STATISTICAL CRYPTANALYSIS ON BLOCK CIPHER

BIMAL ROY AND SOURAV MUKHOPADHYAY
Applied Statistics Unit, Indian Statistical Institute, Kolkata,
WB 700 108, INDIA
e-mails: bimal@isical.ac.in, sourav_t@isical.ac.in

Abstract. The aim of this paper is to present the development of statistical cryptanalysis and design of block cipher. A few design principles are recalled before we go over to describe some of the basic statistical cryptanalysis. Finally, we review some cryptanalysis techniques, which are natural extensions of some of the basic statistical attacks.

Keywords : Block cipher, Cryptanalysis, Encryption standard, Statistical attack.

1 Introduction

Cryptography is the science or art of secret writing. The fundamental objective of cryptography is to enable two people to communicate over an insecure channel (a telephone line or a computer network for example) in such a way that an opponent can not understand what is being communicated. The "plaintext" is transformed to "ciphertext" by means of an "encryption" function and a "secret key". The ciphertext is communicated and the receiver recovers the plaintext by using a "decryption" function. Study of cryptography concentrates on designing "secured" encryption and decryption function. The basic mathematical tools used are Algebra, Number Theory, Combinatories etc. Cryptanalysis (popularly known as code breaking) is the other side of the coin. It is assumed that ciphertexts and the model for the encryption is known to the attackers. In addition some plaintexts may also be available. There are four kinds of attacks.

1. Ciphertext only attack: This is the most weakest cryptanalytic attack, since it requires only passive eavesdropping from the attacker in order to obtain the ciphertext. The knowledge of the plaintext is minimal and consists of some information about the distribution of the plaintexts. For example, an attacker may know that the encrypted plaintext is in English. Ciphers that succumb to this attack are useful examples of how not to build ciphers and as puzzles for cryptography students.

2. Known plaintext attack: This scenario assumes that the attacker knows a portion of the encrypted text. The aim is either to derive from this known portion the secret key or at least to be able to obtain some unknown portion of the message text. This scenario is still highly realistic, since it is hard to prevent the attacker from guessing part of the plaintext (something that in the good old days was called a "probable word method").

3. Chosen plaintext attack: In this case one assumes that the attacker has the ability to encrypt a text of his choice. In practice this can be achieved in

the case when an encryption box with the unknown secret key falls in the hands of the attacker or when it is possible to send chosen plaintext to the owner of the secret key and then tap the transmission of this text in encrypted form to a third party. This scenario is less common since it requires active action of the attacker.

4. Chosen ciphertext attack: This is similar to the previous case, but requires an ability to choose ciphertexts for a decryption device.

In the cryptanalysis attempts are made to "estimate" the secret key. The methodologies generally used are probabilistic/statistical in nature. For more on cryptanalysis of stream cipher one may refer to Roy and Palit (2004). An attempt is made in this servey to highlight several attacks on block cipher.

2 Block cipher

There are two kinds of secret key ciphers; stream ciphers and block ciphers. In stream ciphers a long sequence of key bits are generated and exclusive or'ed (addition modulo 2) with the plaintext. In block ciphers the plaintext is divided in to blocks of a fixed length and encrypted into blocks of ciphertext using the same key. The mathematical definition of a block cipher is:

Definition: An n-bit block cipher is a function $E : V_n \times K \rightarrow V_n$ such that for each key $k \in K$, $E(p, k)$ is an invertible mapping (encryption function for k) from V_n to V_n , written $E_k(P)$. The inverse mapping is the decryption function, denoted $D_k(C)$. $C = E_k(P)$ denotes the ciphertext C that results from plaintext P under k. The variable V_n is the space containing all the possible bit strings of length n.

An n-bit block cipher with a fixed key is a permutation $p : GF(2)^n \rightarrow GF(2)^n$. It would require $log_2(2^n!)$ bits to represent the key such that all permutations p were possible, or roughly 2^n times the number of bits in a cipher block. With an ordinary block size, e.g. 64 bits, this is a much too big number for practical use, therefore the key size in a practical block cipher is much smaller, typically 128 bits or 256 bits. A good encryption function must contain some non-linear component, and this is often a substitution box or s-box. An s-box is defined as a mapping $GF(2)^n \rightarrow GF(2)^m$, usually defined by a $n \times m$ lookup table. Almost all block ciphers used today are iterated block ciphers. These ciphers are based on iterating a function several times, each iteration is called a round.

Figure 1: A typical r-round block cipher

In Figure 1, we show the process of encrypting the plaintext X_0 under a typical r-round block cipher to obtain the cipher text X_r. Here X_i denotes the intermediate value of the block after i rounds of the encryption, so that $X_i =$

$F_i(X_{i-1}, k_i)$, where $(k_1, k_2, ..., k_r)$is the list of round keys which is derived from the secret key K using a policy known as key scheduling algorithm.

