Guesswork and Variation Distance as Measures of Cipher Security

John O. Pliam

Department of Control Science & Dynamical Systems,
University of Minnesota
Minneapolis, MN 55455, U.S.A.
pliam@ima.umn.edu

Abstract. Absolute lower limits to the cost of cryptanalytic attacks are quantified, via a theory of guesswork. Conditional guesswork naturally expresses limits to known and chosen plaintext attacks. New inequalities are derived between various forms of guesswork and variation distance. The machinery thus offers a new technique for establishing the security of a cipher: When the work-factor of the optimal known or chosen plaintext attack against a cipher is bounded below by a prohibitively large number, then no practical attack against the cipher can succeed. As an example, we apply the technique to iterated cryptosystems, as the Markov property which results from an independent subkey assumption makes them particularly amenable to analysis.

1 Introduction

Research on provably secure ciphers often focuses on specific cipher properties or resistance to specific families of attacks (see e.g. [16], [9] and [17]). When general attacks are considered, the adversary's resource limitations are typically built into the equation. In the Luby-Rackoff model (see [11] or more recently [15]), the adversary is assumed to have bounded computational resources. In the *Decorrelation Theory* of Vaudenay (see e.g. [21] and the references in [22]), the adversary may have restricted data complexity (such as a bound on the number of plaintext-ciphertext pairs) or may be carrying out a constrained attack (such as *Differential Cryptanalysis* [2]). In this paper, we summarize a different approach to provable cipher security which is developed more fully in [18]. Our approach is to model a cipher as a group-valued random variable — following Shannon — and derive absolute lower limits on the work-factor for discovering its secret key.

This technique naturally applies to product ciphers and iterated cryptosystems. Figure 1 below depicts a hypothetical security profile for the behavior of the product of finitely many ciphers as a function of the number of terms. In order to begin to quantify this profile, we must find a meaningful measure of security for which establishing the profile's shape in certain places is a tractable problem. Our primary interest is in the non-asymptotic shape of the curve — because iterated cryptosystems cannot iterate forever.

Howard Heys and Carlisle Adams (Eds.): SAC'99, LNCS 1758, pp. 62–77, 2000.

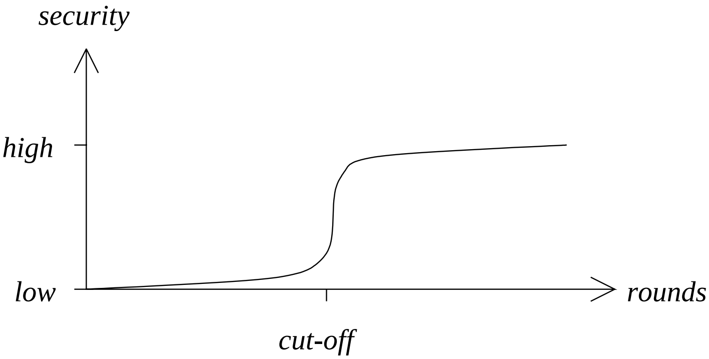

Fig. 1. A hypothetical profile of security as a function of the number of terms in a product, or equivalently the number of rounds in an iterated cryptosystems, assuming subkey independence.

We posit that a reasonable security measure is the expected work involved in "guessing" the cipher's key from the set of all keys which remain consistent with acquired plaintext-ciphertext pairs. We call this measure *guesswork* (or more generally *conditional guesswork*) and demonstrate its tractability by making use of techniques from the modern theory of random walks on symmetric structures.

Starting in Sect. 3 and continuing in Sect. 4, a formal theory of guesswork is developed which parallels information theory in a number of interesting ways. In particular, (logarithmically) tight bounds involving guesswork are derived in Theorem 1, and variation distance plays a role similar to Kullback-Leibler distance. With the help of these tools, we turn our attention in Sect. 5 to quantifying the shape of the security profile of Fig. 1 *non-asymptotically.* That is to say, rather than in some unknown neighborhood of the point at infinity, we establish provable security after a finite number of rounds.

2 Preliminaries

A basic familiarity with group theory [19], and random variables and probability spaces [8] is assumed. We develop an abstract form of Shannon's model of private key ciphers [20], in which the invertible encryption functions are taken as elements of a group G. For a block cipher with a message space $\mathscr{M}$ consisting of all n-bit strings, the group G is naturally seen as a subgroup of the symmetric group $\mathfrak{S}_{\mathscr{M}}$ (whose elements consist of all permutations of $\mathscr{M}$).

2.1 Shannon's Model

Secret keys and messages must be considered random from the viewpoint of a cryptanalytic adversary. Thus, the eavesdropper on an insecure channel may be thought of as performing a probabilistic experiment in which the message and key are values drawn at random according to certain probability distributions. It is assumed that the key is statistically independent of the message, and that the individual plaintext blocks of the message are statistically independent of one another[1]. This allows the cipher itself to be treated as an independent random variable. Furthermore, we may dispense with a distinction between the key space and the group G generated by the encryption functions. Each possible key corresponds to an element of G, and any element of G which is not identified with a key is taken to have probability 0. We may now formally define a cipher.

