by K. Kurosawa et al. [5], and (t-private, t-resilient) scheme which can achieve $(n/3, n/3)$ by Ben-Or et al.'s fault- tolerant protocol [2] or $(n/2, n/2)$ with exponentially small probability of error by Rabin's information checking protocol [13]. However, a specific bound to inform the designer about what levels can be achieved in reconstrutability, privacy, and resiliency is lacking.

As Shannon pointed out there is a unified model for communication systems such as radio, television, satellite, etc. even though their physical aspects vary [10, pp. 144]. In this paper, we attempt to analogize the secret sharing systems to communication systems and present a unified model US^2 which contains a pair of encoder-decoder and a channel to emulate the secret sharing systems of various schemes. Depending on collusion is permitted or not in the particular system, the encoder-decoder pair and channel vary. The channel for collusion- nonpermissible US^2 is developed. Its capacity is derived and used to establish the bounds on the (l,p,r,n) secret sharing scheme presented by Lin and Dunham [3]. The channel and a possible coding scheme are investigated as well for US^2 which allows players collude.

2. Preliminary

- A *Secret sharing scheme* is a method of sharing a piece of secret information among a finite set of players in such a way that only certain specified subsets of players can recompute the secret information [8].
- A (k,n) *threshold scheme*, which was devised independently by Shamir [14] and Blakley [7] in 1979, divides the secret S into n pieces of information called *shares* or *shadows* $(s_1, s_2, \ldots, s_n)$ in such a way that the following two properties hold:
  (1), knowledge of any k or more shares makes S easily computable; and
  (2), knowledge of any $k-1$ or fewer shares leaves S completely undetermined in a sense that all possible values of S are equally likely.
- A *t-private, t-resilient scheme*, which was devised by Ben-Or, Goldwasser, and Wigderson [2] in their distributed fault-tolerant protocol, divides the secret S into n shares such that
  (1), any $t-1$ or fewer shares can't compute S better than a random guess; and
  (2), no set of t or fewer incorrect shares can affect the reconstruction of S.

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

Lin and Dunham combined the (t, n) threshold scheme and *(t-private, t-resilient)* scheme and presented an (l, p, r, n) secret sharing scheme [3].

- An (l, p, r, n) secret sharing scheme, which divides the secret S into n shares in such a way that the following properties hold:
  (1), knowledge of any l or more shares make S easily computable, and l is called *reconstructability*;
  (2), 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
  (3), no set of r or fewer incorrect shares can affect the correctness of S, and r is called *resiliency*.

Notice that knowledge of shares implies that the values of shares and their identifying indices are available. In the following sections, we establish a unified model to emulate a group of players, who may not be totally cooperative, sharing a discrete piece of secret.

3. The unified model: US^2

US^2 is a unified Secret Sharing Model as depicted in Fig. 1 which contains a source encoder and a channel encoder; a discrete, symmetric, erasure, noisy channel (memoryless or with memory); a channel docoder and a source decoder. Depending on whether players will collude or not, that is, whether the players will pool their shares before revealing their values respectively, the channel could be either with memory or memoryless.

Roughly speaking, a secret sharing scheme can be separated into two phases: secret sharing phase and secret revealing phase. Followings are the detailed description of how US^2 models both phases properly.

3.1 US^2 models secret sharing

Only the source and channel encoders are required to model the secret sharing phase. At first, the source encoder encodes the secret information to an information word using a proper source coding technique, then the channel encoder uses an efficient channel coding method to encode the information word to a code word with code rate less than or equal to the channel capacity. These steps emulate the honest dealer, who hides a secret in some finite pieces of information by using an arithmetic or logical scheme and distributes them to a finite set of players.

3.2 US^2 models secret revealing

The channel, the channel decoder, and the source decoder are involved in the secret revealing phase.

The channel is a discrete, symmetric, erasure, noisy channel having equal probabilities of transmitting erroneous digits, or having some probability of missing the digits. These characteristics emulate each player in the finite set, who holds a piece of information and is equally likely to not participate or make mistakes intentionally or unintentionally. Depending on collusion is allowed or not in the particular secret sharing scheme, the channel can be either with memory or memoryless. When collusion is not allowed, each player acts independently just like the output of a memoryless channel only depends on its current input. While collusion is permitted, the players may pool their shares before revealing their values just like the output of a channel with memory depends not only on the current input but also on the past inputs and outputs.