The round key is derived from the cipher key by a key schedule, which is an algorithm that expands the master key or the cipher key. Key-scheduling function should be a good pseudo-random generator, however the complexity of its design is less restricted than that of the main body of the block-cipher itself. This is so since in most cases a single key is used to encrypt many blocks before it is changed and thus key-scheduling algorithm can spend more time on randomizing things than the encryption function. Due to this reasoning in many cases the analysis of key-scheduling function is hard. It is also hardly worth the effort since in most cases the flawed key-schedule can be replaced without altering the main encryption function. An attacker may assume that subkeys are independent random variables. If the cipher is broken under this assumption, no patch of key-schedule will save it. Interestingly it is possible to avoid the need for a complex key-schedule by using a fixed mixing permutation on a large set of inputs and two keys XORed at the input and at the output of the encryption function [Shannon 1949; Even and Mansour 1997], These keys are now called whitening keys. Many modern ciphers combine both the whitening and the key-scheduling approaches.

The cipher key is usually between 40 and 256 bits for a block cipher, and for an r-round iterated cipher this is expanded into r-round keys. The round function is usually a combination of substitution and transposition. Substitution is when a block in the plaintext is substituted with another block by some substitution rule. Transposition is to permute the blocks or characters in the plaintext. In earlier ciphers substitution and transposition were used on their own as a cipher, where each plain text symbol was a block, but this proved to be insecure because of the small block size. Most modern ciphers are a combination of substitution and transposition, and are often called product ciphers (Stinson, 2001).

Among the main building blocks of modern block-ciphers are substitutions and permutations, which are primitive ciphers on their own. Substitution ciphers are known from ancient times and can be viewed simply as a change of names of the letters. For example in a cipher attributed to Julius Caesar each letter of the alphabet is exchanged by a letter standing three positions from it (A is encrypted as D, B as E, C as F, etc.). Of course in general the substitution need not have a simple "shift" structure as in Caesar's cipher. However, in-spite of an astronomical number of possible substitution ciphers over the English alphabet (26!), they are easily solvable, using the letter frequency analysis. As a bright illustration of this one can read Edgar Poe's fascinating story "The Golden Bug", or Conan Doyle's "The Dancing Men". A popular element of modern ciphers- a substitution box (S-box) takes a block of m bits as its input and outputs a block of n bits (m not necessarily equals n). S-box can perform any function on a set of its inputs; if $m = n$ it can be a permutation on a set of 2^m inputs, if $m > n$ it can be a collection of several permutations on a set of 2^n inputs. It can be a randomly chosen function, or a carefully designed function with special properties. It is desirable for an S-box to perform non-linear and

non-affine function in order for the whole cipher to be a non-linear function. Linearity in cipher's behavior is the end of a cipher, since it essentially means that information is leaked from the plaintext to the ciphertext. Both expanding ($m < n$) and contracting ($m > n$) S-boxes can be met in modern block-ciphers. Unless being calculated by a compact formula the memory required to store an S-box grows exponentially with the linear increase in the size of its input m. Thus the most typical sizes for S-box input are $m = 4, 6, 8, 12$ bits. The second basic element - permutation (or transposition) cipher keeps plaintext characters as they are but arranges them in a different order. One of the oldest transposition methods was used by ancient Greeks: A leather belt is tightly wound around a cylinder and a message is written on the belt across the length of the cylinder. The belt is then worn by a messenger. The message can be decrypted by a party who has a cylinder of the same diameter as was used during the "encryption". Breaking a basic permutation cipher is an easy task, especially if one knows a part of the encrypted plain text. In modern ciphers permutations of bits are frequently used. Although weak on their own, a line of substitutions followed by a permutation has good "mixing" properties: substitutions add to local confusion and permutation "glues" them together and spreads the local confusion to the more distant sub-blocks. Shannon (1949) in a pioneering work "Communication Theory of Secrecy Systems" suggested to use several mixing layers interleaving substitutions and permutations. Such design is called substitution-permutation or an SP network (SPN). Figure 2 is an example of SPN.

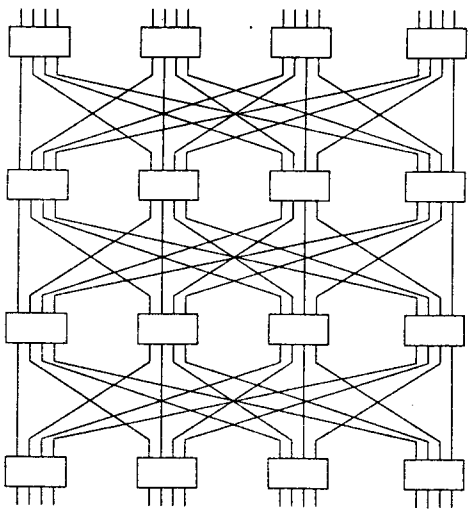

Figure 2: An example of substitution permutation network (SPN)

The Data Encryption Standard (DES) (National Bureau of Standards, 1977) has been the most widely used iterated block cipher since it was published in

1977 by National Bureau of Standards(1977) (now the National Institute of Standards and Technology, or NIST), but it is now replaced by the Advanced Encryption Standard (AES) because of too small key and block size. The DES can be seen as a special implementation of a Feistel cipher, named after Horst Feistel, where the input to each round is divided into two halves, as in the following description.

