J. Cryptology (1990) 2: 13–22

Journal of Cryptology

Probability To Meet in the Middle[1]

Kazuo Nishimura
Department of Business Administration, Komazawa University,
Komazawa, Setagaya-ku, Tokyo, Japan 154
Nishimura@math.keio.ac.jp

Masaaki Sibuya
Department of Mathematics, Keio University,
Hiyoshi, Kohoku-ku, Yokohama, Japan 223

Abstract. This paper proposes "Matching Models A, B, and C" for the meet-in-the-middle attacks against a message with digital signature to provide a more sound foundation for the calculation of the probability of success. The typical procedures by Yuval and by Merkle are regarded as Model A or "birthday paradox in two groups," and are different from the classical birthday paradox. Models B and C are applicable for other similar procedures.

The relationship of Matching Models and probabilistic models for testing the algebraic structure of DES is also discussed.

Key words. Authentication, Digests, Forgery, Hash functions, Data Encryption Standard, Birthday problem, Occupancy, Urn models.

1. Introduction

The high probability of success in the "meet-in-the-middle attack" is often credited to the classical birthday paradox. The credit is not exact and is misleading. Although the asymptotic conclusions in the literature are correct, finite probabilistic models should be considered for each version of the attack as theoretical foundations. In this note Matching Models A, B, and C are proposed for some versions of the attack, and the exact probabilities of success are calculated. Model A, which is adequate for the typical attack, can be regarded as "birthday paradox in two groups."

Digital signature using compressed encoding was proposed by Rabin [12] to confirm efficiently the contents of a message, document, or file as the unchanged original. For this purpose a conventional encryption function $E(K, M)$, a mapping of the pair of a κ-bit key K and a μ-bit message M to a μ-bit cipher, is used for hashing, or compressed encoding, of a message to produce its digest, or a sort of checksum. Divide a message W into a sequence of κ-bit fragments $(W_i)_{i=1}^{l}$. Starting

[1] Date received: September 20, 1988. Date revised: October 7, 1989.

from some initial μ-bit code H_0, using W_i's as keys, generate

$$H_i = E(W_i, H_{i-1}), \qquad i = 1, 2, \ldots, l,$$

and the pair (H_0, H_l) is the digest of W. Then the digest is signed by a public-key authentication method. It is expected that any change of W will produce a different H_l and that to find a different message which produces the same H_l is intractable.

Yuval discussed the weak points of Rabin's procedure, and pointed out a possible attack as follows [14, Appendix B]. Two parties A and B are going to have a contract. Party B writes up one good contract which A is willing to sign, and one bad contract which A would never sign. Then B generate $2^{\mu/2}$ random perturbations on both of them, and the digest for all these perturbed documents, hoping to get a pair of good and bad ones with the same digest. Yuval tried to justify his procedure by the birthday paradox. However, his proof was vague. In fact, he was not discussing matching within a group like the birthday paradox but between good contracts and bad contracts. This point has been overlooked in the literature and the phrases "birthday effect" or "birthday attack" have been used without critical evaluation.

Subsequently, the attack methods were devised further. Ralph Merkle (attributed to by Davies and Price [3]) showed a way to tamper with a message to obtain another one having the same digest. This procedure was called the meet-in-the-middle attack, and appears to be a more plausible method. Other attacks have been reviewed by Akl [1] and Winternitz [13]. All the attacks are based on the random perturbation of a given message. If many fragments of the message are perturbed, independent random codes on the cipher space will be expected.

In Section 2 of this paper, Yuval's and Merkle's procedures are formulated as Matching Model A. This is a new probabilistic model showing the "birthday paradox in two groups." It is shown that if m $(= 2^{\mu})$ is the cardinality of the message space, the meet-in-the-middle attack will succeed with probability p by $(-m \log(1-p))^{1/2}$ trials, while the corresponding straightforward attack needs $-m \log(1-p)$ trials (Propositions 1 and 2). The mathematical details of the analysis were published in another paper [11].

In Section 3 a simpler procedure involving tampering of two consecutive fragments of a message is considered. Actually, perturbation of a word or phrase is difficult and the procedure is limited to tampering with numbers or codes. Matching Model B, a candidate for this procedure, has simply a hypergeometric distribution. In Section 4 another procedure, studied by Mueller-Schloer [9], is considered. The third Matching Model C for this procedure is a sort of mixture of Models A and B and is also new, however, the necessary analysis is simple. In these three models asymptotic behavior of the matching probabilities are the same.

