Convergence in Differential Distributions

Luke O'Connor*[1,2]

[1] Distributed Systems Technology Centre, Australia
[2] Information Security Research Centre, QUT, Australia

Abstract. Differential cryptanalysis is a general attack based on the notion of differences. The success of the attack is derived from the probability of a *differential*. While it has been observed that the distribution of differentials can be modeled as a Markov chain, there have been few analyses that take advantage of this observation because of the prohibitive computations involved. In this paper we apply the Markov approach to the differentially 2-uniform mappings, and show that they converge exponentially fast with high probability.

1 Introduction

Differential cryptanalysis is a general attack based on the distribution of differences in a cipher [3, 2]. The notion of difference can be defined arbitrarily, but in this paper we will assume it to mean the XOR (exclusive-or) of two binary strings. The probability of an r-round differential $\Delta P, \Delta C_r$ is the probability that a pair of plaintexts of difference ΔP have a ciphertext difference of ΔC_r after r-rounds. A cipher is called *iterated* if there is a function $\mathbf{F}$, the round function, such that the cipher operates by applying $\mathbf{F}$ repeatedly. Lai, Massey and Murphy [10] have observed that it is possible in some ciphers to model the distribution of differentials in an iterated cipher as a homogeneous Markov chain $\mathbf{P}$ when the subkeys are assumed to be independent. The states of the chain correspond to the set of nonzero differences. Such ciphers are called *Markov ciphers*, examples of which include DES [14] and IDEA [9]. If $\mathbf{P}^{(r)} = [P_{ij}^{(r)}]$ is the rth power of $\mathbf{P}$, then the probability of the differential $\Delta P = i, \Delta C_r = j$ is given as $P_{ij}^{(r)}$. If it can be shown that all entries of $\mathbf{P}^{(r)}$ are tending to some small value ϵ as r becomes large, this is taken as *strong evidence* that product ciphers built from the round function $\mathbf{F}$ will be resistant to differential cryptanalysis.

A typical analysis of a Markov chain would then proceed to classify the states so as to determine the asymptotic behaviour of $\mathbf{P}^{(r)}$. State classification is usually performed by inspection but this is not possible when $\mathbf{P}$ is large, as is the case for DES and IDEA with 2^{128} entries each. However, some general properties of the $\mathbf{P}$ matrix suggest an approach to approximate its asymptotic

* The work reported in this paper has been funded in part by the Cooperative Research Centres program through the Department of the Prime Minister and Cabinet of Australia. Correspondence should be sent to DSTC, ITE Building, QUT GP, GPO Box 2434, Brisbane Q 4001, Australia. Email: oconnor@dstc.edu.au.

L.C. Guillou and J.-J. Quisquater (Eds.): Advances in Cryptology - EUROCRYPT '95, LNCS 921, pp. 13-23, 1995.

behaviour. It is known [9] that $\mathbf{P}$ is doubly stochastic when $\mathbf{F}$ is bijective, and $\mathbf{P}^{(r)}$ tends to the uniform distribution when $\mathbf{P}$ is *ergodic* (defined below). This means that if $\mathbf{P}$ could be shown to be ergodic, all entries of $\mathbf{P}^{(r)}$ would be tending towards the value $1/2^n$ where n is the block size. We are then confronted with the following two problems: (a) demonstrate that $\mathbf{P}$ is ergodic, and (b) determine the rate at which $\mathbf{P}$ approaches the unifrom distribution if it is ergodic. Note that (a) will determine if the differences are distributed uniformly, while (b) will determine an appropriate number of rounds for the cipher. Hornauer, Stephan, and Wernsdorf [8] were able to prove that certain round functions $\mathbf{F}$ yield ergodic transition matrices $\mathbf{P}$ by examining the group of mappings that could be formed by iterating the round function. Results for more general chains are reported by O'Connor and Golić [13, 12] where a combinatorial approach is used to show that if $\mathbf{F}$ is selected uniformly then $\mathbf{P}$ is ergodic with probability tending to one. On the other hand, there are no known results related to the rate of convergence of specific chains. However, Lai [9] has performed experiments on 'mini' verions of IDEA (8-bit block length) and shown the convergence to be rapid.