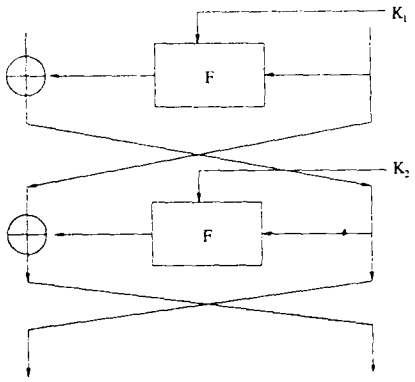

Figure 3: Two round DES

2.1 Description of DES

DES cipher is so important to the development of modern cryptanalysis that it might be worth while to describe this construction in some greater detail. It usually looks "monstrous" to the first time reader. Surprisingly almost every bit of design in DES seems to have a security reason, and most of the changes seem to weaken the cipher considerably. Biham and Shamir (1993) gave a thorough study of DES and its modifications. DES was designed by IBM crypto group from its predecessor Lucifer in early seventies and was published in the Federal Register of 17 March, 1975. DES was adopted as a standard for "unclassified" information on January 1977. Since then it became the most widely used and the most analyzed cipher. DES is an iterative block cipher. It encrypts blocks of 64 bits into ciphertext blocks of 64 bits under control of the 56-bit secret key. DES performs 16 iterations of the round function, which is called the F-function. Figure 3 shows the basic structure of DES reduced to two rounds, one can see that it is a Feistel cipher. The F-function has a relatively simple structure and is based on the substitution-permutation sandwich idea of Shannon (described above).

Each round takes the 64-bit output of the previous round, divides it into two 32-bit halves- the left half L and the right half R. The F-function (described in Figure 4) takes R as its input, expands it (by $E(R)$) from 32 bits in to 48 bits and XORs the result with the 48-bit subkey derived from the 56-bit secret key K by the key scheduling algorithm. Then the result enters eight substitution

boxes (S-boxes). Each S-box takes as input six bits and outputs four bits. The 32-bit result from the row of S-boxes is permuted by the permutation P. The permuted value is the output of the F-function. In the round function, the output of the F-function - $F(R, K_i)$ is XORed with L, and the right and the left halves are swapped. Thus, the output of the i-th round is $(R, L \oplus F(R; K_i))$. Note that the tables P, S_i, i= 1.... 8 , E are defined and fixed in the standard. so the only variable part of DES is the secret key K.

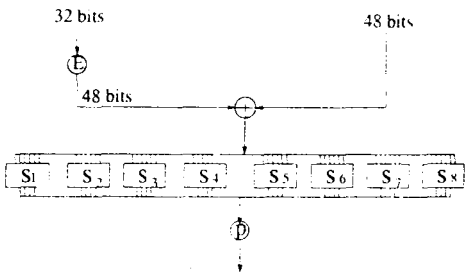

Figure 4: The F function of DES

The key scheduling algorithm of DES is as follows: The 64-bit key is permuted by the permutation $PC-1$ (Stinson, 1995, p. 75). This permutation performs two functions: strips eight parity bits and then distributes the remaining 56 bits into two 28-bit registers C and D. On each round 28-bit registers C and D are left shifted by one or two places (according to a fixed schedule). After the shift the permutation $PC-2$ (Stinson, 1995, p. 76) is performed over C and D , selecting 24 bits out of each 28-bit register. These 48 bits form the subkey of the corresponding round.

AES is the successor of DES. NIST replaced DES by the new standard which is called Advanced Encryption Standard or AES in 1997. At the "First AES candidate conference" on 1998, 15 AES candidate were selected by NIST. On 1999, five of them (MARS, RC6, Rijndael, Serpent and Twofish) were selected at the "Second AES candidate conference". Finally, Rijndael (Daemen and Rijmen, 1998) was ultimately selected as the AES by NIST (National Institute of Standards and Technology, 2001).

2.2 Description of AES

We now give a short description of AES (for details see Daemen and Rijmen, 2001). Rijndeal is a 128-bit block cipher with one of the three different key sizes, 128 or 192 or 256 bits. 128-bit block is viewed as $(b_0, b_1, ..., b_i, ..., b_{15})$, where b_i is the i'th byte of the block. The bytes are organized in a matrix form:

$$\begin{pmatrix} b_0 & b_4 & b_8 & b_{12} \\ b_1 & b_5 & b_9 & b_{13} \\ b_2 & b_6 & b_{10} & b_{14} \\ b_3 & b_7 & b_{11} & b_{15} \end{pmatrix}$$

The number of round is 10 (for key size 128) or 12 (for key size 192) or 14 (for key size 256).
The round function is composed by the following consecutive operations:
SUBBYTES: An S box is applied to each byte of the data (16 times in parallel).
SHIFTROWS: Perform a permutation to change the order of bytes in the data.
MIXCOLUMNS: Every 4 consecute bytes (column) are mixed by a linear operation.
ADDROUNDKEY: The data is XORed with a 128-bit subkey.
For the details of the above operation see Stinson (2001, pp. 103-107).
S-box of Rijndael can be defined algebraically (Algebraic formulation involves operation in a finite field (Lidl and Niederreiter, 1994). S-box is taking the multiplicative inverse of the input in $GF(2^8)$ (modulo the irreducible polynomial of Rijndael $x^8 + x^4 + x^3 + x + 1$).