Our discussion is based on the randomness assumption of the encryption and decryption functions to be used, actually the Data Encryption Standard (DES) [10]. The question whether DES has some algebraic structure or can be considered random in a sense has been studied extensively by Kaliski *et al.* [7]. To break DES itself they considered a version of the meet-in-the-middle attack. Since its success depends on the real characteristics of DES, they proposed the use of the attack for

the experimental testing of the DES characteristics. In Section 3 their attack is related to Matching Model B.

Another "cycling test" for DES, which is better than the meet-in-the-middle test in the space complexity comparison, was also proposed by Kaliski *et al.* [7]. In Section 5 probabilistic aspects of the test are mentioned, as the classical birthday paradox is related to this test. In the final section open questions for future research are presented.

Before completing this section we review the classical birthday problem for completeness, see [4]. Assume that the birthdays of a group of n persons are distributed independently on m ($= 365$) days, and let T denote the number of days which represent someone's birthday. Then

$$\Pr[T < n] = 1 - \Pr[T = n] = 1 - \frac{m^{(n)}}{m^n}, \tag{1}$$

where $m^{(n)} = m(m-1)\cdots(m-n+1)$. The probability is paradoxically large for even small n—when $m = 365$: 0.507 at $n = 23$ and 0.970 at $n = 50$. If $m \to \infty$ the probability (1) is

$$1 - \exp\left[-\frac{n(n-1)}{2m} + O(m^{-1/2})\right] \quad \text{if} \quad n = O(m^{1/2}) \tag{2}$$

or

$$\frac{n(n-1)}{2m} + O\left(\frac{1}{m^2}\right) \quad \text{if} \quad n = O(1).$$

Further, T follows the "classical occupancy distribution";

$$\Pr[T = t; n] = \left\{ {n \atop t} \right\} \frac{m^{(t)}}{m^n}, \qquad 1 \le t \le n, \tag{3}$$

where $\left\{ {n \atop t} \right\}$ denotes the Stirling number of the second kind, defined by the polynomial identity,

$$x^n = \sum_{t=1}^{n} \left\{ {n \atop t} \right\} x^{(t)} \tag{4}$$

[5], [6]. Expression (1) is a special case of (3).

In terms of urns and balls [6] the distribution (3) is obtained as follows: An urn contains m balls which are numbered $1, 2, \ldots, m$. A ball is taken out at random and its number is recorded. Repeat this n times independently replacing a selected ball each time, i.e. sampling "with replacement." The number T of distinct selected balls has the distribution (3). Another version which is used in the following is: n balls are randomly thrown into one of m urns and the number T of urns occupied by one or more balls has the probability distribution (3).

Now, the games are played sequentially until the "hitting time." In the last game, for example, balls are thrown one by one until the $(N+1)$st ball falls, for the first time, into an urn which is already occupied by another ball. Since, using the notation (3),

$$\Pr[N \ge t] = \Pr[T = t; t],$$

it is shown that

$$\Pr[N = t] = \Pr[N \geq t] - \Pr[N \geq t+1] = t\frac{m^{(t)}}{m^{t+1}}, \qquad 1 \leq t < \infty. \tag{5}$$

From (2), it is shown that $N^2/2m$ follows asymptotically the standard exponential distribution as $m \to \infty$. This is *not* the "coupon collector's problem," in which the "hitting" means the number of occupied urns reaches a constant fixed in advance.

2. Meet-in-the-Middle Attack; Matching Model A

To introduce notations Rabin's procedure is repeated in a more formal way. A cryptosystem of the conventional type consists of an encryption function E: $\mathcal{K} \times \mathcal{M} \to \mathcal{C}$ and a decryption function D: $\mathcal{K} \times \mathcal{C} \to \mathcal{M}$, where $\mathcal{K}$ denotes a key space, $\mathcal{M}$ denotes a message space, and $\mathcal{C}$ denotes a cipher space. We assume $\mathcal{M} = \mathcal{C} = \{0, 1\}^\mu$, the set of μ-bit codes, $\mathcal{K} = \{0, 1\}^\kappa$, and $D(K, \cdot)$ is the inverse function of $E(K, \cdot)$ for any $K \in \mathcal{K}$; then $\{E(K, M)\colon M \in \mathcal{M}\}$ is a permutation of $\mathcal{M}$ and $\{D(K, M)\colon M \in \mathcal{M}\}$ is its inverse for a fixed $K \in \mathcal{K}$. Typically, DES satisfies the assumption with $\kappa = 56$ and $\mu = 64$.