1.1 Results

In this paper we examine round functions that are differentially 2-uniform [11], meaning that the XOR table entries for nonzero input differences are either zero or two. As we will show, the answer to (a) depends on the *density* of nonzero entries in $\mathbf{P}$, while (b) depends on the *magnitude* of the nonzero entries in $\mathbf{P}$, both of which are known when $\mathbf{F}$ is differentially 2-uniform. Our main result is then to show that transition matrices $\mathbf{P}$ derived from differentially 2-uniform mappings $\mathbf{F}$ are ergodic with high probability and are expected to converge exponentially fast to the uniform distribution. This is proven by bounding the second largest eigenvalue of $\mathbf{P}$.

The paper proceeds as follows. In section 2 we review concepts related to finite Markov chains, and show that differentially 2-uniform mappings are highly likely to have ergodic transition matrices $\mathbf{P}$. In section 3 we introduce the concept of a rapidly mixing Markov chain, and in section 3.1 show that $\mathbf{P}$ rapidly approaches the uniform distribution with high probability.

Throughout the paper we will use bold notation to refer to objects related to the round function $\mathbf{F}$, such as $\mathbf{P}$ and $\mathbf{G}$. As many definitions and concepts will apply to all Markov chains we will use $P = [p_{ij}]$ to denote an arbitrary N-state chain, referring to it generally using normal (not bold) notation.

2 Finite Markov Chains

General definitions of Markov chains can be found in Feller [6], but we will review some concepts briefly. A chain $P = [p_{ij}]$ is *ergodic* if it is finite, aperiodic and irreducible. A sufficient condition for aperiodicity of an N-state chain P is that $p_{ii} > 0$ for some i, $1 \leq i \leq N$, while P is irreducible if for all i, j there exists

an r such that $p_{ij}^{(r)} > 0$, $1 \leq i, j \leq N$. If P is ergodic then there exists a unique distribution $\Pi = (\pi_1, \pi_2, \ldots, \pi_N)$ such that

$$\pi_j = \lim_{r \to \infty} p_{ij}^{(r)}. \tag{1}$$

The distribution Π is said to be the *limiting distribution* for P and is known to be the uniform distribution for P that are doubly stochastic.

Let $\mathbf{F} : Z_2^n \to Z_2^n$, a bijective mapping, be the round function of an r-round iterated cipher E, such that at round i, the round mapping is $C_{i+1} = \mathbf{F}(C_i + K_i)$ where C_i is the ciphertext at round i and K_i is the subkey at round i, $1 \leq i \leq r$. It can be verified that E is a Markov cipher when '+' denotes XOR and the K_i are assumed to be independent. The differential transition matrix $\mathbf{P} = [P_{ij}]$ is obtained from the XOR table of $\mathbf{F}$ as follows. For each input difference $\Delta X = i$ and output difference $\Delta Y = j$, $1 \leq i, j \leq 2^n - 1$, P_{ij} is defined as

$$P_{ij} = 2^{-n} \cdot \sum_{\substack{X, X' \in Z_2^n \\ \Delta X = X + X'}} [\mathbf{F}(X) + \mathbf{F}(X') = \Delta Y] \tag{2}$$

where $[\cdot]$ is a boolean predicate evaluating to 0 or 1. Then $\mathbf{P} = [P_{ij}]$ is an $N \times N$ matrix where $N = 2^n - 1$ since the degenerate cases where $i = 0$ or $j = 0$ are excluded. The transition matrix $\mathbf{P}$ is doubly stochastic as $\mathbf{F}$ is bijective [9].

Since $\mathbf{P}$ is clearly finite, to prove ergodicity we must demonstrate that the chain is aperiodic and irreducible. Note that $\mathbf{P}$ would be ergodic if all N^2 entries were nonzero since $P_{ii} > 0$ implies that $\mathbf{P}$ is aperiodic, and irreducibility follows trivially as $P_{ij} > 0$ for all i, j. We will argue that when the number of nonzero entries in $\mathbf{P}$ exceeds some bound $B < N^2$, then $\mathbf{P}$ is ergodic with high probability. In particular, using results from random graph theory, we will show that B is approximately $N \log N$.