Definition 1. *Given a finite message space $\mathscr{M}$ and a subgroup $G \leq \mathfrak{S}_{\mathscr{M}}$, a G-**cipher** or a **cipher over** G is a G-valued random variable.*

Shannon wrote down a cipher as a linear combination of encryption functions, with the coefficients taken to be the probabilities of the corresponding functions. He naturally defined the product of ciphers by merely enforcing the distributive laws. Shannon was essentially defining what is now commonly called the *group algebra*. The natural product in the group algebra is equivalent to a kind of convolution of probability distributions.

2.2 Product Ciphers and Convolution

Consider the situation where an encryption operation is the composition of two independent encryption operations. This leads to the formal notion of the product of two ciphers over a group. Let X and Y be independent G-ciphers with probability distributions $x(g) = \mathsf{P}[X = g]$ and $y(g) = \mathsf{P}[Y = g]$. The G-cipher $Z = XY$ is called a *product cipher*, Y is called the *first component* and X is called the *second component*.

Let us examine the distribution $z(g) = \mathsf{P}[Z = g]$:

$$\begin{aligned} z(g) &= \sum_{h \in G} \mathsf{P}\left[X = gh^{-1} \mid Y = h\right] \mathsf{P}[Y = h] \\ &= \sum_{h \in G} \mathsf{P}\left[X = gh^{-1}\right] \mathsf{P}[Y = h] \;\; = \;\; \sum_{h \in G} x(gh^{-1}) y(h). \end{aligned}$$

Notice how much this last expression looks like convolution. In fact, if G were the abelian group of integers modulo n and the multiplicative notation were replaced by additive notation, $z(g)$ would literally be the *circular convolution* of the functions x and y. So, $z(g)$ is a kind of generalized convolution and will be written

$$x * y(g) = \sum_{h \in G} x(gh^{-1}) y(h). \tag{1}$$

Thus the distribution of a product is described by the convolution of the component distributions. Intuitively, convolutions "smooth out" distributions.

[1] Technically, plaintext blocks are typically independent only in the limit of large block length, but the emphasis here is on chosen plaintext attacks which are always faster than known plaintext attacks

2.3 Variation Distance

Let $\mathscr{X}$ be a finite set with probability distributions p and q defined on it. Recall ([4] or [5]) the *variation distance* between p and q defined by

$$\|p - q\| = \max_{\mathscr{Y} \subseteq \mathscr{X}} |p(\mathscr{Y}) - q(\mathscr{Y})|. \tag{2}$$

It is a standard observation that variation distance is half of the ℓ_1-norm, i.e.

$$\|p - q\| = \frac{1}{2}\|p - q\|_1,$$

and that the maximum in (2) is achieved on the set

$$\mathscr{Y} = \{x \in \mathscr{X} \mid p(x) \geq q(x)\} .$$

If X is a G-cipher and u is the uniform distribution on G, then the closer p_X is to uniformity, the harder it will be for any adversary to determine the value of X. This general statement, which holds whether or not the adversary is in possession of plaintext-ciphertext pairs, is formalized in Theorem 1 and Corollary 2 below.

3 Guesswork: The Uncertainty of Guessing

In this section, we develop the means to quantify fundamental statistical limits to the amount of work required to determine the value of a random variable. The notions of work discussed here have appeared before. In a broad sense they are intimately connected to Lorenz's theory of wealth distribution [10] (see also [12]). Massey [13] was the first to formulate, in open cryptology, what we shall call the *guesswork* of a random variable.[2] While it has correctly been pointed out (e.g. [3]) that guesswork is not a meaningful predictor of practical attack performance, we shall show that it is a very useful and tractable measure of the fundamental limits to practical attacks.

3.1 Optimal Brute-Force Attacks

Let $\mathscr{X}$ be a finite set and suppose that X is the $\mathscr{X}$-valued random variable determined by probability distribution p. We may arrange $\mathscr{X}$ so that the probabilities $p_i = p(x_i)$ satisfy

$$p_1 \geq p_2 \geq \cdots \geq p_{|\mathscr{X}|}. \tag{3}$$

[2] We resist calling guesswork "guessing entropy", as is done in [3], because Theorem 1 below is so closely analogous to Shannon's First Theorem [4] that the natural analogue of guesswork is really the expected codeword length of Shannon's theorem, as discussed in Remark 1 below. It perhaps makes more sense to call variation distance a kind of (relative) entropy, because *it* appears in the upper and lower bounds of Theorem 1, just as entropy does in Shannon's theorem. Thus to call guesswork "guessing entropy" might lead to confusion.

Many situations in cryptology and computer security force an adversary to conduct a brute-force attack in which the values of $\mathscr{X}$ are enumerated and tested for a certain success condition. The only possible luxury afforded the adversary is that he or she may know which events are more likely. For example, UNIX passwords are routinely guessed with the aid of a public-domain software package called `crack` [14], which can be configured to test the most likely passwords first. The safest bet for the cryptographer is to assume that the adversary has complete knowledge of p and will conduct any brute-force attack *optimally*, i.e. in the order given by (3). This suggests the following definitions.

Definition 2. *Let X be an $\mathscr{X}$-valued random variable whose probabilities are arranged according to (3). The* **guesswork** *of X is given by*

$$W(X) = \sum_{i=1}^{|\mathscr{X}|} i p_i .$$

The following simple algorithm demonstrates the computational meaning behind guesswork. The adversary is assumed to have access to the necessary optimal enumerator and an oracle which tells whether they have guessed correctly.

