Size of Shares and Probability of Cheating in Threshold Schemes*

Marco Carpentieri Alfredo De Santis Ugo Vaccaro

Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (SA), Italy

Abstract

In this paper we study the amount of secret information that must be given to participants in any secret sharing scheme that is secure against coalitions of dishonest participants in the model of Tompa and Woll [20]. We show that any (k, n) threshold secret sharing algorithm in which any coalition of less than k participants has probability of successful cheating less than some $\epsilon > 0$ it must give to each participant shares whose sizes are at least the size of the secret plus $\log \frac{1}{\epsilon}$.

1 Introduction

In 1979 Blakley [2] and Shamir [15] gave protocols to solve the following problem: divide a secret s in n shares $d_1, \ldots, d_n$ in such a way that:

i) the knowledge of k or more d_i's makes s computable,

ii) the knowledge of $k-1$ or less d_i's leaves s *completely* indeterminate.

This problem, known in the literature as "(k, n) Threshold Secret Sharing", has received considerable attention in the last few years because of its many applications to several fields, as data security, secure computation and others [10]. For an extensive bibliography and illustration of the main results in the area we refer the reader to [17] and [18].

Let $\mathcal{P} = \{P_1, \ldots, P_n\}$ be a set of participants, S be the set of secrets and $D_1, \ldots, D_n$ be the sets in which the shares to participants $P_1, \ldots, P_n$ are taken. Any probability distribution $\{p(s)\}_{s \in S}$ on the set of secrets S and a sharing algorithm for secrets in S (both known by each participant) naturally induce a probability distribution $\{p(d_1, \ldots, d_n)\}_{d_1 \in D_1, \ldots, d_n \in D_n}$ on the joint space $D_1 \times \ldots \times D_n$ of the possible values of the shares. Therefore, we will consider each D_i as a random variable. Formally, a (k, n) Threshold Secret Sharing Scheme is a method to distribute shares to the n participants such that:

*Partially supported by Italian Ministry of University and Research (M.U.R.S.T.) and by National Council for Research (C.N.R.).

T. Helleseth (Ed.): Advances in Cryptology - EUROCRYPT '93, LNCS 765, pp. 118-125, 1994.

1. for any k-tuple of distinct indexes $i_1, \ldots, i_k$, $1 \leq i_j \leq n$, for each $(d_1, \ldots, d_k) \in D_{i_1} \times \ldots \times D_{i_k}$ such that $p(d_1, \ldots, d_k) > 0$ there exists an unique secret $s \in S$ such that $p(s \mid d_1, \ldots, d_k) = 1$,
2. for any $j < k$, for any j-tuple of distinct indexes $i_1, \ldots, i_j$, $1 \leq i_j \leq n$, for each $d_1, \ldots, d_j \in D_{i_1} \times \ldots \times D_{i_j}$, such that $p(d_1, \ldots, d_j) > 0$, for each $s \in S$ it holds $p(s \mid d_1, \ldots, d_j) = p(s)$.

The first property implies that the shares held by any group of k participants univocally determines the secret $s \in S$. Notice that the second property means that the probability that the secret is equal to s given that the shares held by any group of $j < k$ participants are $d_1, \ldots, d_j$ is the same as the *a priori* probability that the secret is s. Therefore, no amount of knowledge of shares of less than k participants enables a Bayesian opponent to modify an *a priori* guess regarding which the secret is.

Using the information theoretic concepts of entropy (see Appendix for definitions and properties) the two previous conditions can be stated as follows [12], [8], [3]:

1'. for any k-tuple of distinct indexes $i_1, \ldots, i_k$, $1 \leq i_j \leq n$, it holds $H(S \mid D_{i_1}, \ldots, D_{i_k}) = 0$,

2'. for any $j < k$, for any j-tuple of distinct indexes $i_1, \ldots, i_j$, $1 \leq i_j \leq n$, it holds $H(S \mid D_{i_1}, \ldots, D_{i_j}) = H(S)$.