3 Statistical Attack on Block cipher

3.1 Linear cryptanalysis

Linear Cryptanalysis is a known plaintext attack that is based on effective linear approximate relations between the plaintext, the cipher text, and the key. Another powerful method of cryptanalysis is linear cryptanalysis introduced by Matsui (1993, 1994). It is a known plain text attack in which the attacker studies linear approximations of parity bits of the plain text, ciphertext and the secret key. Given an approximation with high probability and counting on the parity bits of the known plain texts and ciphertexts one obtains estimate of the parity bit of the key. Using auxiliary techniques one can usually extend the attack to find more bits of the secret key. In slightly more detail: In the basic form of linear cryptanalysis of an r round iterated block cipher, the analyst tries to find a linear approximation over $r-2$ rounds from the second round to the second last round, that is, an approximation of the form $a.X + b.Y + c.k = 0$ where $X = F_1(x, k_1), y = F_r(Y, k_r)$ is the ciphertext, x is plaintext and $k = (k_2, k_2, ..., k_{r-1})$ is a vector of all the unknown round keys used in rounds 2 to $r-1$. Given N known plaintext, the parts of the round keys k_1 and k_r relevant to the approximation can be found by trying all possible round subkeys at rounds 1 and r, and counting the number N_0 of plaintext for which $a.F_1(x, k_1) + b.F_r^{-1}(y, k_r) = 0$ holds. The round subkeys that maximizes $|N_0/N - 1/2|$ are chosen as the most likely candidates.

3.2 Differential cryptanalysis:

Differential cryptanalysis is a chosen plaintext attack that studies the propagation of input differences to output differences in iterated transformations. These difference propagations are formalized in the following definition.

Definition: Let $f : GF(2)^n \rightarrow GF(2)^m$, and let $a, a^* \in GF(2)^n$. The difference $a' = a \oplus a^*$ is said to propagate to the difference $b' = f(a) \oplus f(a^*)$ though f. This is denoted by $a' \xrightarrow{f} b'$. An expression of the form $\alpha \xrightarrow{f} \beta$ is called a differential. If the input difference of a pair is α, the differential $\alpha \rightarrow \beta$ can be used to predict the corresponding output difference. It is thus natural to measure the efficiency of a differential as the fraction of all inputs with difference α that results in the output difference β. Following Daemen(1995), we call this fraction the propagation ratio of the differential.

Definition: The propagation ratio R_p of the differential $\alpha \xrightarrow{f} \beta$ is defined by

$$R_p(\alpha \xrightarrow{f} \beta) = 2^{-n}|\{x \in GF(2)^n | f(x) \oplus f(x \oplus \alpha) = \beta\}|.$$

Discovery of differential cryptanalysis (1990) was a major breakthrough in the field of cryptanalysis in the last decade. It is a very powerful method of cryptanalysis. The main idea is to study the propagation of the differences from round to round in a pair of encryptions instead of studying a single encryption. This study is usually performed with a specially written program capable of searching for differential patterns in a cipher. This allows to make statistical predictions of the output difference of the pair. The attacker then encrypts a pool of pairs with the chosen difference and filters those pairs that support the prediction (in a simplistic case those that have expected ciphertext difference). These pairs reveal internal behavior of the cipher which is otherwise hidden from the attacker and thus help to find bits of the secret key. For example, the knowledge of the difference between two encryptions before the last round (due to the prediction) combined with the knowledge of the difference from the ciphertext and the knowledge of the ciphertexts themselves provides a simple equation for one round of a cipher. This equation contains the subkey of the last round as an unknown. Round function of an iterative cipher is usually not designed to be cryptographically strong (strength of a cipher relies on many iterations of a relatively weak round function). Thus given one or few such equations (all involving the same unknown secret key) it is possible to derive the subkey of the last round. Having achieved this result the attacker is left with a cipher which is shorter by one round. He proceeds with further analysis which becomes much easier and usually does not require additional data. Differential attack is a chosen plain text attack, which can be converted to a known plaintext attack scenario.

A generic differential attack against an r round iterated block cipher is the following.

Step 1. Find an $r-1$ round differential $\alpha \rightarrow \beta$ with high enough propagation ratio.

Step 2. Keep a counter for each possible round subkey k_r at round r. Initialize the counters to zero.

Step 3. Pick a plaintext x uniformly at random and set $x^* = x \oplus \alpha$. Encrypt the plaintexts under the unknown key k obtaining the ciphertexts y and y^*. For each possible round subkey k_r compatible with the assumed input difference

J and the observed outputs y, y^* at round r, add one to the corresponding counter.