2.1 Differentially uniform mappings

A mapping $\mathbf{F}$ is differentially δ-uniform [11] if each entry in the XOR table for $\mathbf{F}$ is at most δ. Since each entry in an XOR table is even, a differentially 2-uniform mapping $\mathbf{F}$ has an XOR table that consists entirely of zeros and twos, with exactly 2^{n-1} nonzero entries in each row of the table. Thus the XOR table for differentially 2-uniform mappings has the maximum number of nonzero entries for a bijective mapping, as does the corresponding transition matrix $\mathbf{P}$, which is $(2^n - 1) \cdot 2^{n-1} = N(N/2 + 1/2)$.

Example 1. Let $\rho : GF(2^3) \to GF(2^3)$ be a bijective mapping defined as $\rho(x) = x^3 \bmod f(x)$ where $f(x) = x^3 + x + 1$ is irreducible over $GF(2)$. The mapping ρ then corresponds to $0 \to 0, 1 \to 1, 2 \to 3, 3 \to 4, 4 \to 5, 5 \to 6, 6 \to 7, 7 \to 2$, and is known to be differentially 2-uniform [11]. The XOR table for ρ and the corresponding transition matrix $\mathbf{P}$ are then

$$XOR_\rho = \begin{bmatrix} 8&0&0&0&0&0&0&0 \\ 0&2&0&2&0&2&0&2 \\ 0&0&2&2&2&2&0&0 \\ 0&2&2&0&2&0&0&2 \\ 0&0&0&0&2&2&2&2 \\ 0&2&0&2&2&0&2&0 \\ 0&0&2&2&0&0&2&2 \\ 0&2&2&0&0&2&2&0 \end{bmatrix} \quad \mathbf{P} = \frac{1}{8} \cdot \begin{bmatrix} 2&0&2&0&2&0&2 \\ 0&2&2&2&2&0&0 \\ 2&2&0&2&0&0&2 \\ 0&0&0&2&2&2&2 \\ 2&0&2&2&0&2&0 \\ 0&2&2&0&0&2&2 \\ 2&2&0&0&2&2&0 \end{bmatrix}. \tag{3}$$

Note that $\mathbf{P}$ *is obtained by deleting the first row and column of* XOR_ρ *and dividing by 8. Observe that* $\mathbf{P}$ *is aperiodic since* $P_{11} > 0$. □

2.2 Random Graph Theory

As has been observed by many authors, a transition matrix P can be considered as the adjacency matrix for a directed graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_N\}$ and there is a directed edge from v_i to v_j if and only if $p_{ij} > 0$. We will call G the *underlying graph* of P. A directed graph G is strongly connected if for all v_i, v_j, there is a directed path from vertex v_i to vertex v_j. Also, an edge in G is called a loop if it connects a vertex to itself.

Lemma 1. The matrix P is irreducible if and only if G is strongly connected. The matrix P is ergodic if G is strongly connected and has a loop.

Proof. If vertices v_i and v_j in G are connected by the path $v_i v_1' v_2' \cdots v_r' v_j$ then by construction $p_{ij}^{(r)} > 0$. When G is strongly connected this is true for all vertex pairs v_i, v_j which implies that $p_{ij}^{(r)} > 0$ for some r, and P must be irreducible. Further, if G has a loop then there must exist an i for which $p_{ii} > 0$, and hence P is aperiodic. The lemma now follows. □

Example 2. The underlying graph $\mathbf{G}$ *corresponding to the* $\mathbf{P}$ *matrix in (3) is shown in Figure 1. It can be verified that* $\mathbf{G}$ *is strongly connected, and since* $\mathbf{G}$ *contains 4 loops, it follows that* $\mathbf{P}$ *is ergodic.* □

Clearly the probability of G having a loop and being strongly connected increases as the number of edges in G increases. We will say that almost all graphs with N vertices and m edges possess a certain property if the fraction of graphs with the property tends to one when $N \to \infty$. Using simple combinatorial arguments, it can be shown that almost all directed graphs G with N vertices and m edges have a loop when $m = N \cdot \gamma_N$, $\gamma_N \to \infty$. Further, Palásti [15] has shown that almost all directed graphs G with m edges are strongly connected when $m = N(\log N + \gamma_N)$, $\gamma_N \to \infty$. Here γ_N is any function that tends to infinity as N does such as $\log \log N$.