Based on the schemes that the source encoder and the channel encoder use, the proper decoding algorithms for both decoders can be chosen accordingly. After the entire code word has been transmitted, the decoders, which act as any group of legitimate players, can reconstruct the original information with arbitrary small error probability, according to Shannon's fundamental theorem of information theory [16]. Thus, the channel and decoders act together to emulate secret revealing.

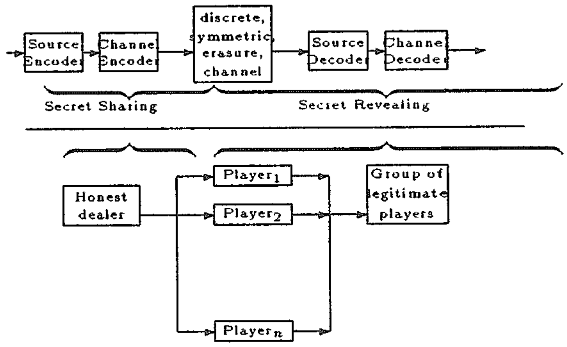

Fig. 1. A Unified Model for Secret Sharing System (US^2).

4. Collusion-nonpermissible US^2

In collusion-nonpermissible secret sharing systems, all players act independently. That is, each player (honest or dishonest) reveals the value based on his/her share solely just like the output of a memoryless channel only depends on its current input. The shares can be any number in a large finite Galois field $GF(N)$ just as some channel block codes can be generated from Galois fields. Thus, the channel in collusion-nonpermissible US^2 is an N-ary, discret, symmetric, erasure, memoryless channel (NSEC).

4.1 N-ary, discret, symmetric, erasure, memoryless channel (NSEC)

An N-ary, discret, symmetric, erasure, memoryless channel (NSEC) is shown in Fig. 2. The set $\mathcal{A} = \{a_0, a_1, \ldots, a_{N-1}\}$ is its N-symbol input alphabet, and $\mathcal{A} \cup \{\mathcal{E}\}$ is its output alphabet, where $\mathcal{E}$ is the erasure symbol. Each input symbol x_i and output symbol y_j has input/output transition probability of

$$p(y_j|x_i) = \begin{cases} 1-\epsilon-\delta & \text{for } y_j = x_i, \\ \dfrac{\epsilon}{N-1} & \text{for } y_j \neq x_i \text{ and } y_j \neq \mathcal{E}, \\ \delta & \text{for } y_j = \mathcal{E}, \end{cases} \tag{1}$$

where $i = 0, \ldots, N-1$ and $j = 0, \ldots, N$; and ϵ and δ are the probabilities of error and erasure, respectively, with $\epsilon + \delta \leq 1$. A sequence of symbols $x_0 x_1 \ldots x_{n-1}$ has input/output transition probability of

$$p(y_0 y_1 \ldots y_{n-1} \mid x_0 x_1 \ldots x_{n-1}) = \prod_{i=0}^{n-1} p(y_i \mid x_i). \tag{2}$$

4.2 The capacity of NSEC

Theorem 1: The capacity of an N-ary, symmetric, erasure, discrete, and memoryless channel with error rate ϵ and erasure rate δ is $C = (1-\epsilon-\delta)\log N + h_b(\delta) - h(\epsilon, \delta, 1-\epsilon-\delta) + \epsilon \log(\frac{N}{N-1})$, where $h_b(\cdot)$ is the binary entropy function. and $h(x, y, z)$ is the entropy function of x, y, and z. For the asymptotic case, $C = (1-\epsilon-\delta)\log N$.

Proof: Theorem 4.5.1 in [11, pp. 91] states that the necessary and sufficient conditions on input probabilities $\{Q(x)\}$ to achieve capacity on a discrete memoryless channel with transition probability $P(y_j|k)$ is that for some number C,

$$\begin{cases} I(x = k; Y) &= C \quad \text{for all } k \in \mathcal{A} \text{ with } Q(k) > 0 \\ I(x = k; Y) &\leq C \quad \text{for all } k \in \mathcal{A} \text{ with } Q(k) = 0 \end{cases}$$

where $\mathcal{A}$ is the input alphabet, $I(x = k, Y)$ is the mutual information for input k averaged over the outputs, that is