In Rabin's scheme for compression, a message W is divided into a sequence of κ-bit fragments $(W_i)_{i=1}^l$. Starting from some initial value H_0, generate

$$H_i = E(W_i, H_{i-1}), \qquad i = 1, 2, \ldots, l,$$

and the pair (H_0, H_l) is a digest of W. The digest is signed by a public-key authentication method, and the resulted authenticator is sent with the message W. The receiver can verify W and identify the sender by regenerating the digest.

A forger, knowing (H_0, H_l), wants to tamper with the message, keeping the digest and the authenticator unchanged. (At least the sender and the receiver of the message know the digest and can become the forger.) He tries to generate randomly another sequence $(\tilde{W}_i)_{i=1}^l$ which reaches the same H_l. An example of forging $\tilde{W}_i$'s by perturbation, namely random rephrasing, is illustrated by Davies and Price [3].

In the meet-in-the-middle attack, the forger fixes a middle step r, $0 < r < l$, and starting from both ends, $\tilde{H}_0^+ = H_0$ and $\tilde{H}_l^- = H_l$, he randomly generates forward sequences

$$\tilde{H}_i^+ = E(\tilde{W}_i, \tilde{H}_{i-1}^+), \qquad i = 1, 2, \ldots, r, \tag{6a}$$

and backward sequences

$$\tilde{H}_{i-1}^- = D(\tilde{W}_i, \tilde{H}_i^-), \qquad i = l, l-1, \ldots, r+1. \tag{6b}$$

If a forward result $\tilde{H}_r^+$ and a backward result $\tilde{H}_r^-$ coincide, then the attack is accomplished. In fact, the total length l need not be the original one. We use the symbol l, however, for simplicity.

Now, our Matching Model A is formulated as follows. Assume that the forger wants to tamper with many fragments of the message, for example, the whole message. Let a trial mean one generation of $\tilde{H}_r^+$ by (6a) or $\tilde{H}_r^-$ by (6b). Since $\tilde{H}_r^+$ and $\tilde{H}_r^-$ are obtained by repeated encryptions and decryptions randomly changing the

keys, $\tilde{H}_r^+$ and $\tilde{H}_r^-$ can be regarded as random variables uniformly distributed on $\mathcal{M}$ and independent trial by trial. In this attack, the generated $\tilde{H}_r^+$ and $\tilde{H}_r^-$ correspond to balls, all the possible μ-bit codes at the rth stage correspond to urns, and there are $m = 2^\mu$ urns. Unlike the birthday problem, there are two types of balls: One corresponds to $\tilde{H}_r^+$ and the other to $\tilde{H}_r^-$. Regard these types as balls of different colors, say, white and red. The event of concern is a collision between the two colors. So, modify the classical birthday problem as follows: There are two groups, say, n_1 boys and n_2 girls. What is the probability that a boy's birthday coincides with a girl's birthday? This can be called the "birthday problem in two groups."

Yuval's procedure, mentioned in Section 1, also fits Model A: perturbed good contracts and bad contracts correspond to white and red balls.

Under uniformity and independence assumptions, the probability of no collision between n_1 white and n_2 red balls is shown to be

$$q_1 := \frac{1}{m^{n_1}} \sum_t \left\{ {n_1 \atop t} \right\} m^{(t)} \left(1 - \frac{t}{m}\right)^{n_2} = \frac{1}{m^{n_2}} \sum_t \left\{ {n_2 \atop t} \right\} m^{(t)} \left(1 - \frac{t}{m}\right)^{n_1}$$

$$= \frac{1}{m^{n_1+n_2}} \sum_{v=t_1+t_2} m^{(v)} \left\{ {n_1 \atop t_1} \right\} \left\{ {n_2 \atop t_2} \right\}, \tag{7}$$

where the symbol defined by (4) is used. Because when n_1 white balls are thrown at random into one of m urns, the number T of the urns occupied by the white balls follows the classical occupancy distribution:

$$\Pr[T = t; n_1] = \frac{1}{m^{n_1}} \left\{ {n_1 \atop t} \right\} m^{(t)}, \qquad 1 \le t \le n_1.$$