Tompa and Woll [20] considered the following scenario: let us suppose that $k-1$ participants $P_{i_1}, \ldots, P_{i_{k-1}}$ want to cheat a k-th participant P_{i_k}. Let $d_1, \ldots, d_{k-1}, d_k$ be the shares held by participants $P_{i_1}, \ldots, P_{i_k}$ and s be the correct secret, that is, the secret the participants would reconstruct if they pooled together their shares. The $k-1$ cheaters $P_{i_1}, \ldots, P_{i_{k-1}}$, not knowing d_k, could return $d'_1, \ldots, d'_{k-1}$ forged shares in a tentative to force the k-th participant P_{i_k} to reconstruct a secret $s' \neq s$. Tompa and Woll showed that Shamir's scheme [15] is insecure against this attack, in the sense that even a single participant, with high probability, can deceive other $k-1$ honest participants. Tompa and Woll, however, modified Shamir's scheme to make it secure against cheating. Briefly, they proposed a sharing algorithm that specifies a subset S_{legal} of the set S of possible secrets. A secret will be accepted as authentic *only if it is an element of* S_{legal}. *If a set of* k *participants calculate the secret to be an element* of $S_{illegal} = S - S_{legal}$, then they realize that at least one of them is cheating. In other words S_{legal} is the set of legal secrets that each participant would expect to reconstruct. $S_{illegal}$ is a set of illegal secrets that is introduced only to reveal cheating. Other papers that addressed the problem of coping with cheaters in secret sharing schemes are [1], [7], [14], [16] and [20].

An important issue in the implementation of secret sharing schemes is the size of shares distributed to participants since the security of a system degrades as the amount of the information that must be kept secret increases. Recently, several papers studied this topic and both upper bounds and lower bounds on the size of the shares have been provided [3], [4], [5], [6], [8], [19]. In this paper we study the amount of secret information that must be given to participants in terms of the probability that the previously described attack be successful. Our motivations are based on the observation that the Tompa and Woll secret sharing scheme requires that each participant must receive an amount of secret information that grows with the level of security one imposes against dishonest coalitions. We show that this phenomenon is unavoidable, in the sense that in any secret sharing scheme that has probability of successful cheating less than some $\epsilon > 0$, it must give to each participant shares whose size is at least the size of the secret plus $\log \frac{1}{\epsilon}$.[1]. The security of the schemes presented in this paper is unconditional, since they are not based on any computational assumption.

2 Robust Secret Sharing Schemes

Tompa and Woll [20] defined the cheating probability as "*the probability that from $k-1$ forged shares $d'_1, \ldots, d'_{k-1}$ and any d_k the secret s' reconstructed is legal, but not a correct one*". In order to formally define the problem let us introduce some notations. For each k-tuple of distinct indexes $i_1, \ldots, i_k$, $1 \leq i_j \leq n$, and for any $(d_1, \ldots, d_k) \in D_{i_1} \times \ldots \times D_{i_k}$ such that $p(d_1, \ldots, d_k) > 0$ (i.e., for any k-tuple of shares that the secret sharing algorithm can possibly give to participants $P_{i_1}, \ldots, P_{i_k}$ to share a particular secret $s \in S$), let us denote by $(d_1, \ldots, d_k) \rightarrow s$ the fact that the values $d_1, \ldots, d_k$ force participants $P_{i_1}, \ldots, P_{i_k}$ to reconstruct the secret $s \in S$. Since the sharing algorithm and the probability distribution $\{p(s)\}_{s \in S}$ are known to all participants, it follows that the probability distribution $\{p(d_1, \ldots, d_n)\}_{d_1 \in D_1, \ldots, d_n \in D_n}$ is also known. Assume that the $k-1$ cheaters $P_{i_1}, \ldots, P_{i_{k-1}}$ know the correct secret s and their legal shares $d_1, \ldots, d_{k-1}$. From Decision Theory it is well known (see for example [11]) that the decision rule that minimizes the probability of error is the Bayesian decision rule that chooses the hypothesis with largest "a posteriori" probability. Therefore, the best strategy the $k-1$ cheaters $P_{i_1}, \ldots, P_{i_{k-1}}$ can follow to cheat a k-th participant P_{i_k} is to give him forged values $d'_1, \ldots, d'_{k-1}$ that maximize the following quantity