Algorithm 1. *Optimal brute-force attack against X which will always succeed and has expected time complexity $O(W(X))$.*

```
input: (i). An enumerator of the values of $\mathscr{X}$ in order of
    nonincreasing probability. (ii). An oracle which answers
    whether $X = x$.
output: The value of $X$.

for $x \in \mathscr{X}$ do
  if $X = x$ then
    return $x$.
  endif
done
```

Clearly, the average computation time of Algorithm 1 is $W(X)$. Thus, guesswork may be interpreted as the optimal expected work involved in guessing the value of a random variable.

3.2 Guesswork and Variation Distance

It is easily seen that guesswork is bounded above by

$$W(X) \leq \frac{|\mathscr{X}| + 1}{2}, \tag{4}$$

and that equality is achieved if and only if X is uniformly distributed on $\mathscr{X}$ (see [12] or [18]). The next theorem offers tight upper and lower bounds on the difference between guesswork and its maximum. The situation is analogous to

Shannon's First Theorem in which the average codeword length (the thing you want to know) is bounded above and below by expressions involving entropy (the thing you can often compute).

Theorem 1. *Let $\mathscr{X}$ be a set of size n, and let X be an $\mathscr{X}$-valued random variable defined by probability distribution p. Then,*

$$\frac{n}{2}\|p-u\| \leq \frac{n+1}{2} - W(X) \leq n\|p-u\|. \tag{5}$$

The theorem is proved in [18]. Note that when $\|p-u\|$ is sufficiently small, the upper and lower bounds of

$$\frac{n+1}{2} - n\|p-u\| \leq W(X) \leq \frac{n+1}{2} - \frac{n}{2}\|p-u\|, \tag{6}$$

are both positive. We see that as $\|p-u\| \rightarrow 0$, $W(X)$ approaches its maximum within increasingly tight bounds.

Remark 1. For small values of the variation distance, (6) admits the approximation

$$W(X) \approx \frac{n}{2}(1-\|p-u\|).$$

In this form, an analogy to Shannon's First Theorem [4] is rather apt, because the optimal expected codeword length L^* may similarly be approximated by

$$L^* \approx \log(n) - D(p\|u),$$

where $D(p\|u) = \log(n) + H(p)$ is the *Kullback-Leibler* distance to uniformity, and $H(p)$ is the *entropy* of p. Notice that $\|p-u\|$ has a supremum of 1, while $D(p\|u)$ has a maximum of $\log(n)$.

Furthermore in Shannon's theorem, the optimal codeword length is within 1 bit of the entropy. However entropy is a logarithmic quantity relative to guesswork. In that sense, (5) says that the cost of being naive in a non-optimal search is within 1 bit of the quantity $\log(n\|p-u\|)$.

4 Security Measures for Known and Chosen Plaintext Attacks

In this section we consider a cipher's capacity for resisting known and chosen plaintext attacks.

4.1 Conditional Guesswork and the Security Factors

In a known or chosen plaintext attack, a single encryption key is used to encrypt a number of different plaintexts. An adversary who observes the corresponding plaintext-ciphertext pairs is privy to partial information about the key. In this section we quantify the intuitive notion that the resilience of a cipher against known or chosen plaintexts attacks can be measured by the amount of work required to guess the key *after* information about the plaintext-ciphertext pairs has been taken into account.

Formally if G is a subgroup of $\mathfrak{S}_{\mathscr{M}}$, let X be a G-cipher representing a single choice of a cipher's key. Again let $x(g) = \mathsf{P}\left[X = g\right]$, and let $P^\ell = (P_1, P_2, \ldots, P_\ell)$ be an ℓ-tuple of i.i.d. random variables describing a sequence of *distinct* plaintexts in $\mathscr{M}$. Following [6], let $\mathscr{M}^{(\ell)}$ denote the set of ℓ-tuples with distinct elements of $\mathscr{M}$. $\mathfrak{S}_{\mathscr{M}}$ and hence G acts on $\mathscr{M}^{(\ell)}$ in the natural way, namely $\sigma(m_1, \ldots, m_\ell) = (\sigma m_1, \ldots, \sigma m_\ell)$. Now define

$$C^\ell = (XP_1, XP_2, \ldots, XP_\ell) = (C_1, C_2, \ldots, C_\ell).$$

In other words, P^ℓ and C^ℓ are $\mathscr{M}^{(\ell)}$-valued random variables. We write $p, c \in \mathscr{M}^{(\ell)}$ for instances of P^ℓ and C^ℓ. When contemplating the loss of security due to observations of plaintext-ciphertext pairs, it is natural to define, by analogy to conditional entropy, notions of conditional guesswork.

Definition 3. *Given the quantities described above, the* **conditional guesswork** *of X given C^ℓ and P^ℓ is defined as*

$$W(X|C^\ell, P^\ell) = \sum_{c,p \in \mathscr{M}^{(\ell)}} W(X|C^\ell = c, P^\ell = p) p_j(c, p).$$

The **conditional guesswork** *of X given C^ℓ and that $P^\ell = p$ is defined as*

$$W(X|C^\ell, P^\ell = p) = \sum_{c \in \mathscr{M}^{(\ell)}} W(X|C^\ell = c, P^\ell = p) p_c(c|p).$$

Here $p_j(c, p)$ is the joint distribution of C^ℓ and P^ℓ, while $p_c(c|p)$ is the conditional distribution of C^ℓ given P^ℓ. These are respectively given by

$$p_j(c, p) = \mathsf{P}\left[C^\ell = c, P^\ell = p\right], \quad \text{and} \quad p_c(c|p) = \mathsf{P}\left[C^\ell = c \mid P^\ell = p\right].$$

The two kinds of conditional guesswork will be used to quantify the performances of optimal known and chosen plaintext attacks, justifying the following definitions.