Step 4. Repeat Step 3 until some round subkeys are counted significantly more often than the others. Output these keys as the most likely subkey at the last round.

Biham and Shamir (1993) proposed Differential cryptanalysis of DES.

3.3.1 Differential attack on DES

Differential cryptanalysis of DES was the first method capable of breaking DES faster than exhaustive search. It is a statistical attack which requires 2^{47} chosen plaintexts to break the DES cipher. It is based on the linearity of most of the operations used in DES: $E(X) \oplus E(X^*) = E(X \oplus X^*)$, $P(X) \oplus P(X^*) = P(X \oplus X^*)$. Where E is the expansion operation, P is the permutation, and K is any subkey. The only nonlinear operations are the S-boxes, for which the equation $S(X) \oplus S(X^*) = S(X \oplus X^*)$, does not hold. However, it was observed that for any particular input X OR not all the output X OR values are possible, and the possible ones do not appear uniformly, some of them appear more frequently then others. Using this observation the difference distribution table of an S-box can be defined as follows:

Definition: A table that shows the distribution of the input XORs and output XORs of all the possible pairs of an S-box is called the difference distribution table of the S-box. In this table each row corresponds to a particular input XOR and each column corresponds to a particular output XOR. The entries themselves count the number of pairs out of 64 possible pairs with the particular input XOR that yield the particular output XOR.

Each line in a difference distribution table contains 64 pairs distributed over 16 entries. Thus an average of the entries in each line of the table is exactly four. The first line of the difference distribution table of S_1 (Stinson, 2001) of DES shows that for the zero input XOR the output XOR must be zero. Also different lines in the table have different distributions and tables for different S-boxes are of course different. For example, for $X \oplus X^* = 34_x$, $S_1(X) \oplus S_1(X^*) = 2_x$ for 16 pairs out of 64. Or in other words the input XOR difference 34_x causes the output XOR difference to be 2 with probability $p = 16/64 = 1/4$. By using the linearity of the rest of the operations in the cipher we receive probabilistic approximation of the difference of output of the F-function and thus of one-round of DES, These approximations are called one round characteristics. It is possible to concatenate one-round characteristics in order to get longer characteristics. Here is a more strict definition of an n-round characteristic:

Definition: Associated with any pair of encryptions are the XOR value of its two plaintexts (denoted by P'), the XOR of its ciphertexts (denoted by C') and the XORs of the inputs and of the outputs of each round in the two executions. These values form an n-round characteristic (denoted by Y'). For a given input XOR P', the probability that a randomly chosen input pair with P' difference leads to Y' is called the probability of Y'. It can be expressed as $P(Y'|P')$.

We assume that in the process of concatenation of characteristics the probabilities of the characteristics are multiplied. This assumption can be justified empirically. It is important to note that there exist characteristics that can be concatenated with themselves. These characteristics are called iterative characteristics. We search for characteristics which have the highest probabilities. The higher is the probability of the characteristic that covers the whole cipher the less is the number of chosen plaintexts required for the attack. A useful notion of an active S-box may be introduced here.

Definition: An S-box S_i is said to be active in round k with respect to differential characteristic $Y^{'}$ if it has non-zero input difference in round k of $Y^{'}$.

The less is the number of active S-boxes in the differential characteristic - the higher is its probability. It can be shown that for DES the best characteristic can be built by iterating eight times a particular two-round characteristic. See Figure 5 for one such characteristic. The first round of this characteristic has $\alpha \rightarrow 0$ - XOR difference on the input of the F-function causes the output XOR difference of the F-function to be zero (with some probability). The second round of this characteristic has the form $0 \rightarrow 0$, which holds with probability one. In DES such a characteristic takes place for the difference $\alpha = 19600000$. It involves three adjacent active S-boxes S_1, S_2, S_3 with input differences of $3_x = 000011$, $32_x = 110010$, $2C_x = 101100$ respectively (after α has been expanded). The probability of this characteristic is $(14.8.10)/64^3 = 1/234$ which is rather low. This is due to the precautions taken by the designers of DES. They claim that they were aware of the high potential of differential cryptanalytic attacks.

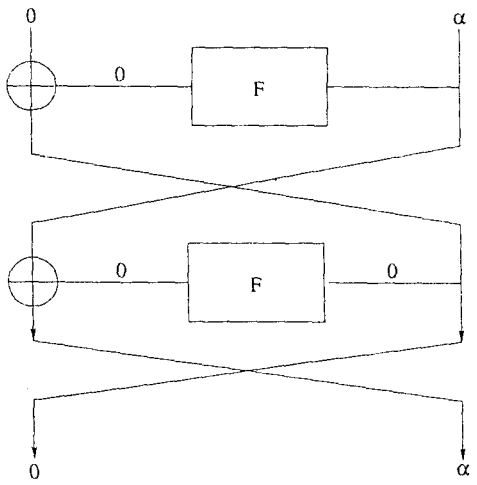

Figure 5: Characteristic of 2-DES

The Attack: Given the ideas described above, how the actual attack may work? In the simplest form, given a characteristic of probability $p >> 2^{-64}$ of