$$\sum_{d_k \in D_{i_k} : (d'_1, \ldots, d'_{k-1}, d_k) \rightarrow S - \{s\}} p(d_k \mid d_1, \ldots, d_{k-1}, s) .$$

Averaging on all secrets in S and on the possible shares the sharing algorithm could give, we have that the maximum average probability $P(Cheat \mid D_{i_1}, \ldots, D_{i_{k-1}}, S)$ that $k-1$ participants $P_{i_1}, \ldots, P_{i_{k-1}}$, knowing the correct secret, succedees in cheating the k-th participant P_{i_k} is:

[1] All logarithms in this paper are of base 2

$$P(Cheat \mid D_{i_1},\ldots,D_{i_{k-1}},S)$$
$$= \sum_{d_1\in D_{i_1},\ldots,d_{k-1}\in D_{i_{k-1}},s\in S} p(d_1,\ldots,d_{k-1},s)\max{}_{d'_1,\ldots,d'_{k-1}}\left[\sum_{d_k\in D_{i_k}:(d'_1,\ldots,d'_{k-1},d_k)\rightarrow S-\{s\}} p(d_k \mid d_1,\ldots,d_{k-1},s)\right]$$

For a fixed ϵ, $1 \geq \epsilon > 0$, we define a (k,n,ϵ) Robust Secret Sharing Scheme as secret sharing scheme that satisfies the following properties:

P1) for any k-tuple of distinct indexes $i_1,\ldots,i_k$, it holds $H(S \mid D_{i_1},\ldots,D_{i_k}) = 0$,

P2) for any j-tuple of indexes $i_1,\ldots,i_j$, $1 \leq j < k$, it holds $H(S \mid D_{i_1},\ldots,D_{i_j}) = H(S)$,

P3) $P(Cheat \mid D_{i_1},\ldots,D_{i_{k-1}},S) \leq \epsilon$.

Properties P1 and P2 are those of a (k,n) threshold scheme. The property P3 assures that any cheating tentative has arbitrarily small probability of succeeding, even though the cheaters know the correct secret.

Note that the condition $\epsilon > 0$ for (k,n,ϵ) Robust Secret Sharing Schemes is necessary, since the probability of cheating $P(Cheat \mid D_{i_1},\ldots,D_{i_{k-1}},S)$ cannot be 0 in any (k,n) Threshold Secret Sharing Scheme. For a proof of this fact see the remark following Lemma 2.

A quantity that will play an important role to derive our result is the probability that $k-1$ participants $P_{i_1},\ldots,P_{i_{k-1}}$ can guess the share of the k-th participant P_{i_k}, given that they know the correct secret s besides their own shares $d_1,\ldots,d_{k-1}$. Again, the best strategy that the $k-1$ participants can follow is to choose the value $\underline{d}_k \in D_{i_k}$ that maximizes the conditional probability

$$p(d_k \mid d_1,\ldots,d_{k-1},s).$$

Averaging on all the possible shares and secrets it follows that the maximum average probability $P(GuessD_{i_k} \mid D_{i_1},\ldots,D_{i_{k-1}},S)$ that $k-1$ participants $P_{i_1},\ldots,P_{i_{k-1}}$, knowing the correct secret, succeeds in guessing the value of the share of the k-th participant P_{i_k} is:

$$P(GuessD_{i_k} \mid D_{i_1},\ldots,D_{i_{k-1}},S)$$
$$= \sum_{d_1\in D_{i_1},\ldots,d_{k-1}\in D_{i_{k-1}},s\in S} p(d_1,\ldots,d_{k-1},s)\max{}_{d_k\in D_{i_k}} p(d_k \mid d_1,\ldots,d_{k-1},s).$$

The following lemma holds.