Definition 4. *The* **known plaintext security factor** *of X against the observation of ℓ plaintext-ciphertext pairs is defined as*

$$\nu_\ell(X) = W(X|C^\ell, P^\ell),$$

and the **chosen plaintext security factor** *of X against a choice of ℓ plaintext-ciphertext pairs is defined as*

$$\theta_\ell(X) = \min_{p \in \mathscr{M}^{(\ell)}} W(X|C^\ell, P^\ell = p).$$

Finally, we define $\nu_0(X) = \theta_0(X) = W(X)$.

Notice that the chosen plaintext security factor is independent of the plaintext statistics, as one would expect. The next proposition establishes $\theta_\ell(X)$ as a principal figure of merit.

Proposition 1. *$\theta_\ell(X) \leq \nu_\ell(X)$.*

Proof. Simply expand out the formulas,

$$\nu_\ell(X) = \sum_{p \in \mathscr{M}^{(\ell)}} W(X|C^\ell, P^\ell = p) p_P(p) \geq \sum_{p \in \mathscr{M}^{(\ell)}} \theta_\ell(X) p_P(p) = \theta_\ell(X),$$

where $p_P(p) = \mathsf{P}\left[P^\ell = p\right]$. □

4.2 Observing Plaintext-Ciphertext Pairs

An elementary observation about group action leads to a fundamental fact. Put simply, the guesswork $W(X|C^\ell = c, P^\ell = p)$ is entirely determined by the distribution of X on a coset of a certain subgroup of G.

Lemma 1 (Coset Work Lemma). *With X, C^ℓ, P^ℓ and $c, p \in \mathscr{M}^{(\ell)}$ defined as before, there is a $g_{cp} \in G$ such that $c = g_{cp}p$. The conditional guesswork*

$$W(X|C^\ell = c, P^\ell = p),$$

is determined by the distribution of X on the left coset $g_{cp}H$, where H is the stabilizer subgroup $\mathrm{Stab}_G(p)$*.*

Proof. The proof uses familiar group action observations discussed in [18]. Let g_{cp} be the value of X. By definition $c = g_{cp}p$, and it is standard that

$$\{g \in G \,|\, c = gp\} = g_{cp}\,\mathrm{Stab}_G(p) = g_{cp}H.$$

If $\widehat{X}$ is the random variable $(X|C^\ell = c, P^\ell = p)$, we have

$$\mathsf{P}\left[\widehat{X} = g\right] = \begin{cases} \frac{x(g)}{x(g_{cp}H)}, & \text{if } g \in g_{cp}H, \\ 0, & \text{otherwise,} \end{cases}$$

by Bayes's theorem. Now $W(\widehat{X}) = W(X|C^\ell = c, P^\ell = p)$, which completes the proof. □

The coset work lemma suggests optimal algorithms for attacking a cipher X. Given ℓ plaintext-ciphertext pairs $c, p \in \mathscr{M}^{(\ell)}$, we may restrict the optimal search for the value of X in Algorithm 1 to a coset of $H = \mathrm{Stab}_G(p)$. Thus we have the following algorithms.

Algorithm 2 (Optimal Known Plaintext Attack). *Defines algorithm* $\mathtt{kpa}_\ell$.

> *1. Collect into tuple p, ℓ random plaintexts according to their natural statistics.*
> *2. Collect into tuple c, the corresponding ciphertexts.*
> *3. Invoke Algorithm 1 to optimally search $g_{cp}\,\mathrm{Stab}_G(p)$ for the value of X, where $c = g_{cp}p$.*

Algorithm 3 (Optimal Chosen Plaintext Attack). *Defines algorithm* $\mathtt{cpa}_\ell$.

> *1. Let $\widehat{p}$ minimize $W(X|C^\ell, P^\ell = \widehat{p})$.*
> *2. Let $c = X\widehat{p}$.*
> *3. Invoke Algorithm 1 to optimally search $\widehat{g}\,\mathrm{Stab}_G(\widehat{p})$ for the value of X, where $c = \widehat{g}\widehat{p}$.*

The next proposition justifies the definitions of security factors $\nu_\ell(X)$ and $\theta_\ell(X)$. See [18] for a formal proof of this intuitive statement.

Proposition 2. *Under the assumption that the various oracles respond instantaneously, the expected computation time of attacks $\mathtt{kpa}_\ell$ and $\mathtt{cpa}_\ell$ against X are given by $\nu_\ell(X)$ and $\theta_\ell(X)$, respectively.*

4.3 Uniformly Distributed and (Conditionally) Perfect Ciphers

Ciphers for which every achievable message permutation is equally likely have extraordinary properties, making them worthy of special attention. There is one such cipher for every subgroup G of the symmetric group $\mathfrak{S}_{\mathscr{M}}$. Their security factors greatly simplify and can often be explicitly computed. When $G = \mathfrak{S}_{\mathscr{M}}$, we shall show that the resulting uniformly distributed cipher is *perfect* in meaningful ways.