the full cipher, it is possible to distinguish a cipher from a random permutation. This can be done by querying the pairs of plaintexts with the difference P' as in the characteristic and counting the number of pairs that arrived at the ciphertext difference C' predicted by the characteristic. Such a distinguisher will use $O(p^{-1})$ pairs. Indeed, given M =C/p pairs (for some constant $C > 1$) chosen independently with the difference P', the probability that no one of them will follow the characteristic is $(1-p)^M = (1-p)^{C/p} < e^{-C}$ which can be made arbitrarily small by choosing sufficiently large C. On the other hand the probability that C' will not occur for similarly chosen pairs passed through a random permutation is $(1-2^{-64})^M$. This probability is very close to one if $p >> C \times 2^{-64}$.

Cryptanalysis on AES: The S-box of Rijndael is taking the multiplicative inverse of the input in $GF(2^8)$. This finite field inversion operation yields linear approximation and difference distribution table with the entries close to uniform distribution. This gives the security against linear and differential cryptanalysis. AES is secured against all known cryptanalysis. Cryptanalysis on AES is currently a very important research area. Using self-dual property Barkan and Biham (2002) gave an attack on AES which is slightly better than exhaustive search. They also proved that the choice of irreducible polynomial of Rijndael is arbitrary, hence it is irrelevant if the irreducible polynomial is primitive or not.

4 Statistical Attack on RC6

4.1 Description of RC6

RC6 is a block cipher submitted to NIST for consideration as the new AES. RC6 is designed by Rivest et al(2000). RC6-$w/r/b$ means that four w-bit word plaintext are encrypted with r round by b-byte keys. The (w, r, b) are called parameters. The nominal parameters for AES are (32,20,16), (32,20,24), (32,20,32) respectively for a 128, 196 and 256-bit user key.

Notation:
$+$: integer addition modulo 2^w,
$-$: integer subtraction modulo 2^w,
$\oplus$: bitwise exclusive-or,
$\times$: integer multification modulo 2^w,
$a <<< b$:cyclic rotation of a to left by b-bit,
$a >>> b$:cyclic rotation of a to right by b-bit,
There is a key scheduling algorithm which extends the original b-byte key into an $2r+4$-word array $S[0, 1, ..., 2r+3]$. The encryption is performed by using four register A, B, C, D.The algorithms is described as below:
Input:
Plaintext stored in four w-bit input register A, B, C, D.

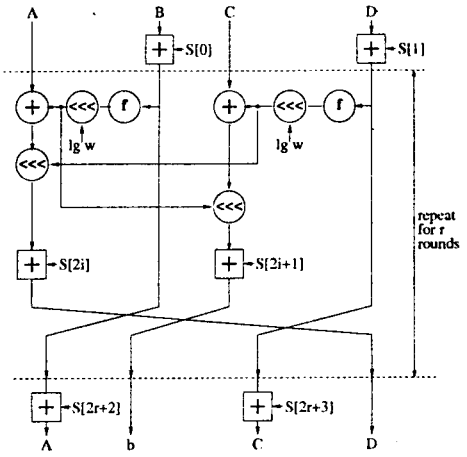

Figure 6: Encryption with RC6-$w/r/b$. Here $f(x) = x(2x+1)$.

w-bit round keys $S[0, 1, ..., 2r+3]$
Output:
Cipher stored in A, B, C, D.
Procedure:
$B = B + S[0]$, $D = D + S[1]$
for i=1 to r do
{
$t = (B \times (2B+1)) <<< lgw$
$u = (D \times (2D+1)) <<< lgw$
$A = ((A \oplus t) <<< u) + S[2i]$
$C = ((C \oplus u) <<< t) + S[2i+1]$
$(A, B, C, D) = (B, C, D, A)$
}
$A = A + S[2r+2]$, $C = C + S[2r+3]$.

Table 1

Attack	Rounds	Number of plaintexts
Linear attack (Borst et al, 1999)	16	2^{119}
Differential attack (Contini et al, 1998)	12	2^{117}
Multiple Linear attack (Borst et al, 1999)	14	$2^{119.68}$
Multiple Linear attack (Borst et al, 1999)	18	$2^{126.94}$
χ^2 attack (Handschuh and Gilbert, 1997)	15	2^{119}
χ^2 attack (Handschuh and Gilbert, 1997)	17	2^{118}

Table 1 summarizes some of the cryptanalysis on RC6. Up to the present, linear attacks, Differential attacks, and χ^2-attacks against RC6 and some simplified variants of RC6 have been analyzed intensively. The secuirty of RC6 against

the linear and Differential cryptanalysis is discussed in Contini et al (1998) paper. They estimated that 12 round RC6 is not secure against the Differential cryptanalysis and RC6 with 16 or more round is secure against linear cryptanalysis. Currently, RC6 with paprameter (31,20,..) is recommended to give suffcient resistance against the Borst et al (1999), Contini et al (1998), Contini et al (1999), Gilbert et al (2000), Knudsen and Meier (2001), Shimoyama et al (2002), Shimoyama et al (2000) attack. χ^2-attack is one of the most effective attacks on RC6. The χ^2-attack was first proposed by Vaudenay (1996) as an attack on DES. Gilbert et al (2000), Knudsen and Meier (2001), Borst et al (1999) applied χ^2-attacks to RC6 or a simplied variant of RC6.
The χ^2-attack can be used for both distinguishing attacks and key recovery attacks. Distinguishing attacks handle plaintexts in such a way that the χ^2-value of a part of ciphertext becomes significantly a higher value. Key recovery attack have to rule out all wrong keys, single out exactly a correct key by using the χ^2-value, and thus they often require more work and memory than distinguishing attacks.