Lemma 1 *In any (k,n) Threshold Secret Sharing Schemes one has*

$$H(D_{i_k} \mid D_{i_1},\ldots,D_{i_{k-1}},S) \geq \log \frac{1}{P(GuessD_{i_k} \mid D_{i_1},\ldots,D_{i_{k-1}},S)}.$$

Proof. Since the function $\log x$ is convex, the lemma follows from Jensen inequality [13]. Indeed, we have

$$\begin{aligned}
-H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S) &= \sum_{d_1,\ldots,d_k,s} p(d_1,\ldots,d_k,s) \log p(d_k \mid d_1,\ldots,d_{k-1},s) \\
&\leq \log \left[\sum_{d_1,\ldots,d_k,s} p(d_1,\ldots,d_k,s)\, p(d_k \mid d_1,\ldots,d_{k-1},s) \right] \\
&\qquad\qquad \text{(from Jensen inequality)} \\
&\leq \log \left[\sum_{d_1,\ldots,d_{k-1},s} p(d_1,\ldots,d_{k-1},s) \max_{d_k} p(d_k \mid d_1,\ldots,d_{k-1},s) \right] \\
&= \log \; P(GuessD_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S).
\end{aligned}$$

□

The relationship between the cheating probability $P(Cheat \mid D_{i_1}, \ldots, D_{i_{k-1}}, S)$ and the guess probability $P(GuessD_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S)$ is stated by the following lemma:

Lemma 2 *In any (k,n) Threshold Secret Sharing Schemes one has*

$$P(GuessD_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S) \leq P(Cheat \mid D_{i_1}, \ldots, D_{i_{k-1}}, S) \;.$$

Proof. The proof follows from the following argument. For all $d_1 \in D_{i_1}, \ldots, d_{k-1} \in D_{i_{k-1}}$ and $s \in S$, in correspondence of a value $\underline{d}_k \in D_{i_k}$ that maximizes the guess probability, that is, for which $\max_{d_k \in D_{i_k}} p(d_k \mid d_1 \ldots d_{k-1} s) = p(\underline{d}_k \mid d_1 \ldots d_{k-1} s)$, it must exist a choice of $k-1$ shares $d'_1 \in D_{i_1}, \ldots, d'_{k-1} \in D_{i_{k-1}}$ such that $(d'_1, \ldots, d'_{k-1}, \underline{d}_k) \rightarrow s'$, for some $s' \in S - \{s\}$. In the opposite case, the value $\underline{d}_k$ would univocally identify a secret s, in the sense that the k-th participant P_{i_k} knowing only the share $\underline{d}_k$, could reconstruct the correct secret s. This contradicts Property P2. Therefore,

$$\max_{d_k \in D_{i_k}} p(d_k \mid d_1, \ldots, d_{k-1}, s) \leq \max_{d'_1,\ldots,d'_{k-1}} \left[\sum_{d_k \in D_{i_k} : (d'_1,\ldots,d'_{k-1},d_k) \rightarrow S - \{s\}} p(d_k \mid d_1, \ldots, d_{k-1}, s) \right]$$

and the lemma follows. □

Remark. It is clear that the probability of guessing $P(GuessD_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S)$ is positive. Therefore, from Lemma 2 it follows that $P(Cheat \mid D_{i_1}, \ldots, D_{i_{k-1}}, S) > 0$.

The following theorem represents our main result.