$$I(x = k; Y) = \sum_j p(y_j|k) \log \frac{P(y_j|k)}{\sum_i Q(i) P(y_j|i)},$$

and C is the capacity of the channel. Let

$$Q(x = i) = \frac{1}{N} > 0 \text{ for } i = a_0, a_1 \ldots, a_{N-1}.$$

Then,.for any input $x = k \in \mathcal{A}$, we have

$$\begin{aligned} & I(x = k; Y) \\ &= \sum_{j=0}^{N-1} P(y_j|k) \log \frac{P(y_j|k)}{\sum_i Q(i) P(y_j|i)} \\ &\quad + P(\mathcal{E}|k) \log \frac{P(\mathcal{E}|k)}{\sum_i Q(i) P(\mathcal{E}|i)} \\ &= \sum_{j=0, y_j \neq k}^{N-1} P(y_j|k) \log \frac{P(y_j|k)}{\frac{1}{N}\sum_i P(y_j|i)} \\ &\quad + P(k|k) \log \frac{P(k|k)}{\frac{1}{N}\sum_i P(k|i)} \\ &\quad + P(\mathcal{E}|k) \log \frac{P(\mathcal{E}|k)}{\frac{1}{N}\sum_i P(\mathcal{E}|i)} \\ &= (N-1)\frac{\epsilon}{N-1} \log \frac{\frac{\epsilon}{N-1}}{\frac{1-\delta}{N}} \\ &\quad + (1-\epsilon-\delta) \log \frac{1-\epsilon-\delta}{\frac{1}{N}(1-\delta)} + \delta \log(\frac{\delta}{\frac{1}{N} N \delta}) \\ &= (1-\delta-\epsilon) \log N + h_b(\delta) - h(\epsilon, \delta, 1-\epsilon-\delta) \\ &\quad + \epsilon \log \frac{N}{N-1}. \end{aligned} \tag{3}$$

Since a uniform distribution was assumed for input x,

$$Q(x = k) > 0 \text{ for all } x = k \in \mathcal{A},$$

and hence $I(x = k; Y)$ in (3) is the capacity. Thus

$$C = (1-\delta-\epsilon)\log N + h_b(\delta) - h(\epsilon, \delta, 1-\epsilon-\delta) + \epsilon \log \frac{N}{N-1}. \tag{4}$$

For sufficiently large N, $\log \frac{N}{N-1} \simeq \log 1 = 0$ and $0 \leq h_b(\cdot), h(\cdot) \leq \log 3$, and hence can be neglected. Thus, (4) becomes

$$\lim_{N \to \infty} C = (1-\epsilon-\delta) \log N.// \tag{5}$$

Fig. 3 shows the normalized channel capacity for size of alphabet N =2,3,8 and 65536 respectively, given error rate ϵ and erasure rate δ. Notice that the capacity approaches to $1-\epsilon-\delta$ as $N = 65536$ and is constant when both ϵ and δ are kept fixed.

After succesfully emulated the (l, p, r, n) secret sharing scheme by collusion-nonpermissible US^2, we are ready to establish its performance bound.

4.3 The performance bound of (l, p, r, n) secret sharing scheme

Theorem 2: Viewing an (n, k) code as an (l, p, r, n) secret sharing scheme, we have that $l \geq k + t \geq p + t$, and $p + r \leq n - \rho$ for sufficiently large n and N, where t is the number of errors and ρ is the number of missing pieces and N is the size of alphabet.

Proof: Apply Shannon's fundamental theorem to our model, we have that for sufficiently large n and N, the secret can be reconstructed with arbitrary small probability of error if a proper (n, k) code is used and the rate R is less than the channel capacity C. Thus, for sufficiently large n and N, we have

$$R \leq (1 - \epsilon - \delta) \log N. \qquad (6)$$

Since $N^k = 2^{nR}$, so $R = (k/n) \log N$. Substituting R in (6), we have

$$k \leq n - n\epsilon - n\delta. \qquad (7)$$

But for sufficiently large n, $n\epsilon$ is the number of erroneous shares, and $n\delta$ is the number of missing shares based on Chernoff bound [11, pp. 127]. Thus, let $t = n\epsilon$ and $\rho = n\delta$, we have $k \leq n - t - \rho$, and so,

$$n - \rho \geq k + t, \qquad (8)$$

which implies that the minimum number of shares required for recovery of the secret with arbitrary small probability of error is $k+t$. Thus, for sufficiently large n and N, we have

$$l \geq k + t. \qquad (9)$$

McEliece and Sarwate [17] had related an non-systematic linear (n, k) code to Shamir's (k, n) threshold scheme, thus there exists an non-systematic linear (n, k) code which has $p = k$. But linear code can be systematic, which means that the secret S could possibly be one of the share. Thus we have $p \leq k$. Combine this with (9), we have $l \geq k + t \geq p + t$.

