SMALLEST POSSIBLE MESSAGE EXPANSION IN THRESHOLD SCHEMES

G. R. Blakley
Department of Mathematics
Texas A&M University
College Station, Texas 77843-3368

R. D. Dixon
Department of Computer Science
Wright State University
Dayton, Ohio 45435

A k out of n threshold scheme of any sort known to date involves at least nfold message expansion at the source (where the shadows of the original message are produced). Also, at least k times as much text must be input to the recovery process as is output from it. Linear ramp schemes are more economical but they give only Shannon relative security [BL85]. Is it possible to retain Shannon perfect security and yet cut down the message expansion from at least nfold at the source and at least kfold at the time and place of message recovery? No. In fact the message expansion attained by a Shamir scheme [SH79] or the rigid version [BL85] of a Blakley linear scheme (these two are merely duals of each other) is, in a sense, best possible. Moreover this best possible expansion is slightly larger than just nfold and kfold. The actual expansion factors involve an additive log term. We assume that the reader is familiar with [SH79] and [BL85], and their terminology

Let k be a positive integer. Let P and N be finite sets such that

$$1 \leq k \leq n = \text{card}(N) \leq \text{card}(P) - 1$$

Here $\text{card}(N)$ stands for the cardinality of the set N. Similarly $\text{card}(P)$. Fix any α which does not belong to N. Let

A.M. Odlyzko (Ed.): Advances in Cryptology - CRYPTO '86, LNCS 263, pp. 266-274, 1987.

$$V = \bigcup P^{I \cup \{\alpha\}}$$

$$W = \bigcup P^{H}$$

where the unions are over all (k-1)-member subsets I of N, and over all k-member subsets H of N, respectively. A k out of N threshold scheme over P is a pair (E,D) of maps

$$E : V \to P^{N \cup \{\alpha\}}$$

$$D : W \to P$$

with the properties that

$$E(\phi)\big|_{I \cup \{\alpha\}} = \phi$$

$$D(E(\phi)\big|_G) = \phi(\alpha)$$

for every k-entry list ϕ belonging to V, and every k-member subset G of N. As usual, the restriction $E(\phi)\big|_G$ of the function

$$E(\phi) : N \cup \{\alpha\} \to P$$

is the k-member sublist of the n-member list $E(\phi)$ which consists of only those pairs $(t, E(\phi)[t])$ for which t belongs to G. Similarly $E(\phi)\big|_{I \cup \{\alpha\}}$.

Comment: If the k-entry list ϕ belongs to V then its E-image $E(\phi)$ is a member of $P^{N \cup \{\alpha\}}$ and, thus, amounts to a list with n+1 entries. The list $\lambda = E(\phi)$ can have two equal entries, i.e. it is possible to have $\lambda(i) = \lambda(j)$ for two distinct members i, j of $N \cup \{\alpha\}$. But when you consider λ as a set of exactly n+1 ordered pairs belonging to $(N \cup \{\alpha\}) \times P$ then no two members of this set of ordered pairs can coincide. For any choice of ϕ belonging to V we will define a shadow of $\phi(\alpha)$ belonging to P to be a member of $\lambda\big|_N$, considered as an n-member subset of $N \times P$. Thus when enough random material has been chosen so that the substance (i.e. message) $\phi(\alpha)$ has given rise to n shadows, no two of these shadows can coincide. This is not a subtlety. Consider a rigid Shamir 3

out of $\{1, 2, 3, 6, 8\}$ scheme on GF(13). Let the substance be $\phi(0) = 2$ and suppose two appeals to the random number generator yield

$$\phi(2) = \phi(8) = 2.$$

Then the quadratic polynomial ϕ happens to be the constant polynomial

$$\phi(x) = 2 = 2 + 0x + 0x^2$$

As we have noted, the five shadows of $\phi(0) = 2$ are not the numbers

$$\phi(1) = 2$$
$$\phi(2) = 2$$
$$\phi(3) = 2$$
$$\phi(6) = 2$$
$$\phi(8) = 2$$

but are, instead, the five ordered pairs

$$(1, \phi(1)) = (1, 2)$$
$$(2, \phi(2)) = (2, 2)$$
$$(3, \phi(3)) = (3, 2)$$
$$(6, \phi(6)) = (6, 2)$$
$$(8, \phi(8)) = (8, 2).$$

And no two of these five ordered pairs are equal.