Theorem 1 *In any (k, n, ϵ) Robust Secret Sharing Scheme, for any k-tuple of distinct indexes $i_1, \ldots, i_k$ it holds*

$$H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}) \geq H(S) + \log \tfrac{1}{\epsilon}.$$

Proof. From (3) of Appendix we have:

$$\begin{aligned} H(D_{i_k}, S \mid D_{i_1}, \ldots, D_{i_{k-1}}) &= H(S \mid D_{i_1}, \ldots, D_{i_{k-1}}) + H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S) \\ &= H(S \mid D_{i_1}, \ldots, D_{i_k}) + H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}) \ . \end{aligned}$$

As a consequence of Properties P1 and P2 it follows that

$$H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}) = H(S) + H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S).$$

From Lemmas 1 and 2, and Property P3 of (k, n, ϵ) Robust Secret Sharing Scheme, the theorem follows. Indeed,

$$\begin{aligned} H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}) &= H(S) + H(D_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S) \\ &\geq H(S) + \log \frac{1}{P(GuessD_{i_k} \mid D_{i_1}, \ldots, D_{i_{k-1}}, S)} \quad \text{(by Lemma 1)} \\ &\geq H(S) + \log \frac{1}{P(Cheat \mid D_{i_1}, \ldots, D_{i_{k-1}}, S)} \quad \text{(by Lemma 2)} \\ &\geq H(S) + \log \frac{1}{\epsilon} \quad \text{(by Property P3).} \end{aligned}$$

□

Corollary 1 *In any (k, n, ϵ) Robust Secret Sharing Scheme, if the secret is uniformly chosen it holds*

$$\log \mid D_i \mid \geq \log \mid S \mid + \log \tfrac{1}{\epsilon}.$$

Proof. The proof is immediate from Theorem 1 and properties (1) and (2) of the entropy in the Appendix. □

The corollary shows that in a (k, n, ϵ) Robust Secret Sharing Scheme the size of the shares given to participants — measured as the number of bits necessary to their representation — necessarily grows as ϵ decreases. For completeness, we recall that in the Tompa and Woll algorithm the size of shares $\log |D_i|$ satisfies the bound

$$2\log\left(\frac{(|S|-1)(k-1)}{\epsilon} + k\right) < \log|D_i| < 2\log\left(\frac{(|S|-1)(k-1)}{\epsilon} + k\right) + 1.$$

We have proved a tradeoff between the size of the shares and the probability of successful cheating in perfect (i.e., properties 1. and 2. hold) threshold schemes. The same technique can be used for non-perfect schemes, such as, for example, ramp schemes.

References

[1] M. Ben-Or, T. Rabin *Verifiable Secret Sharing and Multiparty Protocols with Honest Majority*, Proc. 21st ACM Symposium on Theory of Computing, pp. 73-85, 1989.

[2] G. R. Blakley, *Safeguarding Cryptographic Keys*, Proceedings AFIPS 1979 National Computer Conference, pp. 313-317, 1979.

[3] C. Blundo, A. De Santis, L. Gargano, and U. Vaccaro, *On the Information Rate of Secret Sharing Schemes*, "Advances in Cryptology - CRYPTO 92", Ed. E. Brickell, "Lecture Notes in Computer Science", Springer-Verlag, (to appear).

[4] C. Blundo, A. De Santis, D. R. Stinson, and U. Vaccaro, *Graph Decomposition and Secret Sharing Schemes*, "Advances in Cryptology - EUROCRYPT 92", Ed. R. Rueppel, "Lecture Notes in Computer Science", Springer-Verlag, (to appear).

[5] E. F. Brickell and D. M. Davenport, *On the Classification of Ideal Secret Sharing Schemes,* J. Cryptology, Vol.**4**, 123–134, 1991.

[6] E. F. Brickell and D. R. Stinson, *Some Improved Bounds on the Information Rate of Perfect Secret Sharing Schemes*, J. Cryptology, Vol. **5**, pp. 153–156, 1992.

[7] E. F. Brickell, D. R. Stinson, *The Detection of Cheaters in Threshold Schemes*, SIAM J. Disc. Math, Vol. **4**, pp. 502-510, 1991.