Also, from (8), $t \leq n - \rho - k$, which implies that the maximum number of erroneous pieces that can not affect the correct reconstruction of the secret is $n - \rho - k$. Thus, for sufficiently large n and N, we have $r \leq n - \rho - k$. Since $p \leq k$, so $r \leq n - \rho - p$ and therefore, $p + r \leq n - \rho$. //

The bold lines in Fig. 4 depict the bound of the trade-off between (reconstructability l, privacy p), and between (privacy p, resiliency r) of the (l, p, r, n) secret sharing scheme.

5. Examples

5.1 Non-systematic linear-coding (l, p, r, n) secret sharing scheme

An non-systematic linear-coding (l,p,r,n) secret sharing scheme is an (l,p,r,n) SSS realized by an non-systematic linear code. From [12], an non-systematic linear (n, k) code word, when expressed explicitly, is $s_i = f(\alpha_i)$ for $i = 1, \ldots, n$, where $f(x) = a_0 + a_1 x + \ldots + a_{k-1}x^{k-1}, a_0, a_1, \ldots, a_{k-1}$ are randomly chosen from a uniform distribution of integers in a large finite field $GF(N)$ and α_i's are non-zero, distinct elemets in $GF(N)$. Let the secret S be a_0, the information word be $(S, a_1, \ldots, a_{k-1})$ and the channel code word be $(s_1, \ldots, s_n)$, such non-systematic linear codes can emulate the honest dealer sharing the secret S with n players.

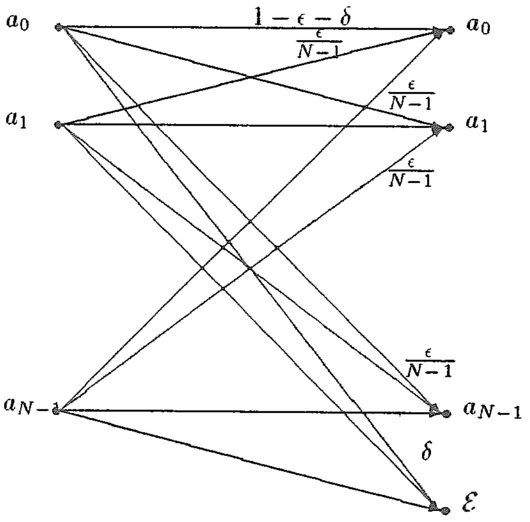

Fig. 2. An N-ary, discrete, symmetric, erasure, memoryless channel.

Since the players in this scheme are possibly cheating and nonparticipating, they can be emulated by the channel in collusion-nonpermissible US^2.

McEliece and Sarwate had related the above non-systematic linear (n, k) code to Shamir's (k, n) threshold scheme with t-cheater extension [17]. They pointed out that $k+2t$ shares suffice to reconstruct the secret based on the theory of error correcting codes and hence we have reconstructability $l \geq k + 2t$ and privacy $p = k$. Since $l \geq k + 2t$ implies $n - \rho \geq k + 2t$, hence $t \leq \frac{1}{2}(n - \rho - k)$ and so the resiliency $r = \frac{1}{2}(n - \rho - k)$. The shaded areas in Fig. 4 depict this result.

5.2 Shamir's (k, n) threshold scheme

Shamir's scheme is a special case of non-systematic linear (n, k) coding (l, p, r, n) scheme. In his realization, the share $s_i = f(i)$ for $i = 1, \ldots, n$. Since all players are assumed honest, the channel is an N-ary discrete, memoryless, erasure channel with error rate $\epsilon = 0$. Thus the capacity $C = (1 - \delta) \log N$ asymptotically. Similar to the proof in Theorem 2, we have $k \leq n - \rho$, so $l \geq k$, and $l \geq p$. Since (k, n) threshold scheme requires $l = p$, and hence it is an (l, p, r, n) scheme with $l = p = n/2$. The point marked by an 'X' as shown in Fig. 4 depicts this result when $t = 0$.