Theorem 2. Let $m = N(\log N + \gamma_N)$ where $\gamma_N \to \infty$. Then almost all directed graphs G have a loop and are strongly connected.

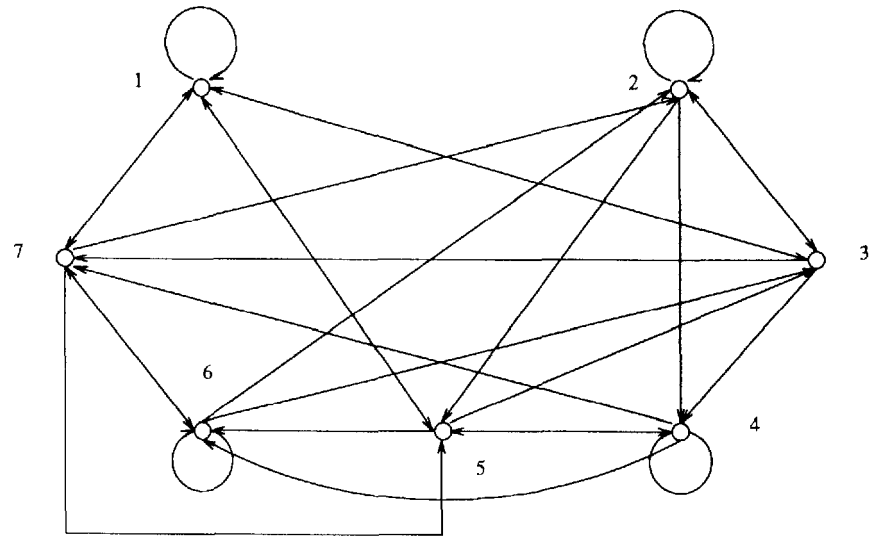

Fig. 1. The underlying graph of ρ, $0 \rightarrow 0,1 \rightarrow 1,2 \rightarrow 3,3 \rightarrow 4,4 \rightarrow 5,5 \rightarrow 6,6 \rightarrow 7,7 \rightarrow 2$,

Proof. For any two events α_1, α_2, $\Pr(\alpha_1, \alpha_2) \geq \Pr(\alpha_1) + \Pr(\alpha_2) - 1$. If the events of interest are 'G has a loop' and 'G is strongly connected', then the joint probability tends to one when $m = N(\log N + \gamma_N)$ and $\gamma_N \rightarrow \infty$. □

Recall that the **P** matrix associated with a differentially 2-uniform mapping $\mathbf{F} : Z_2^n \rightarrow Z_2^n$ has $(2^n - 1) \cdot 2^{n-1} = N(N/2 + 1/2) > N^2/2$ nonzero entries. Theorem 2 states that a transition matrix with at least $N \log N$ *randomly distributed* nonzero edges is ergodic with high probability. Even though the nonzero entries of **P** are *not* randomly distributed, it is still highly likely that **P** is ergodic since the number of edges in **G** exceeds the required $N \log N$ bound of Theorem 2. We therefore make the following assumption

Proposition 2.1 Let $\mathbf{F} : Z_2^n \rightarrow Z_2^n$ be a bijective differentially 2-uniform round function. Then the transition matrix **P** derived from **F** is assumed to be ergodic. □

To support this assumption we have verified that from the 40,320 bijective mappings $\mathbf{F} : Z_2^3 \rightarrow Z_2^3$, 10,752 are differentially 2-uniform, and each has an ergodic transition matrix **P**. There are only 7 distinct transition matrices **P**, with 1536 mappings **F** yielding the same **P** matrix. (We note that 3 is the smallest n for which differentially 2-uniform bijective mappings exist).