[8] R. M. Capocelli, A. De Santis, L. Gargano, U. Vaccaro, *On the Size of Shares for Secret Sharing Schemes*, Advances in Cryptology – CRYPTO '91, J. Feigenbaum (Ed.), Lectures Notes in Computer Science, Vol. **576**, pp. 101–113, 1992, Springer-Verlag. Also to appear in Journal of Cryptology.

[9] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems,* Academic Press, 1981.

[10] D. Denning, *Cryptography and Data Security,* Addison–Wesley, Reading, MA, 1983.

[11] T.S. Ferguson, *Mathematical Statistics*, Academic Press, New York, 1967.

[12] E. D. Karnin, J. W. Greene, and M. E. Hellman, *On Secret Sharing Systems,* IEEE Trans. on Inform. Theory, Vol. **IT-29**, pp. 35–41, 1983.

[13] A.W. Marshall, I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.

[14] R. J. McEliece, D. V. Sarwate, *On Sharing Secrets and Reed-Solomon Codes*, Communications of the ACM, Vol. **24**, pp. 583-584, 1981.

[15] A. Shamir, *How to Share a Secret*, Communication of the ACM, Vol. **22**, pp. 612-613, 1979.

[16] G. Simmons, ***Robust Shared Secret Schemes or "How to be Sure You Have the Right Answer Even Though You Do Not Know the Question"***, Congr. Numer., Vol. **68**, pp. 215-248, 1989.

[17] G. J. Simmons, ***An Introduction to Shared Secret and/or Shared Control Schemes and Their Application***, Contemporary Cryptology, IEEE Press, pp. 441–497, 1991.

[18] D. R. Stinson, ***An Explication of Secret Sharing Schemes***, Design, Codes and Cryptography, Vol. **2**, pp. 357–390, 1992.

[19] D. R. Stinson, ***Decomposition Constructions for Secret Sharing Schemes***, Technical Report UNL-CSE-92-020, Department of Computer Science and Engineering, University of Nebraska, September 1992.

[20] M. Tompa, H. Woll, *How to Share a Secret with Cheaters*, Journal of Cryptology, Vol. **1**, pp. 133-139, 1988.

3 Appendix

In this section we shall review the information theoretic concepts we used in the paper. For a complete treatment of the subject we refer the reader to [9].

Given a probability distribution $\{p(x)\}_{x \in X}$ on a finite set X, define the *entropy* of X, $H(X)$, as:

$$H(X) = -\sum_{x \in X} p(x) \log p(x).$$

The entropy $H(X)$ is a measure of the average information content of the elements in X or, equivalently, a measure of the average uncertainty one has about which element of the set X has been chosen when the choices of the elements from X are made according to the probability distribution $\{p(x)\}_{x \in X}$. The entropy $H(X)$ enjoys the following property:

$$0 \leq H(X) \leq \log |X|, \tag{1}$$

where $H(X) = 0$ if and only if there exists $x_0 \in X$ such that $p(x_0) = 1$; $H(X) = \log |X|$ if and only if $p(x) = 1/|X|$, for each $x \in X$.

Given two sets X and Y and a joint probability distribution $\{p(x, y)\}_{x \in X, y \in Y}$ on their cartesian product, the *conditional entropy* $H(X|Y)$ of X given Y is defined as:

$$H(X|Y) = -\sum_{y \in Y} \sum_{x \in X} p(y) p(x|y) \log p(x|y).$$

The conditional entropy satisfies the following inequalities

$$H(X) \geq H(X|Y) \geq 0. \tag{2}$$

The entropy of the joint space XY satisfies:

$$H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y).$$

Analogously, the conditional entropy of XY given Z satisfies:

$$H(XY|Z) = H(X|Z) + H(Y|XZ) = H(Y|Z) + H(X|YZ). \tag{3}$$