5.3 Ben-Or et al's (t-private, t-resilient) scheme

Ben-Or et al.'s (t-private, t-resilient) scheme is a primitive (n, k) Reed-Solomon code [2], which is also a special case of non-systematic linear (n, k) coding (l, p, r, n) scheme. Since a full enrollment is as-

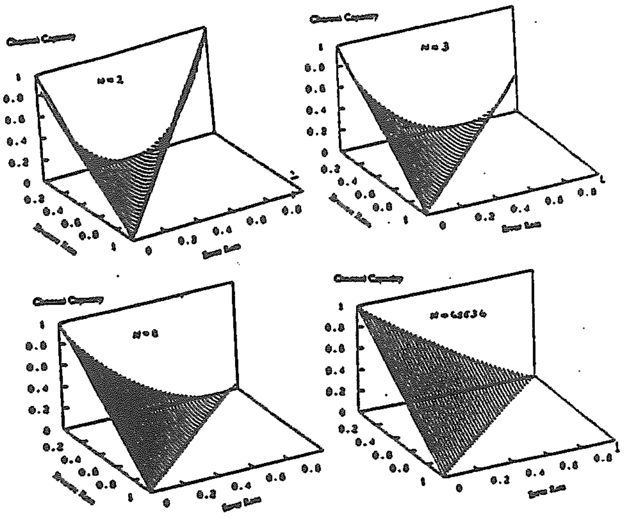

Fig. 3. The capacity of an N-ary, discrete, symmetric, erasure, memoryless channel for N =2,3,8 and 65536.

sumed, the channel is an N-ary, discrete, memoryless, symmetric channel with $\delta = 0$. Thus the capacity $C = (1-\epsilon)\log N$ asymptotically. Similar to the proof in Theorem 2, we have capacity rate $k_c \leq n - t$. But an (n, k) Reed-Solomon code can correct $\frac{1}{2}(n-k)$ errors [6], that is, $t \leq \frac{1}{2}(n-k)$. So the code rate $k \leq n - 2t < k_c$. But Reed-Solomon codes also have $p = k$ based on Shamir's argument, hence $r \leq \frac{1}{2}(n-p)$. Since (t-private, t-resilient) scheme requires $p = r$, and hence it is an (l, p, r, n) scheme with $p = r = n/3$. The point marked by a '•' as shown in Fig. 4 depicts this result when $\rho = 0$.

5.4 Rabin et al.'s information checking protocol

Unlike the schemes of non-systematic linear coding and Ben-Or et al. which use error-correcting codes to detect and correct the erroneous shares, Rabin and Ben-Or [13] used an information checking protocol to validate the shares and hence to detect and identify the cheaters.

Since the model US^2 enables us to realize the secret sharing schemes by error-correcting and detecting codes, the information checking protocol are not fitted in US^2 properly.

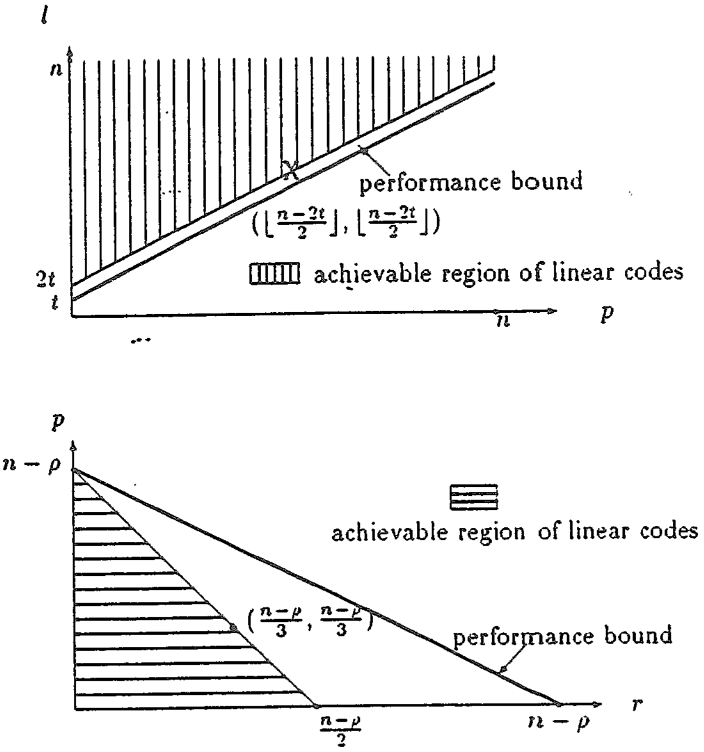

Fig. 4. The performance capacity for a collusion-nonpermissible (l, p, r, n) secret sharing scheme.