3 Convergence and Rapid Mixing

Let $P = [p_{ij}]$ be ergodic with $\Pi = (\pi_1, \pi_2, \ldots, \pi_N)$ its limiting distribution. Also let $P_i^{(r)}$ be the distribution of the states after r steps when started in state i. The *variation* [5] between $P_i^{(r)}$ and Π is defined as

$$\| P_i^{(r)} - \Pi \| = \frac{1}{2} \cdot \sum_{j=1}^{N} |p_{ij}^{(r)} - \pi_j| = \frac{1}{2} \cdot \sum_{j=1}^{N} e_{ij}^{(r)}. \tag{4}$$

A chain is *rapidly mixing* if the error terms $e_{ij}^{(r)}$ in (4) converge to zero as a fast function of r (see [16] for a survey). Initially the results related to rapid mixing only applied to special chains, such as those that were *time reversible*. A chain P is time reversible if $\pi_i p_{ij} = \pi_j p_{ji}$ for all states i, j. If P is nonreversible then define $M(P) = P \cdot \tilde{P}$ where $\tilde{P} = [\tilde{p}_{ij}]$ and $\tilde{p}_{ij} = \pi_j p_{ji} / \pi_i$. It can be shown [7] that $M(P)$ is time reversible, is ergodic if P is ergodic, and has limiting distribution Π if Π is the limiting distribution of both P and $\tilde{P}$. The usefulness of $M(P)$ is that it is possible to bound $\| P_i^{(r)} - \Pi \|$ by considering its eigenvalues.

For an ergodic chain P the Perron-Frobenius theorem [1] states that the largest eigenvalue is 1 while all other eigenvalues are less than 1 in modulus. In particular, the N eigenvalues of $M(P)$ are real and nonnegative [7]. Consequently, the convergence of the chain is determined by the magnitude of the second largest eigenvalue. Let the N eigenvalues of $M(P)$ be $1 = \beta_1 > \beta_2 \geq \cdots \geq \beta_N > 0$.

Theorem 3 Fill [7]. For any state i, $4\pi_i \cdot \| P_i^{(r)} - \Pi \|^2 \leq (\beta_2)^r$. □

There are several methods to bound β_2 when $M(P)$ is time reversible, and the method we will use is based on the Poincaré inequality [5] and canonical paths [16]. One result from the investigation of rapidly mixing Markov chains has been to show that the convergence of the chain P depends on the geometric properties of G_M, the underlying graph of $M(P)$. Let $M(P) = [q_{ij}]$, and from the time reversible property define

$$d_{ij} \stackrel{\text{def}}{=} \pi_i q_{ij} = \pi_j q_{ji}. \tag{5}$$

For $G_M = (V, E)$, let $\delta(v_i, v_j)$ be a (directed) path between vertices v_i and v_j with no repeated edges. Let Γ be a collection of paths $\delta(v_i, v_j)$ containing one path for each vertex pair v_i, v_j in G_M. At least one such path set Γ will exist since P, and hence $M(P)$, is irreducible. Now define the length of the path $\delta(v_i, v_j)$ to be

$$|\delta(v_i, v_j)| \stackrel{\text{def}}{=} \sum_{e : e \in \delta(v_i, v_j)} d(e)^{-1} \tag{6}$$

where the sum is over all edges e in the path $\delta(v_i, v_j)$ and $d(e) = d_{ij}$. Finally define κ as

$$\kappa \stackrel{\text{def}}{=} \kappa(\Gamma) = \max_{e} \sum_{\delta(v_r, v_s) : e \in \delta(v_i, v_j)} \pi_i \pi_j \cdot |\delta(v_i, v_j)| \tag{7}$$

where the maximum is taken over all directed edges in the graph and the sum is over all paths that traverse the edge e. Note that κ is essentially a measure of 'bottlenecks'. A bottleneck S in a graph G is a set of vertices for which there are relatively few edges directed in or out of S as compared to $|S|$. Intuitively, if the chain enters a state corresponding to a vertex in S then the process gets 'stuck' in S and does not mix rapidly. Consequently, in any path set Γ, the edges joining S to the rest of the graph will be traversed frequently. We are now ready to state the Poincaré inequality.