Under the condition that $T = t$, n_2 red balls are thrown at random into the urns. Then the number S of the red balls falling into the urns that are occupied by white balls is a binomial random variable:

$$\Pr[S = s \mid T = t; n_2] = \binom{n_2}{s} \left(\frac{t}{m}\right)^s \left(1 - \frac{t}{m}\right)^{n_2 - s}.$$

Unconditionally,

$$q_1 = \Pr[S = 0; n_1, n_2] = \sum_t \Pr[S = 0 \mid T = t; n_2] \Pr[T = t; n_1],$$

and this is the first expression of (7). Because the above random event is symmetric with respect to the white and red balls the second expression is equivalent to the first, and definition (4) leads to the third expression. It is shown that for fixed $n_1 + n_2$ the probability q_1 is increasing in $|n_1 - n_2|$. The matching probability $1 - q_1$ is surprisingly large for small $n_1 = n_2 = n$—when $m = 365$: 0.504 at $n = 16$, 0.915 at $n = 30$, and 0.999 at $n = 50$.

In [11] it was shown that q_1 of expression (7) is bounded as follows when $n_1 = n_2 = n > 1$:

$$\left(1 - \frac{1}{m}\right)^{n_2} < q_1 < \exp\left\{-\frac{n^2}{m} + \lambda\left[\exp\left(\frac{n}{m}\right) - 1\right]\right\},$$

where $\lambda = n^2/2(m-n)$. As $m \to \infty$, the above inequalities lead to

$$q_1 = \exp\left\{-\frac{n^2}{m}[1 + O(m^{-1/2})]\right\} \quad \text{if} \quad n = O(m^{1/2}). \tag{8}$$

The corresponding probability by the one-way straightforward attack of n trials is the probability of missing n times to hit a specific H_l:

$$q_2 := \left(1 - \frac{1}{m}\right)^n = \exp\left\{-\frac{n}{m}\left[1 + O\left(\frac{1}{m}\right)\right]\right\}. \tag{9}$$

In summary,

Proposition 1. *In the meet-in-the-middle attack of Matching Model A, the forger's failure probability is given by q_1 in (7). In other words, approximately*

$$n_1 = n_2 = \left(m \log \frac{1}{1-p}\right)^{1/2}$$

trials give the success probability p. In the corresponding one-way attack, the forger's failure probability is given by q_2 in (9). In other words, approximately

$$n = m \log \frac{1}{1-p}$$

trials give the success probability p.

When the attack is tried sequentially, the argument for obtaining (5) is applied. From (8) and (9) we obtain the following result. See [11] for the exact meaning of the asymptotic distribution.

Proposition 2. *In Matching Model A, let $N = N_1 = N_2$ be the hitting time for success of the sequential meet-in-the-middle attack. Then the asymptotic distribution of N^2/m is the standard exponential distribution.*

Let N^* be the hitting time for success of the straightforward attack. Then the asymptotic distribution of N^*/m is the standard exponential distribution.

Thus, a lucky forger can succeed without enormous efforts if Rabin's original scheme for making a digest is adopted.

3. Meet-in-the-Middle Attack; Matching Model B

Matching Model B is as follows. Assume that H_{r-1} and H_{r+1} are fixed, that $\tilde{H}_r^+ = E(\tilde{W}_r, H_{r-1})$ forms a simple random sample without replacement from $\{E(K, H_{r-1}): K \in \mathcal{K}\}$ for n_1 trials of random $\tilde{W}_r$, and that $\tilde{H}_r^- = D(\tilde{W}_{r+1}, H_{r+1})$ forms a similar one, which is independent of the forward sample, for n_2 trials of random $\tilde{W}_{r+1}$. Then the probability that there is no overlap in the two samples is

$$q_3 := \frac{(m-n_1)!\,(m-n_2)!}{m!\,(m-n_1-n_2)!} = \frac{m^{(n_1+n_2)}}{m^{(n_1)}m^{(n_2)}}. \tag{10}$$

Further, the number S of overlapping codes follows the hypergeometric distribution

$$\Pr[S = s; n_1, n_2] = \binom{n_1}{s}\binom{m-n_1}{n_2-s}\Big/\binom{m}{n_2} = \binom{n_2}{s}\binom{m-n_2}{n_1-s}\Big/\binom{m}{n_1}$$
$$= \frac{n_1!\, n_2!\, (m-n_1)!\, (m-n_2)!}{m!\, s!\, (n_1-s)!\, (n_2-s)!\, (m-n_1-n_2+s)!},$$

where $\max(0, n_1 + n_2 - m) \le s \le \min(n_1, n_2)$, and (10) can be obtained as $q_3 = \Pr[S = 0; n_1, n_2]$.