Definition 5. *For any $G \leq \mathfrak{S}_{\mathscr{M}}$, the uniformly distributed G-cipher denoted U_G is called the* **uniform G-cipher**. *In case $G = \mathfrak{S}_{\mathscr{M}}$, $U_{\mathfrak{S}_{\mathscr{M}}}$ will simply be denoted U and called the* **perfect cipher**.

The coset work lemma admits an immediate simplification for uniform ciphers.

Theorem 2. *For every $p \in \mathscr{M}^{(\ell)}$,*

$$W(U_G|C^\ell, P^\ell = p) = \frac{1}{2}\left(1 + |\mathrm{Stab}_G(p)|\right).$$

Proof. For specific $c, p \in \mathscr{M}^{(\ell)}$, the coset work lemma tells us that

$$W(U_G|C^\ell = c, P^\ell = p) = W(\widehat{U_G}),$$

where $\widehat{U_G}$ is uniformly distributed on a coset of $H = \mathrm{Stab}_G(p)$. By (4),

$$W(\widehat{U_G}) = \frac{1}{2}(1 + |g_{cp}H|) = \frac{1}{2}(1 + |H|),$$

which is independent of c. The desired result follows. □

For a uniform cipher, we immediately have that the chosen plaintext security factor is a function *only* of the size of the smallest ℓ-message stabilizer.

Corollary 1. *For any* $G \leq \mathfrak{S}_{\mathscr{M}}$*,*

$$\theta_\ell(U_G) = \frac{1}{2}\left(1 + \min_{p \in \mathscr{M}^{(\ell)}} |\mathrm{Stab}_G(p)|\right).$$

For the perfect cipher U, which is uniformly distributed over the entire symmetric group, we can obtain precise formulas for $\nu_\ell(U)$ and $\theta_\ell(U)$.

Proposition 3.

$$\nu_\ell(U) = \theta_\ell(U) = \frac{1 + (|\mathscr{M}| - \ell)!}{2}.$$

Proof. For any tuple $p \in \mathscr{M}^{(\ell)}$, the stabilizer subgroup of p in $\mathfrak{S}_{\mathscr{M}}$ is the symmetric group on the remaining messages $\mathscr{M} - \{p\}$. Each stabilizer therefore has $(|\mathscr{M}| - \ell)!$ elements, and we may apply Corollary 1 to obtain $\theta_\ell(U)$. Furthermore, we have

$$\nu_\ell(U) = \sum_{p \in \mathscr{M}^{(\ell)}} W(U|C^\ell, P^\ell = p)p_P(p) = \theta_\ell(U) \sum_{p \in \mathscr{M}^{(\ell)}} p_P(p) = \theta_\ell(U).$$

□

What Proposition 3 tells us is that we can determine exactly the expected performance of the optimal known and chosen plaintext attacks $\mathtt{kpa}_\ell$ and $\mathtt{cpa}_\ell$ against a perfect cipher. Provided $\ell \ll |\mathscr{M}|$, these attacks reduce to very long brute-force searches. The addition of a new plaintext-ciphertext pair reduces the size, in bits, of the effective search space by

$$\log \theta_\ell(U) - \log \theta_{\ell+1}(U) \approx \log \left[\frac{(|\mathscr{M}| - \ell)!}{(|\mathscr{M}| - \ell - 1)!}\right] < \log |\mathscr{M}|.$$

Thus, for a cipher of block length n, $|\mathscr{M}| = 2^n$ and each new plaintext-ciphertext pair reduces the search space by no more than n bits. But by Stirling's formula

$$\log \theta_0(U) \approx \log |\mathfrak{S}_{\mathscr{M}}| \approx n2^n \text{ bits}.$$

In other words, in order to reduce the search space to within a reasonably attackable size, on the order of 2^n distinct plaintext-ciphertext pairs must be obtained. By that time the adversary has a table of all 2^n possible ciphertexts from which they can look up any desired target plaintext. One cannot expect a block cipher to perform better than this. We now explicitly prove what this discussion suggests, namely that the perfect cipher is as secure as any cipher.

Theorem 3. *For any* G*-cipher with* $G \leq \mathfrak{S}_{\mathscr{M}}$*,*

$$\nu_\ell(X) \leq \nu_\ell(U), \quad \textit{and} \quad \theta_\ell(X) \leq \theta_\ell(U).$$

Proof. Since $G \leq \mathfrak{S}_{\mathscr{M}}$, $\mathrm{Stab}_G(p) \leq \mathrm{Stab}_{\mathfrak{S}_{\mathscr{M}}}(p)$, so that any distribution on a coset of $\mathrm{Stab}_G(p)$ can be thought of as distribution on a coset of $H = \mathrm{Stab}_{\mathfrak{S}_{\mathscr{M}}}(p)$. Once again invoking the coset work lemma along with the fact that the uniform distribution is majorized by any other distribution (see [12] or [18]), we obtain

$$W(X|C^\ell = c, P^\ell = p) \leq W(U|C^\ell = c, P^\ell = p) = \frac{1+|H|}{2}.$$

The theorem is essentially proved. Pedantically, one should expand out the formulas for $\nu_\ell(X)$ and $\theta_\ell(X)$ (see [18]). □

4.4 The Security Factors and Variation Distance

Theorem 1 gave us, in terms of variation distance, tight bounds on the difference between guesswork and its maximum value. If we wish variation distance to have a deeper security meaning, it is natural to seek similar bounds on conditional guesswork.