Proposition 3.1 (Poincaré inequality) For an ergodic time reversible chain

$$\beta_2 \leq 1 - \frac{1}{\kappa}.$$

□

Since in general the transitions matrices $\mathbf{P}$ describing differential distributions are not time reversible, our goal is to show that the convergence of $\mathbf{P}$ can be bound using Theorem 3 and the Poincaré inequality.

3.1 Bounding eigenvalues

We begin by showing that the $M(\mathbf{P})$ matrix derived from a differentially 2-uniform mapping $\mathbf{F}$ has the *complete graph* (all vertex pairs connected by an edge) as its underlying graph.

Lemma 4. $M(\mathbf{P}) = \mathbf{P}\mathbf{P}^T$.

Proof. Recall that $M(\mathbf{P}) = \mathbf{P}\tilde{\mathbf{P}}$ is dervied from $\mathbf{P} = [P_{ij}]$ by defining $\tilde{\mathbf{P}} = [\tilde{P}_{ij}]$ where $\tilde{P}_{ij} = \pi_j P_{ji}/\pi_i$. But since $\pi_i = \pi_j$ for all states i, j it follows that $\tilde{\mathbf{P}} = \mathbf{P}^T$, the transpose of $\mathbf{P}$. □

Corollary 3.1 $M(\mathbf{P})$ has no zero entries. Equivalently, the underlying graph $\mathbf{G}_M$ of $M(\mathbf{P})$ is complete.

Proof. Let $M(\mathbf{P}) = [q_{ij}]$ and observe that

$$q_{ij} = \sum_{k=1}^{N} P_{ik} \cdot \tilde{P}_{kj} = \sum_{k=1}^{N} P_{ik} \cdot P_{jk}. \tag{8}$$

But by construction, each row $P_{i1}, P_{i2}, \ldots, P_{iN}$ of $\mathbf{P}$ has $2^{n-1} = N/2 + 1/2$ nonzero entries. Since the majority of entries in each row are nonzero, there must be at least one k in (8) for which $P_{ik} \cdot \tilde{P}_{kj} > 0$, implying $q_{ij} > 0$. Since this is true for all pairs of states i, j, it follows that $M(\mathbf{P})$ has no nonzero entries. □

Now consider applying the Poincaré inequality to bound the variation in $\mathbf{P}$, requiring a path set Γ for $\mathbf{G}_M$. But since $\mathbf{G}_M$ is complete we simply take the path between v_i and v_j to be the directed edge connecting these vertices so that

$\delta(v_i, v_j) = e_{ij}$. Obviously each edge is used only once in the path set which implies that the summation on the RHS of (7), the equation defining κ, will only have one term. To determine κ we need only determine the length $|\delta(v_i, v_j)|$ of each path (i.e. edge).

Recall that the length of an edge $e = e_{ij}$ is the inverse of $d_{ij} = \pi_i q_{ij}$ where $1/\pi_i = N$ and q_{ij} is given as in (8). Since $P_{ij} = 2/2^n$ if $P_{ij} > 0$ then q_{ij} can be written as

$$q_{ij} = \frac{4}{2^{2n}} \cdot \sum_{k=1}^{N} [P_{ik} > 0] \cdot [P_{jk} > 0] \tag{9}$$

where $[\cdot]$ is a boolean predicate evaluating to either 1 or 0. When $i = j$ the value of the summation in (9) is $2^{n-1} = N + 1/2$. However when $i \neq j$, we will model the summation as a random variable α_e distributed binomially as $b(p, N)$, where $p = (\frac{1}{2} + \frac{1}{2N})^2$ is the probability that two rows from $\mathbf{P}$ are nonzero in the same entry. There are $N^2 - N$ random variables α_e to consider, corresponding to q_{ij}, $i \neq j$.

Given the restrictions on $M(\mathbf{P})$, the criterion (7) for κ reduces to

$$\kappa = \kappa(\Gamma) = \max_e \frac{1}{N^2} \cdot \frac{N \cdot 2^{2n}}{4 \cdot \alpha_e} = \max_e \frac{(N+1)^2}{4N \cdot \alpha_e}. \tag{10}$$

So the maximum is obtained when α_e is minimized. Observe that $\beta_2 \leq 2^{-t}$ when $\kappa = \frac{2^t}{2^t - 1}$, implying that there is some α_e for which

$$\alpha_e = \frac{2^t - 1}{2^t} \cdot \frac{(N+1)^2}{4N} = (1 - 2^t) \cdot pN. \tag{11}$$

It follows that a good bound on β_2 is obtained if the smallest α_e is just slightly less than the mean pN of its distribution $b(p, N)$.