4.2 χ^2 cryptanalysis

4.2.1 χ^2 test

The block cipher is a random permutation (here the randomness comes from the random choice of the secret key). The goal of the block cipher designer is to make it "look like" a truely random permutation i.e like a random permutation with a uniform distribution among the set of permutation. To distinguish a random source with some unknown distribution from a random source with uniform distribution, a common tool for this is χ^2 test.
Suppose X_i, $i = 1, 2, ..., n$ be *iid* (independent and identically distributed random variable) observation from a population X, which takes values in the set $\{x_1, x_2, ..., x_m\}$. The χ^2 test is used to decide if an observation X_1, X_2 ,...,X_n is consistent with the hypothesis $P\{X = x_j\} = p(j)$, $j = 1, 2....m$, where $\sum_j p(j) = 1$. Let N_j denote the number of times the observation X takes on the value x_j. The χ^2 statistic is the random variable defined by:
$\chi^2 = \sum_j (N_j - np(j))^2 / np(j)$, where $\sum_j N_j = n$. In a χ^2 test, the observed χ^2 statistic is compared to $\chi^2{}_{a,m-1}$, the threshold for the χ^2 test with $m - 1$ degrees of freedom and with significance level a.

4.2.2 χ^2 statistics of RC6

To investigate the nonrandomness of r-round of RC6, the analysis is based on systematic experiments on increasing numbers of rounds of RC6 with varying word lenth w. The method is used to demonstate that detecting and quantifying nonrandomness is experimentally feasible up to 6 rounds of RC6. For this purpose, the least significant $log_2 w$ bits of words A and C of the input are fixed to zero. Depending on the experiment and the number of round, the remaining input bits are either chosen randomly or more of the remaining input bits are

suitably fixed so that one(or both) of the data dependent rotations are zero. Knudsen and Meier (2000) persued the χ^2 statistic of the integer of size twice $log_2 w$ bits as obtained by concatenating the least significant $log_2 w$ bits in the words A and B every two rounds later. In the experiments, they consider $w = 8$, 16 and 32 bits, respectively. It is instructive to see that the general behaviour of the χ^2 test for increasing number of rounds in all three cases is very similar. To judge the outcome of these χ^2 tests note that for the word sizes w as considered, 6-bit, 8bit and 10-bit integers are tested at the output. Hence the degrees of freedom are 63, 255 and 1023 respectively, and these number coincide with the expected value of the χ^2 statistic, provided the distribution to be tested is uniform.

5 Attack on other block ciphers

There are several improvements to the basic differential attack that reduce the number of plaintexts needed. There are also attacks using $r - 2$ round differentials that counts on the round subkeys of the last two rounds. Using straightforward statistical analysis, it can be shown that the correct round key can be distinguished from a randomly selected key with sufficient confidence, provided that the number of plaintexts available is inversely proportional to the propagation ratio of the differential used. Thus, a necessary condition for resistance against conventional differential attacks is that there does not exist any differential ranging over all but a few (say, 3) rounds with propagation ratio significantly larger than 2^{-n}, where n is the block size. If we consider COCONUT98 cipher, Vaudenay (1998) proved by decorrelation technique that the full COCONUT98 cipher admits no good differential characteristics using decorrelation technique. So COCONUT98 is resistant against differential cryptanalysis. But we observe that there are differential characteristics of very high probability for the half of the cipher. Wagner(1999) made extensive use of these characteristics in his boomerang attack. Note that resistance against conventional differential attacks does not imply anything about resistance against natural extensions to differential cryptanalysis, such as impossible [Biham et al, 1999a, b], higher order [Knudsen, 1995; Lai, 1994] and truncated (Knudsen, 1995) differentials, and the boomerang attack.

5.1 The Boomerang Attack

The main idea behind the boomerang attack is to use two short differentials with high probabilities instead of one differential of more rounds with low probability. The motivation for such an attack is quite apparent, as it is easier to find short differentials with a high probability than finding a long one with a high enough probability. We assume that a block cipher $E : \{0,1\}^n \times \{0,1\}^k \rightarrow \{0,1\}^n$ can be described as a cascade $E = E_1 \circ E_0$, such that for E_0 there exists a differential $\alpha \rightarrow \beta$ with probability p, and for E_1 there exists a differential $\gamma \rightarrow \delta$ with probability q. The boomerang attack uses the first characteristic

$\alpha \to \beta$ for E_0 with respect to the pairs (P_1; P_2) and (P_3; P_4), and uses the second characteristic $\gamma \to \delta$ for E_1 with respect to the pairs (C_1; C_3) and (C_2; C_4).