As $m \to \infty$, Stirling's formula and the Taylor expansion of the logarithmic function show that

$$q_3 = \exp\left\{-\frac{n_1 n_2}{m}\left[1 + \frac{n_1 + n_2 - 1}{2m} + O\left(\frac{1}{m}\right)\right]\right\} \quad \text{if } n_i = O(m^{1/2}),\ i = 1, 2, \quad (11)$$

or

$$q_3 = 1 - \frac{n_1 n_2}{m} + \frac{n_1 n_2 (n_1 n_2 - n_1 - n_2 + 1)}{2m^2} + O\left(\frac{1}{m^3}\right) \quad \text{if } n_i = O(1),\ i = 1, 2.$$

If the attack proceeds sequentially, the probabilistic distribution of the hitting time N, the number of trials both forward and backward when the first coincidence occurs, is obtained from (11):

$$\Pr[N > n] = \Pr[S = 0; n, n] = \exp\left\{-\frac{n^2}{m}[1 + O(m^{-1/2})]\right\} \quad \text{if } n = O(m^{1/2}).$$

The variate N^2/m follows asymptotically the standard exponential distribution.

Model B is effective for the attack against the cryptosystem itself [7]. Given a small set of plaintexts $M_0, M_1, \ldots, M_l$ and their ciphertexts $C_i = E(K, M_i)$, $i = 0, 1, \ldots, l$, instead of disclosing K, the cryptanalyst tries to find a pair of keys (K_1, K_2) such that

$$E(K, \cdot) = E(K_2, E(K_1, \cdot)) \quad (12)$$

by the meet-in-the-middle attack. That is, for randomly selecting K_1 and K_2, he tries to find a pair which satisfies $E(K_1, M_0) = D(K_2, C_0)$. When such a pair is found, (12) should be checked by $E(K_2, E(K_1, M_i)) = C_i$, $i = 1, \ldots, l$.

If the cryptosystem is assumed to be "closed" in the sense that for any $K_1, K_2 \in \mathcal{K}$ there exists a key $K \in \mathcal{K}$ satisfying (12), this means that $\{E(K, \cdot): K \in \mathcal{K}\}$ forms a permutation group, and there are $k = 2^\kappa$ pairs (K_1, K_2) satisfying (12), and the above discussions are applied. We remark that the discussions in the first part of this section do not assume the cryptosystem to be closed. (The last expression of Section 3.4 in Kaliski *et al.* [7] is incorrect and it affects their Proposition 4.1.)

Now assume that $\{E(K, \cdot): K \in \mathcal{K}\}$ is a simple random sample without replacement from all the permutations of $\mathcal{M}$. From this sample take similarly two subsamples $\mathcal{K}_i$ of size n_i, $i = 1, 2$. The set $\{E(K_2, E(K_1, \cdot)): K_1 \in \mathcal{K}_1, K_2 \in \mathcal{K}_2\}$ contains at most $n_1 n_2$ permutations, and may be close to $n_1 n_2$. Thus the set includes a particular permutation with the probability at most $n_1 n_2/m!$, which is much smaller than $1 - q_3$. There are some regularities found in DES, yet practically

$\{E(K, \cdot): K \in \mathscr{K}\}$ can be regarded as a random sample from all the permutations of $\mathscr{M}$ [7].