Example 3. For the $\mathbf{P}$ *matrix defined in (3), it can be verified that* $M(\mathbf{P})$ *is a* 7×7 *matrix with 1 on the main diagonal and* $1/8$ *elsewhere. In this case* $pN = 16/49$, *and* $\alpha_e = (1 - 2^{-3}) \cdot pN$ *for all* q_{ij}, $i \neq j$, *implying that* $\| P_i^{(r)} - \Pi \|$ *is the same value for all* i. *The Poincaré inequality states that* $\beta_2 \leq 2^{-3}$, *and Table 1 shows that* $(\beta_2)^r = 2^{-3r}$ *bounds* $4/7 \cdot \| P_i^{(r)} - \Pi \|$ *as predicted from Theorem 3.*

□

For larger n we will not be able to compute $M(\mathbf{P})$ explicitly as we did in the example above, and some probabilistic statement must be made. The next lemma (adapted from [4]) gives a bound on the probability that $|\alpha_e - pN| \geq 2^{-t} \cdot pN$.

Lemma 5. If $0 < p < \frac{1}{2}$, and $(pN)^{-\frac{1}{2}} < \epsilon < \frac{1}{6}$, then

$$\Pr(\, |\alpha_e - pN| \geq \epsilon pN) < e^{-\frac{\epsilon^2 pN}{3(1-p)} + \frac{\epsilon}{1-p}} \tag{12}$$

where e is the base of the natural logarithm. □

r	$\| P_i^{(r)} - \Pi \|$	$4/7 \cdot \| P_i^{(r)} - \Pi \|^2$	2^{-3r}
1	0.10714	0.45918	0.875
2	0.53571	0.11479×10^{-1}	0.10937
3	0.13392×10^{-1}	0.71747×10^{-3}	0.13671×10^{-1}
4	0.66964×10^{-2}	0.17936×10^{-3}	0.17089×10^{-2}
5	0.16741×10^{-2}	0.11210×10^{-4}	0.21362×10^{-3}

Table 1. Convergence bounds for $\mathbf{P}$.

If $2^{2t} = o(N)$ and letting $p = 1/4$, the probability that all $N^2 - N$ α_e deviate from pN by less than a factor of $(1 - 2^{-t})$ is $\left(1 - e^{-\frac{N}{9 \cdot 2^{2t}} + \frac{4}{3 \cdot 2^t}}\right)^{N^2 - N}$ which reduces to

$$1 - e^{\ln N + \ln(N-1) - \frac{N}{9 \cdot 2^{2t}} + \frac{4}{3 \cdot 2^t}} + O\left(e^{2 \ln N - \frac{2N}{2^{2t}}}\right). \tag{13}$$

Using (13) we are able to argue convergence results for ciphers with large given parameters such as block size and number of rounds.

Example 4. Consider a 16-round cipher of block size $n = 64$, or $N = 2^{64} - 1$, for which we wish to show that $\beta_2 \leq \frac{1}{32}$. In this case $t = 5$ and (13) bounds the deviation probability as greater than $1 - e^{128 - 2^{50}}$ which for all practical purposes is 1. Then if at least one α_e is less than the mean, Theorem 3 states that

$$4\pi_i \cdot \| P_i^{(16)} - \Pi \|^2 \leq (\beta_2)^{16} < 2^{64} \cdot 2^{-5 \cdot 16} = 2^{-16}. \tag{14}$$

Since the variation has N terms, the average squared error per state is then approximately $\frac{1}{4} \cdot 2^{-64} \cdot 2^{-16} = 2^{-82}$, giving the average error value $e_{ij}^{(16)}$ to be 2^{-41}. □

4 Conclusion

The main aim of the paper was to show that bounds on the convergence of Markov chains describing differential distributions can be obtained for differential 2-uniform mappings. The convergence is expected to be exponential since there is a large separation between β_1 and β_2, the two largest eigenvalues of the $M(\mathbf{P})$ matrix. This separation is due to the fact that $M(\mathbf{P})$ contains no zero entries for a differential 2-uniform mapping.