Theorem 4. *For permutation group $G \leq \mathfrak{S}_{\mathscr{M}}$, let X be a G-cipher with probability distribution x. If $p \in \mathscr{M}^{(\ell)}$ and $H = \mathrm{Stab}_G(p)$, then*

$$W(X|C^\ell, P^\ell = p) \geq \frac{1+|H|}{2} - |G|\,\|x-u\|.$$

The proof of Theorem 4, which is presented in detail in [18], is essentially based on the decomposition of the group algebra $\mathbb{R}G$ into a direct sum of vector spaces isomorphic to the smaller group algebra $\mathbb{R}H$. This is a special case of a very important construction called the *induced representation* (see [7] and [18]).

We may bound $\theta_\ell(X)$ by an expression which is a function of the variation distance $\|x-u\|$, and as in Corollary 1, the size of the smallest ℓ-message stabilizer.

Corollary 2. *For any $G \leq \mathfrak{S}_{\mathscr{M}}$ and any G-cipher X with probability distribution x,*

$$\theta_\ell(X) \geq \frac{1}{2}\left(1 + \min_{p\in\mathscr{M}^{(\ell)}} |\mathrm{Stab}_G(p)|\right) - |G|\,\|x-u\|.$$

Proof. By definition, $\theta_\ell(X) = W(X|C^\ell, P^\ell = \widehat{p})$, where $\widehat{p}$ minimizes the conditional guesswork. Writing $\widehat{H} = \mathrm{Stab}_G(\widehat{p})$, we observe

$$\begin{aligned}\theta_\ell(X) &= W(X|C^\ell, P^\ell = \widehat{p})\\ &\geq \frac{1+|\widehat{H}|}{2} - |G|\,\|x-u\|\\ &\geq \frac{1}{2}\left(1 + \min_{p\in\mathscr{M}^{(\ell)}} |\mathrm{Stab}_G(p)|\right) - |G|\,\|x-u\|,\end{aligned}$$

which was to be proved. □

Just as the lower bound on $W(X)$ of (6) is vacuously negative unless $\|x-u\|$ is smaller than 1/2, so too the lower bound given in Corollary 2 will be negative unless $\|x-u\|$ is sufficiently small. To see when this happens, let H be one of the smallest ℓ-message stabilizers, and rearrange the inequality of the corollary as

$$\theta_\ell(X) \geq \frac{|H|}{2}(1-2\,[G\!:\!H]\,\|x-u\|). \tag{7}$$

When represented in this way, the lower bound on $\theta_\ell(X)$ becomes meaningful only when $\|x-u\|$ is less than $1/(2[G\!:\!H])$. It is not terribly surprising that resistance to chosen plaintext attacks should come at some measurable cost.

5 Applications to Iterated Cryptosystems

5.1 Generalized Markov Ciphers and the Cut-Off Phenomenon

Under the assumption of subkey independence, an iterated cryptosystem is equivalent to the product of finitely many independent and identically distributed G-ciphers. The sequence of all such products — as the number of rounds r ranges from 0 to ∞ — defines a random walk on G whose underlying Markov chain has many important security properties.

Formally, let $(X_i)_{i=1}^{\infty}$ be an infinite sequence of i.i.d. G-ciphers, each with probability distribution $x(g) = \mathsf{P}\,[X_i = g]$. Define the sequence $(Z_r)_{r=0}^{\infty}$ of G-ciphers by $Z_0 = 1$ and $Z_r = X_r \cdots X_2 X_1$. Applying (1), we see that the distribution of Z_r is given by an r-fold convolution of x with itself

$$\mathsf{P}\,[Z_r = g] = x^{*r}(g).$$

The next result follows from Proposition 5 below.

Proposition 4. *The sequence (Z_i) is a Markov chain with state space G.*

This fact allows us to generalize the definition of a Markov cipher given by Lai, Massey and Murphy [9]. Our motivation is itself a generalization of theirs. The idea of a Markov cipher in [9] was used to model resistance to Differential Cryptanalysis as a function of the number of rounds. Similarly we seek to quantify the resistance of an iterated cryptosystem to *all* known and chosen plaintext attacks, as a function of the number of rounds.

Definition 6. *With x and (X_i) defined as above, the G-cipher $Z_r = X_r \cdots X_2 X_1$, is called the (***generalized***)* **Markov cipher** *generated by r rounds of x, and (Z_i) will be called the* **Markov chain** *generated by x.*

There are a multitude of different Markov chains resulting from the action of (Z_i) on various G-sets. The following proposition is proved in [18].