The reader might feel that our definition admits Shamir and Blakley schemes but rules out the one-time pad. For in this 2 out of {pad, transmission} case it would indeed seem possible to have

$$\text{substance} = 0$$
$$\text{pad} = 0$$
$$\text{transmission} = 0.$$

Thus it would seem that the two shadows (i.e. the pad and the transmission) are both equal to zero. We will not address this point directly. We will, instead, say two things. First, one-time pads obey the conclusions of our theorem even if they do not obey its hypotheses. Second, a decoder could not place any reliance on the inference

$$\begin{aligned} 0 &= \text{substance} \\ &= \text{pad XOR transmission} \\ &= 0 + 0 \end{aligned}$$

with only two bits of information lying around: a 0 and a 0. The decoder would only feel confident if at least one more bit of information were available, namely a yes answer to the question: "Is this 0 really the pad and that 0 really the transmission?" We will return to this idea shortly.

The deeper question of how to formulate a definition of threshold scheme which clearly describes and utilizes all the information really available to the decoder and which, perhaps, leads to a more inclusive theorem with conclusions as strong as ours will be left to the reader.

For every (k-1)-member subset I of N, we define probability density functions

$$\mu : P^{\{\alpha\}} \to [0,1]$$

$$\nu_I : P^I \to [0,1]$$

$$\lambda_I = \mu \times \nu_I : P^{I \cup \{\alpha\}} \to [0,1]$$

in such a way that ν_I is a uniform pdf. Such a threshold scheme is called Shannon perfectly secure through the disclosure of k-1 shadows. The reason for this is that, on the basis of the probability assessment [BL81] based on these measures, we have the equality

$$\begin{aligned} &\text{a posteriori probability that } \phi(\alpha) = \pi, \text{ given that } \phi\big|_G = \psi \\ &\quad = \text{a priori probability that } \phi(\alpha) = \pi \end{aligned}$$

for every π belonging to P, every subset G of N such that card(G) = k-1, and every function ψ: G → P.

Theorem 1: Let (E, D) be a k out of N threshold scheme on P. Suppose that it is Shannon perfectly secure through the disclosure of k-1 shadows. Suppose there is a way of representing shadows as bit strings. Then the average length of a shadow is no less than log ([card(P)][card(N) - k + 1]) bits.

Example 1: In Shamir's scheme a shadow is a pair (x, p(a(x))). The nonnegative integer x tells which shadow it is. The substance is p(0). a(x) is a member of $GF(2^L)$ which has been fixed and published in advance for all x belonging to {1,2, ..., n}. p is a polynomial function p: $GF(2^L) \to GF(2^L)$ of degree k-1. Now it is possible that p(a(x)) = p(a(y)) for $x \neq y$. But no two shadows can coincide. This is true because the equality (x, p(a(x))) = (z, p(a(z))) simply means that you are comparing two copies of the same shadow, not two shadows, since x = z. So Shamir's scheme satisfies our hypotheses. And it is right at the lower bound if L = 3, k = 4, n = 7. A pair (w,p(w)) could be formatted as a 3-bit string prefixed by as many bits as needed to identify one of 7 shadows. If p(w) = b(3) b(2) b(1) is always a string of 3 bits then the shadows are of the forms:

b(3) b(2) b(1);
1 b(3) b(2) b(1);
1 0 b(3) b(2) b(1);
1 1 b(3) b(2) b(1);
1 0 0 b(3) b(2) b(1);
1 0 1 b(3) b(2) b(1);
1 1 0 b(3) b(2) b(1).

their average length is (3+4+5+5+6+6+6)/7 = 5 bits. But the bound here is equal to L + log(n-k+1) = 3 + log(7-4+1) = 5. Somebody might say that you could also omit a high order bit b(3) if it equals 0. But then you would actually have to symbolize the comma in the expression (x,p(a(x))) some way. And this would add

bits to each word. You could go at it the other way and write out all the x values as 3-bit numbers and let the field elements be variable in length. In this approach you have

$$b(3)\ b(2)\ b(1)\ x(3)\ x(2)\ x(1)$$

dropping high order b bits which are zero. But here again the possible shadows are:

$$x(3)\ x(2)\ x(1);$$
$$1\ x(3)\ x(2)\ x(1);$$
$$1\ 0\ x(3)\ x(2)\ x(1);$$
$$1\ 1\ x(3)\ x(2)\ x(1);$$
$$1\ 0\ 0\ x(3)\ x(2)\ x(1);$$
$$1\ 0\ 1\ x(3)\ x(2)\ x(1);$$
$$1\ 1\ 0\ x(3)\ x(2)\ x(1);$$
$$1\ 1\ 1\ x(3)\ x(2)\ x(1).$$

The average length is now (3+4+5+5+6+6+6+6)/8 = 5.125 bits.

This is one example of a general phenomenonon. A Shamir scheme over $GF(2^L)$ achieves the theorem's bound, L + log(n-k+1) when n-k+1 is an integer power of 2.