4. Meet-in-the-Middle Attack; Matching Model C

There can be another type of attack, which is a sort of combination of Matching Model A and B. Assume that the forward sequences $(\tilde{H}_i^+)_{i=1}^r$ are generated as (6a) of Section 2, while the backward sequences (6b) are generated just one step, $\tilde{H}_r^- = D(\tilde{W}_{r+1}, H_{r+1})$, as in Section 3. Thus $\tilde{H}_r^+$'s are independent random variables taken from $\mathscr{M}$ with replacement, while $\tilde{H}_r^-$'s are assumed to be random samples from $\mathscr{M}$ without replacement. If n_1 $\tilde{H}_r^+$'s and n_2 $\tilde{H}_r^-$'s are generated, the number S^+ of $\tilde{H}_r^+$ which matches one of the $\tilde{H}_r^-$'s is the binomial distribution

$$\Pr[S^+ = s; n_1, n_2] = \binom{n_1}{s}\left(\frac{n_2}{m}\right)^s\left(1 - \frac{n_2}{m}\right)^{n_1 - s}.$$

The number S^- of $\tilde{H}_r^-$ which matches one of the $\tilde{H}_r^+$'s, given the number $T = t$ of different $\tilde{H}_r^+$'s, has the hypergeometric distribution

$$\Pr[S^- = s \mid T = t; n_2] = \binom{t}{s}\binom{m - t}{n_2 - s}\Big/\binom{m}{n_2}.$$

Since T has the classical occupancy distribution (3), the unconditional distribution of S^- is

$$\Pr[S^- = s; n_1, n_2] = \sum_t \binom{t}{s}\binom{m - t}{n_2 - s}\left\{{n_1 \atop t}\right\}\frac{m^{(t)}}{m^{n_1}}\Big/\binom{m}{n_2}.$$

We are just interested in the probability of the event $S^- = 0$ which is equivalent to $S^+ = 0$, and if $n_i = O(m^{1/2})$, $i = 1, 2$, as $m \to \infty$,

$$\Pr[S^+ = 0] = \Pr[S^- = 0] = \left(1 - \frac{n_2}{m}\right)^{n_1} = \exp\left\{-\frac{n_1 n_2}{m}[1 + O(m^{-1/2})]\right\}.$$

Matching Model C is effective for the following situation. A hashing scheme, based on a cipher block chaining (CBC) mode of operation, was studied by Mueller-Schloer [9]. Starting from $H_0 = 0$, and using the leading fragment of the message as the key, $K = W_0$, compute

$$H_i = E(K, H_{i-1} \oplus W_i), \qquad i = 1, 2, \ldots, l,$$

where $\oplus$ denotes exclusive-or, and the pair (K, H_l) is used as a digest of the message $(W_i)_{i=0}^l$. This is vulnerable against the following attack. Using perturbed fragments $(\tilde{W}_i)_{i=1}^r$ and $\tilde{W}_{r+1}$, generate

$$\tilde{H}_i^+ = E(K, \tilde{H}_{i-1}^+ \oplus \tilde{W}_i), \qquad i = 1, 2, \ldots, r,$$

and

$$\tilde{H}_r^- = D(K, H_{r+1}) \oplus \tilde{W}_{r+1}.$$

If a pair of $\tilde{H}_r^+$ and $\tilde{H}_r^-$ matches, the attack is accomplished. In this case, $\tilde{H}_r^+$ can be regarded as randomly chosen from $\mathscr{M}$ with replacement, but $\tilde{H}_r^-$ without replacement, if the perturbed $\tilde{W}_{r+1}$'s are different to each other.

5. The Cycling Test

Let $\mathcal{U}$ be any finite set with the cardinality $u = |\mathcal{U}|$, and let $\mathcal{F}$ be the set of all functions $f: \mathcal{U} \to \mathcal{U}$ ($|\mathcal{F}| = u^u$). We select at random one point $U_0 \in \mathcal{U}$ and one function $f \in \mathcal{F}$, and form a sequence $(U_i)_{i=0}^{\infty}$ by $U_{i+1} = f(U_i)$. Let J_1 be the least integer such that $U_{J_1} = U_i$ for some $0 \le J_1 < i$, and let J_2 be the least positive integer such that $U_{J_1+J_2} = U_{J_1}$. The sequence is determined by the leader $(U_i)_{i=0}^{J_1-1}$ and the cycle $(U_i)_{i=J_1}^{J_1+J_2}$. It is known [8] that $J := J_1 + J_2$ has the distribution

$$\Pr[J = t] = t\frac{u^{(t)}}{u^{t+1}}, \qquad 1 \le t < \infty, \tag{13}$$

and given J, J_1 is distributed uniformly on $[0, J-1]$.

In fact, the probabilistic quantities up to $U_{J_1+J_2}$ are completely equivalent to the random walk on $\mathcal{U}$: starting from U_0, each of U_i, $i = 1, 2, \ldots$, are selected independently and randomly from $\mathcal{U}$. This is further equivalent to the sequential version of the classical birthday problem in Section 1. Notice that (13) is the same as (5).