Proposition 5. *Let x be a probability distribution on a finite group G, and let (Z_i) be the Markov chain generated by x. If $\mathscr{Y}$ is a G-set, and Y_0 is an independent $\mathscr{Y}$-valued random variable, then the sequence $(Y_i = Z_i Y_0)_{i\geq 0}$, is a Markov chain on the state space $\mathscr{Y}$. The transition matrix is doubly stochastic and is completely determined by x.*

Recent decades have witnessed a renaissance in Markov chain research spearheaded by Aldous and Diaconis (see [1], [5]). An important frequent observation of this research has been that many random walks on groups and other discrete structures exhibit *cut-off phenomena*, in which there is a rapid transition from order to uniform randomness. The phenomena is often quantified in terms the variation distance $\|x^{*r} - u\|$, and in fact $1 - \|x^{*r} - u\|$ often follows a profile like the one in Fig. 1. Proofs of cut-off phenomena for special cases abound in the literature. They sometimes employ representation theoretic arguments as in [5], and they sometimes employ more probabilistic arguments as in [1].

In the next section, we explore how probabilistic arguments can establish the non-asymptotic behavior of an iterated cryptosystem. In [18] some cryptological implications of the representation theoretic approach are explored.

5.2 Strong Uniform Times

Following [1] and [5, Chap. 4], we introduce some basic definitions useful in making probabilistic arguments for establishing the behavior of $\|x^{*r} - u\|$.

Let $\mathbb{N} = \{1, 2, \ldots\}$, and take $G^{\mathbb{N}}$ to be the set of infinite sequences of elements of G. A *stopping rule* is a function

$$t : G^{\mathbb{N}} \longrightarrow \mathbb{N} \cup \{\infty\},$$

such that if $t(g_1, g_2, \ldots) = i$, then $t(\widehat{g}_1, \widehat{g}_2, \ldots) = i$ whenever $\widehat{g}_j = g_j,\ j \leq i$. If Z_i is a sequence of G-valued random variables, then the $\mathbb{N}$-valued random variable $T = t(Z_1, Z_2, \ldots)$ is called a *stopping time*. In essence, the stopping rule identifies the first time that a certain condition is met in a sequence of group elements, and the stopping time describes random fluctuations in the first occurrence of that condition.

Of course, we are interested in the evolution of Markov ciphers and their approach to uniformity. Let Z_r be the Markov cipher generated by r rounds of x. If the condition being met by a stopping time T is sufficient to guarantee uniformity of Z_r, in other words if

$$\mathsf{P}\left[Z_r = g \mid T \leq r\right] = \frac{1}{|G|}, \quad \text{for all } g \in G,$$

then T is called a *strong uniform time* (*for* x). As one might intuitively expect, the statistics of a strong uniform time can characterize the approach to uniformity of the Markov cipher.

Lemma 2 (Aldous, Diaconis [1]). *Let x be a probability distribution on a finite group G, and let T be a strong uniform time for x. Then*

$$\|x^{*r} - u\| \leq \mathsf{P}\left[T > r\right], \quad \textit{for all } r \geq 0.$$

5.3 An Example: Top-to-Random Shuffle

Consider shuffling a deck of k cards by repeatedly removing the top card and returning it to a random position in the deck. Each step can be modeled by choosing a random permutation in $\mathfrak{S}_k$ of the form $\gamma_i = (i \dots 21)$ with probability $x(\gamma_i) = 1/k$, $1 \leq i \leq k$. There is a strong uniform time for x defined in the following way. Let t_1 be a stopping rule expressing the first time that γ_k is chosen. At this point the specific card j has been moved from the top to the bottom of the deck. Let t_{k-1} be the first time after t_1 that j returns to the top of the deck. At this point each permutation of the remaining cards is equally likely. At $t_k = t_{k-1} + 1$, in other words after the j on top of the deck is placed at random within the deck, each permutation of the deck is equally likely. If (Z_i) is the Markov chain generated by x, then the random time $T = t_k(Z_1, Z_2, \dots)$ is a strong uniform time for x.

Aldous and Diaconis show in [1] that the probability $\mathsf{P}[T > r]$ is governed by the "coupon-collector's problem" and is bounded by

$$\mathsf{P}[T > k \log k + ck] \leq e^{-c}, \quad c \geq 0, \ k \geq 0.$$

Thus we have a cut-off point $r_0 = k \log k$, and

$$\|x^{*r} - u\| \leq e^{-(r - r_0)/k}, \quad r \geq r_0.$$

The simplicity of the top-to-random shuffle allows it to be implemented as an iterated cryptosystem. Consider the shuffle permutations acting on the $k = 2^n$ bit strings of length n. By considering these bit strings as binary representations of the integers $\{0, \dots, 2^n - 1\}$, the following algorithm implements one round of the shuffle.

Algorithm 4. *Defines function* `TopToRand`(n, i)*, which implements one round of the top-to-random shuffle. We assume the existence of a pseudo-random number generator (PRNG) satisfying* $1 \leq \mathtt{random}(n) \leq n$*, and uniformly distributed thereon.*

```
input: The block length $n$, and the plaintext input, represented as
    an integer $0 \leq i < 2^n$.
output: The ciphertext output.

function TopToRand($n, i$):
  $m \leftarrow$ random($2^n$).
  if $i < m$ then
    return $i - 1 \bmod m$.
  else
    return $i$.
  endif
```