Example 2: Consider a one-time pad for transmitting messages belonging to a set P of cardinality 4. Here k = card(N) = 2. On the face of things it would appear that the average shadow size is

$$(1 + 1 + 2 + 2)/4 = 3/2$$

which is smaller than the bound

$$\log(4) * (2-2+1) = 2$$

in the theorem. Does this mean that the theorem is merely a curiosity with so many hypotheses that it cannot usefully apply to the simplest cases? Quite the

contrary. When we look more closely at this example we will find that the decoder must have, on the average, four bits of information when he applies the decode process to the two shadows he requires (i.e. the secret pad and the nonsecret transmitted message. We may as well let P = {0, 1, 10, 11}. There are 16 possibilities

		secret 0	1	shared 10	pad 11
public transmitted message	0	0	1	10	11
	1	1	0	11	10
	10	10	11	0	1
	11	11	10	1	0
		secret message to be communicated			

A naive observer might conclude that the 16 possible outcomes lead to 16 decodes with a total number of input bits equal to

$$(2+2+3+3) + (2+2+3+3) + (3+3+4+4) + (3+3+4+4) = 48.$$

Such an observer would believe the 1.5-bits-per-shadow average alluded to above. The error in such a viewpoint lies in not looking at the whole picture. When two shadow words (e.g. 1 and 11) are ready to be XORed together to produce the reconstructed plaintext word 10 (i.e. the substance 10), the decoder does not have merely three bits of information. Some person or device has inspected the pad and found it to contain a 1, and has monitored the channel and verified that a 11 was actually received in what looks like a legitimate transmission. Thus there is at least one more bit of information available at the decoder. This bit corresponds to a yes answer to the question "Does the pair consisting of 1 and 11 constitute a valid input, one consisting a word from the pad and a word transmitted in the agreed manner down the channel? The decoder thus has 4 bits of information when it forms 1 XOR 11 = 10. When we take this into account and

average over the 16 possible pairs we find the average number of bits available at decode to be 1/16th of the sum

$$(3+3+4+4) + (3+3+4+4) + (4+4+5+5) + (4+4+5+5) = 64.$$

Hence the average number of bits per shadow is 2.

It follows that the one time pad, which does not seem to obey the hypotheses of Theorem 1 (because pad = 0, transmission = 0 is allowable, in violation of the assumption that no two shadows are equal), nevertheless obeys its conclusions.

The purpose of Example 2 is to make the following point. We believe that the bound in Theorem 1 cannot be bettered if one takes into account all the information available to a decoder. In other words we believe that our hypotheses are unduly restrictive and that the message bandwidth expansion attained by, for example, certain rigid Shamir schemes [SH79] is best possible no matter how you define a shadow. Decode isn't possible if you merely have a few members of a message space. You must have some further information. And the amount of further information needed is as much as if you knew where each of your message space members occurred in the output stream of the encoder.

The proof of the theorem.

Now fix any (k-1)-member subset I of N. Note that α is not a member of I. Fix any $\pi \in P$. Fix any $\psi \in P^I$ and let $\beta[\psi,\alpha,\pi] \in P^{I \cup \{\alpha\}}$ be the k-entry list such that

$$\beta[\psi,\alpha,\pi](t) = \psi(t)$$

for every t belonging to I, and such that

$$\beta[\psi,\alpha,\pi](\alpha) = \pi.$$

Define $f: P \to P$ by requiring that

$$f(\pi) = D(\beta[\psi,\alpha,\pi]).$$

Because of Shannon perfect security, f must be a surjection. For, otherwise, possession of $k-1$ shadows would enable somebody to rule out some values of the substance as impossible. In fact for every z belonging to P and every s belonging to $N \setminus I$ there exists $\Upsilon[z,s]$ belonging to P such that

$$D(\delta[\psi,s,\Upsilon[z,s]]) = \pi$$

where

$$\delta[\psi,s,\Upsilon[z,s]](t) = \psi(t)$$

for every $t \in I$ and such that

$$\delta[\psi,s,\Upsilon[z,s]](s) = \Upsilon[z,s].$$

Since no two shadows of π can be equal for any k-member subset $H = I \cup \{s\}$ of N it follows that there are at least $n-k+1$ such shadows $(s,\Upsilon[z,s])$. This is true for every $\pi \in P$. Hence, for any choice of a k-member subset H of N there are at least $\mathrm{card}(P) * (n-k+1)$ preimages.

References

BL81 G. R. Blakley and L. Swanson, Security proofs for information protection schemes, Proceedings of the 1981 Symposium on Security and Privacy, IEEE Computer Society, Los Angeles (1981), pp. 75-88.

BL85 G. R. Blakley and C. Meadows, Security of Ramp Schemes, in G. R. Blakley and D. Chaum, Editors, Advances in Cryptology: Proceedings of Crypto '84, Springer-Verlag, Berlin (1985), pp. 242-268.

SH79 A. Shamir, How to share a secret, Communications of the ACM, vol. 22 (1979), pp. 612-613.