6. Collusion-permissible US^2

Unlike the model in which each player casts his/her value according to his/her own share only, this model deals with players who may not act independently. In other words, the players may pool their shares before casting their values. Depending on how the players collude, the channel that emulates the players varies accordingly. For example, if $player_i$ casts $value_i$ according to $share_i$ and $value_{i-1}$ only, then we can emulate the players by a channel with Markov sources. If k players collude, i.e., $player_k$ casts his/her value based on values of $player_1$ to $player_{k-1}$ and shares of $player_1$ to $player_k$, then we can emulate the players by a channel with the input/output probability $p(y_k|x^k, y^{k-1})$, where x^k denotes the shares of $player_1$ to $player_k$ and y^{k-1} denotes the values casted by $player_1$ to $player_{k-1}$. Since the present state of a channel will in a sense represent a summary of its past history, the channel with finite memory is in fact a finite-state channel.

6.1 A discrete, symmetric, erasure, finite-state channel

A discrete, symmetric, erasure, finite-state channel with a finite set $S = \{s_1, \ldots, s_r\}$ as its set of states and two finite set $\Gamma = \{b_1, \ldots, b_t\}$ and Δ as its input alphabet and output alphabet can be characterized by a collection of r by r transition matrices $M_{b_1}, \ldots, M_{b_t}$, where $M_i(s_j, s_k)$ represents the probability of moving from state s_j to s_k when the input symbol is i, together with a function g that assigns to each pair (b, s), $b \in \Gamma, s \in S$, an element $g(b, s) \in \Delta$ [1, pp. 215].

6.2 Codes for a discrete, symmetric, erasure, finite-state channel

From [9, pp. 212], block codes, such as linear (n, k) codes used in collusion-nonpermissible model as described in the sections above, are designed for channels that there is no dependence on past information bits, convolutional codes are for channels that the ouput block depends not only on the current input block, but also on some of the past inputs as well. This may explains the fact that no good block codes have been proposed yet for secret sharing systems that allow players collude. Tompa and Woll [4] pointed out cheating is possibly undetected if k players collude when using Shamir's scheme even choosing the secret from a field much smaller than the field where the shares reside.

7. Conclusions

A unified model US^2 is presented to model the various secret sharing schemes in which players may or may not collude. Depending on whether collusion is allowed or not, the channel in US^2 is with or without memory. The channel capacity of the collusion-nonpermissible US^2 is developed and used to establish the performance bound for an (l, p, r, n) secret sharing scheme. An non-systematic linear (n, k) code is demonstrated to be an (l, p, r, n) secret sharing scheme but yet not reaching the performance capacity. This suggests possible existence of other codes which might outperform the linear codes when viewed as secret sharing schemes. The channel in collusion-permissible US^2 and possible codes for it are also investigated.

References

[1] Robert Ash. *Information Theory.* Interscience Publishers, New York, 1965.

[2] 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*, pages pp. 1-10, 1988.

[3] P. L. Lin, J. G. Dunhum. *A General Secret Sharing Model-*GS^3. In *Electronics Letters*, pages pp. 2116-2118, 1994.

[4] M. Tompa, H. Woll. *How to Share a Secret with Cheaters. Journal of Cryptology*, vol. 1(2):pp. 133–137, 1994.

[5] K. Kurosawa, S. Obana, W. Ogata. *t-Cheater Identifiable (k,n) Threshold Secret Sharing Schemes. Advances in Cryptology- Crypto'95*, pp. 410–423, 1995.

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

[7] G. R. Blakley. *Safeguarding Cryptographic Keys.* In *Proceedings of the AFIPS 1979 National Computer Conference*, pages pp. 313-317, 1979.

[8] E. F. Brickell and D. R. Stinson. *Some Improved Bounds on the Information Rate of Perfect Secret sharing Schemes. Journal of Cryptology*, vol. 5(3):pp. 153–166, 1992.

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

[10] S. Ihara. *Information Theory for Continuous Systems.* World Scientific, 1993.

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

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

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

[14] A. Shamir. *How to Share a Secret. Communications of the ACM*, vol. 22(11):pp. 612–613, 1979.

[15] C. Shannon. *Communication theory of secrecy systems.* In *Bell Syst. Tech. J.*, pages pp. 565–715, Oct. 1949.

[16] C. Shannon. *A mathematical theory of communication.* In *Bell Syst. Tech. J., 27*, pages pp. 379–423, Oct. 1949.

[17] R.J. McEliece and D.V. Sarwate. *On Sharing Secrets and Reed-Solomon Codes. Communications of the ACM*, vol. 24(9):pp. 583–584, 1981.