Unfortunately, the previous algorithm does not achieve security within a practical number of rounds because for a reasonable block length, the cut-off point $r_0 = n2^n$ is too large. Nevertheless, the example shows that there exists a cipher with an efficient round function, and for which an explicit cut-off point can be computed. Furthermore, using (6) and (7), we can also compute explicit bounds on the guesswork and chosen plaintext security factor. For $r \geq r_0$,

$$W(Z_r) \geq \frac{2^n!}{2} \left[1 - 2e^{-(r-r_0)/2^n}\right],$$

and

$$\theta_\ell(Z_r) \geq \frac{(2^n-\ell)!}{2} \left[1 - 2^{n\ell+1} e^{-(r-r_0)/2^n}\right].$$

The lower bound on $\theta_\ell(Z_r)$ makes use of the fact that all ℓ-message stabilizers of $\mathfrak{S}_{2^n}$ have size $(2^n - \ell)!$. As soon as the quantities in brackets are positive, the lower bounds on $W(Z_r)$ and $\theta_\ell(Z_r)$ grow quickly toward intractably large quantities, forcing even the most endowed adversary to work "forever" guessing the key.

6 Conclusion

We have successfully demonstrated that inequalities involving guesswork, conditional guesswork and variation distance can be used to establish the number of rounds necessary to achieve provable security in an iterated cryptosystem. Though in the example given here, that number of rounds grows exponentially with the block length, the iterations could still be applied to a smaller message space to produce provably secure S-boxes. Ongoing research suggests that iterated cryptosystems exist in which the round function is computationally efficient and number of rounds required for provable security is a polynomial in the block length. Some caveats to this approach include:

- We assume the existence of a cryptographically strong pseudo-random function. To date such functions are based on hard open problems and bounded computational resources.
- Variation distance is relatively sensitive to small deviations away from uniformity. It may therefore prove to be overly conservative as a security measure.
- Direct application of these techniques to existing block ciphers such as DES is not expected to be fruitful because it is known that keys of nonzero probability are sparse in a large group. Furthermore, it took several decades of open research to establish (finally in [23]) the precise group for DES. Nevertheless, in the design of new ciphers, the group G is easily treated as a design parameter. Large candidate groups which are smaller than $\mathfrak{S}_{2^n}$ include various wreath products.

References

1. David J. Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93:333–348, 1986.
2. Eli Biham and Adi Shamir. *Differential Cryptanalysis of the Data Encryption Standard.* Springer-Verlag, New York, 1993.
3. Christian Cachin. *Entropy Measures and Unconditional Security in Cryptography.* PhD thesis, ETH Zürich, 1997.
4. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* John Wiley & Sons, New York, 1991.
5. Persi Diaconis. *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics, Hayward, CA, 1988.
6. John D. Dixon and Brian Mortimer. *Permutation Groups.* Springer-Verlag, New York, 1996.
7. William Fulton and Joe Harris. *Representation Theory: A First Course.* Springer-Verlag, New York, 1991.
8. G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes.* Oxford University Press, Oxford, 2nd edition, 1992.
9. Xuejia Lai, James L. Massey, and Sean Murphy. Markov ciphers and differentical cryptanalysis. In D. W. Davies, editor, *Advances in Cryptology - EUROCRYPT '91*, pages 17–38, Berlin, 1991. Springer-Verlag.
10. M. O. Lorenz. Methods of measuring concentration of wealth. *J. Amer. Statist. Assoc.*, 9:209–219, 1905.
11. Michael Luby and Charles Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM Jour. Comput.*, 75(2), 1988.
12. Albert W. Marshall and Ingram Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, San Diego, 1979.
13. James L. Massey. Guessing and entropy. *Proc. 1994 IEEE Int'l Symp. on Information Theory*, page 204, 1994.
14. Alec Muffett. *Crack Version 5.0a User Manual.*
URL: `ftp://ftp.cert.org/pub/tools/crack/`.
15. Moni Naor and Omer Reingold. On the construction of pseudorandom permutations: Luby-Rackoff revisited. *Journal of Cryptology*, 12:29–66, 1999.
16. Kaisa Nyberg and Lars Ramkilde Knudsen. Provable security against a differential attack. *Journal of Cryptology*, 8:27–37, 1995.
17. Luke O'Connor and Jovan Dj. Golić. A unified Markov approach to differential and linear cryptanalysis. In Josef Pieprzyk and Reihanah Safavi-Naini, editors, *Advances in Cryptology - ASIACRYPT '94*, pages 387–397, New York, 1994. Springer-Verlag.
18. John O. Pliam. *Ciphers and their Products: Group Theory in Private Key Cryptography.* PhD thesis, University of Minnesota, July 1999.
19. Joseph J. Rotman. *An Introduction to the Theory of Groups.* Wm. C. Brown, Dubuque, IA, 3rd edition, 1988.
20. Claude E. Shannon. Communication theory of secrecy systems. *Bell System Tech. Jour.*, 28:656–715, 1949.
21. Serge Vaudenay. Provable security for block ciphers by decorrelation. In *STACS '98*, pages 249–275, Berlin, 1998. Springer-Verlag.
22. Serge Vaudenay. The decorrelation technique, 1999.
URL: `http://www.dmi.ens.fr/~vaudenay/decorrelation.html`.
23. Ralph Wernsdorf. The one-round functions of DES generate the alternating group. In R.A. Reuppel, editor, *Advances in Cryptology - EUROCRYPT '92*, pages 99–112, Berlin, 1993. Springer-Verlag.