In testing DES starting from a random code $M_0 \in \mathcal{M}$, a sequence $(M_i)_{i=0}^{\infty}$ is generated by $M_{i+1} = E(g(M_i), M_i)$, where $g: \mathcal{M} \to \mathcal{K}$ is a pseudorandom function. Let $\mathcal{U}^*$ denote the space on which the sequence $(M_i)_{i=0}^{\infty}$ walks randomly. If DES is algebraically closed, then $|\mathcal{U}^*| \le |\mathcal{K}|$, while if $\{E(K, \cdot): K \in \mathcal{K}\}$ is a random sample from the all permutations of $\mathcal{M}$, then $\mathcal{U}^*$ will cover almost the whole $\mathcal{M}$ [7].

6. Quadratic Efficiency

In the literature, including [7], the phrase "birthday paradox" is used to cover the probabilistic phenomena of several situations as discussed in this note. New schemes of hashing are proposed, and more complicated attacks against them are investigated [1], [2], [9], [13]. These are related as matching models, but it is not clear whether there is a unifying probabilistic principle to explain them.

A possible definition of the phenomena is as follows. Try to tamper with a message or find a key by random trials against a cryptosystem of the size characterized by m and k, the cardinalities of the message space and the key space, respectively. As $m \to \infty$, the number n of naive random trials must be $n = \Theta(m)$ to keep the success probability $p > 0$ fixed (see [5] for the symbol Θ). If a smart attack can proceed with $n = O(m^{1/2})$, then it has an effect like the birthday paradox. Alternatively, m and k being fixed, the failure probability q decreases as $n \to \infty$ such that $-\log(q(n))$ is quadratically dependent on n rather than linearly dependent. Then, the attack is effective. A milder requirement is for $q(n)$ to be a log-concave function of n.

Questions remain as to other possible situations, their models, and a more general way of analysis.

Acknowledgment

The authors thank the referee for his helpful comments.

References

[1] S. G. Akl, On the security of compressed encodings, in *Advances in Cryptology—Proceedings of Crypto* '83, ed. D. Chaum, Plenum, New York, 1984, pp. 209–230.

[2] D. Coppersmith, Another birthday attack, in *Advances in Cryptology—Crypto* '85, ed. H. C. Williams, Lecture Notes in Computer Science, No. 218, Springer-Verlag, New York, 1985, pp. 14–17.

[3] D. W. Davies and W. L. Price, The application of digital signatures based on public key cryptosystems, *5th Int. Conf. on Comput. Commun.*, Oct. 1980, pp. 525–530.

[4] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edn., Wiley, New York, 1968.

[5] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 1989.

[6] N. L. Johnson and S. Kotz, *Urn Models and Their Applications*, Wiley, New York, 1977.

[7] B. S. Kaliski, Jr., R. L. Rivest, and A. T. Sherman, Is the Data Encryption Standard a group? (Results of cycling experiments on DES), *J. Cryptology*, Vol. 1(1) (Apr. 1988), pp. 3–36.

[8] D. E. Knuth, *The Art of Computer Programming*, Vol. 2, 2nd edn., Addison-Wesley, Reading, MA, 1981 (Exercise 3.1.12).

[9] C. Mueller-Schloer, DES-generated checksums for electronic signatures, *Cryptologia*, Vol. 7(3) (July 1983), pp. 257–273.

[10] National Bureau of Standards, *Data Encryption Standard*, Federal Information Processing Standards Publications, No. 46, U.S. Department of Commerce, 15 Jan. 1977.

[11] K. Nishimura and M. Sibuya, Occupancy with two types of balls, *Ann. Inst. Statist. Math.*, Vol. 44(1) (Mar. 1988), pp. 77–91.

[12] M. O. Rabin, Digitalized signatures, in *Foundations of Secure Computation*, ed. R. A. DeMillo *et al.*, Academic Press, New York, 1978, pp. 155–166.

[13] R. S. Winternitz, Producing a one-way hash function from DES, in *Advances in Cryptology—Proceedings of Crypto* '83, ed. D. Chaum, Plenum, New York, 1984, pp. 203–207.

[14] G. Yuval, How to swindle Rabin, *Cryptologia*, Vol. 3(3) (July 1979), pp. 187–189.