The analysis has shown that the density of zero entries in the XOR table for the round function $\mathbf{F}$ will determine if it is ergodic and also the rate of convergence, since $M(\mathbf{P})$ will mostly be nonzero if $\mathbf{P}$ is mostly nonzero. It is then possible to extend the analysis to mappings other than those that are differentially 2-uniform, if the density of zeros can be approximated and a bound is known on the largest entry in the XOR table so that α_e can be approximated.

However, we conjecture that the rate of convergence in differential 2-uniform mappings is optimal (fastest) given that $M(\mathbf{P})$ is totally nonzero and all nonzero entries in $\mathbf{P}$ are bounded by the constant 2.

Strictly speaking, the convergence result states that the probability of a differential can be made arbitrarily close to $1/N = \frac{1}{2^n-1}$. However since each nonzero XOR table entry is even, the lowest nonzero probability of a differential is $\frac{2}{2^n} = \frac{1}{2^{n-1}}$. The discrepancy is introduced by modeling differentials using Markov chains. In practice, if the probability of the most likely differential is at most $3/2^n$, then all 2^n plaintext-ciphertext pairs must be examined, which renders key search redundant [9]. A rapidly converging Markov chain for differential cryptanalysis strongly suggests that all differentials will have a probability approaching the practical minimum.

Acknowledgement

I would like to thank Jovan Golić for many helpful discussions.

References

1. U. Bhat. *Elements in applied stochastic processes.* John Wiley and Sons, 1972.
2. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.
3. E. Biham and A. Shamir. *Differential cryptanalysis of Data Encryption Standard.* Springer–Verlag, 1993.
4. B. Bollobás. *Random graphs.* Academic Press, 1985.
5. P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Annals of Applied Probability*, 1(1):37–61, 1991.
6. W. Feller. *An Introduction to Probability Theory and its Applications.* New York: Wiley, 3rd edition, Volume 1, 1968.
7. J. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Annals of Applied Probability*, 1(1):62–87, 1991.
8. G. Hornauer, W. Stephan, and R. Wernsdorf. Markov ciphers and alternating groups. *Advances in Cryptology, EUROCRYPT 93, Lecture Notes in Computer Science, vol. 765, T. Helleseth ed., Springer-Verlag*, pages 453–460, 1994.
9. X. Lai. *On the design and security of block ciphers.* ETH Series in Information Processing, editor J. Massey, Hartung-Gorre Verlag Konstanz, 1992.
10. X. Lai, J. Massey, and S. Murphy. Markov ciphers and differential analysis. In *Advances in Cryptology, EUROCRYPT 91, Lecture Notes in Computer Science, vol. 547, D. W. Davies ed., Springer-Verlag*, pages 17–38, 1991.
11. K. Nyberg. Differentially uniform mappings for cryptography. *Advances in Cryptology, EUROCRYPT 93, Lecture Notes in Computer Science, vol. 765, T. Helleseth ed., Springer-Verlag*, pages 55–64, 1994.
12. L. J. O'Connor. Designing product ciphers using Markov chains. *proceedings of the Workshop on Selected Areas in Cryptography, Kingston, Canada, May 1994*, pages 2–13, 1994.

13. L. J. O'Connor and J. Dj Golić. A unified markov approach to differential and linear cryptanalysis. to be presented at Asiacrypt, November 1994.
14. National Bureau of Standards. Data Encryption Standard. FIPS PUB 46, Washington, D. C. (January 1977).
15. I. Palásti. On the strong connectedness of random graphs. *Studia Sci. Math. Hungar.*, 1:205–214, 1966.
16. U. Vazirani. Rapidly mixing Markov chains. In B. Bollobás, editor, *Probabilistic combinatorics and its applications, proceedings of Symposia in Applied Mathematics, volume 44*, pages 99–121, 1991.