The attack is based on the following boomerang process:

Step1: Ask for the encryption of pair of plaintexts (P_1; P_2) such that $P_1 \oplus P_2 = \alpha$ and denote the corresponding ciphertexts by (C_1; C_2).

Step2: Calculate $C_3 = C_1 \oplus\delta$ and $C_4 = C_2 \oplus\delta$, and ask for the decryption of the pair (C_3; C_4). Denote the corresponding plaintexts by (P_3; P_4).

Step3: Check whether $P_3 \oplus P_4 = \alpha$.

It is easy to see that for a random permutation the probability that the last condition is satisfied is 2^{-n}. For E, however, the probability that the pair (P_1; P_2) is a right pair with respect to the first differential $\alpha \to \beta$ is p. The probability that both pairs (C_1; C_3) and (C_2; C_4) are right pairs with respect to the second differential is q^2. If all these are right pairs, then they satisfy $E_1^{-1}(C_3)\oplus E_1^{-1}(C_4)= \beta = E_0(P_3)\oplus E_0(P_4)$, and thus, with probability p also $P_3 \oplus P_4 = \alpha$. Therefore, the total probability of this quartet of plaintexts and ciphertexts to satisfy the boomerang conditions is $(pq)^2$. Therefore, $pq > 2^{-n/2}$ must hold for the boomerang attack to work.

5.2 Slide attacks

Slide attack was proposed by Biryukov and Wagner (1999). In the simplest case, we have an r-round block cipher E whose round functions are same and use the same subkey, so that $E = F \circ F \circ ..... \circ F = F^r$. Let (P, C) be a known Plaintext-ciphertext pair for E. The crucial observation is, if $P^{'}$ = F(P) then $C^{'}$ = E($P^{'}$)=F^r(F(P))=F(F^r(P))=F(C). In a slide attack, we try to find pairs (P, C), ($P^{'}$, $C^{'}$) with $P^{'}$ = F(P), we call such a pair a slide pair, and then we will get the extra relation $C^{'}$=F(C).

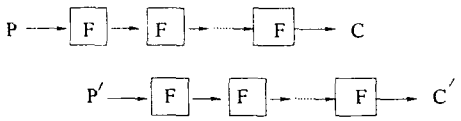

Figure 7: A typical slide attack

A slide attack provides a very general attack on block cipher with repeating round subkeys. The only requirement on F is that it is week against known-plaintext attack with two pairs. More precisely, we call $F_k(x)$ a weak function if given the two equations $F_k(x_1)$=y_1 and $F_k(x_2)$=y_2 it is easy to extract the key k (for example 3-round DES is a weak function). Such a cipher (with a n-bit block) can be broken with only $2^{n/2}$known texts, since then we obtain 2^n possible pairs (P, C), (P', C'). As each pair has a 2^{-n} chance of forming a slide pair, we expect to see one slide pair which discloses the key.

5.3 New types of cryptanalysis attacks using Related Keys

Biham (1994) proposed a new type of attack using related keys. These attacks are based on the observation that in many blockciphers we can view the key scheduling algorithms as a set of algorithms, each of which extracts one particular subkey from the subkeys of the previous few rounds. If all the algorithms of extracting the subkeys are the various rounds are same, then given a key we can shift all the subkeys one round backwards and get a new set of valid subkeys which can be derived from some other key, these keys are called related keys. Biham (1994) also combined this attack with the attacks based on complementation properties.

6 Conclusion

Cryptanalysis of block cipher models, that are in practice, are mostly intuitive and straight forward. Elegant statistical approaches are looked for from statistician. In addition, different statistical methods may be adopted to reduce the computational complexity of the cryptanalysis. For further references to recent work in this area the reader may refer to IACR website (http://www.iacr.org). This area demands attention from the statistics community.

References

Barkan, E. and Biham, E. (2002). In how many ways can you write Rijndael? *Advances in Cryptology - ASIACRYPT* , **2501 / 2002**, Springer Verlag, 160-175.

Biham, E., Biryukov, A., and Shamir, A. (1999a). Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In *Advances in Cryptology—Eurocrypt '99*, LNCS **1592**, 12–23. SpringerVerlag.

Biham, E., Biryukov, A., and Shamir, A. (1999b). Miss in the middle attacks on IDEA and Khufu. *Proceedings Fast Software Encryption*, LNCS **1636**, 124–138.

Biham, E. (1994). New Types of Cryptanalytic Attacks Using Related Keys. *J. Crypt.*, **7(4)**, 229-246.

Biham, E. and Shamir, A. (1993). *Differential Cryptanalysis of the Data Encryption Standard,* SpringerVerlag.

Biryukov, A. and Wagner, D. (1999). Slide attack. *Proceedings Fast Software Encryption,* LNCS **1636**, Springer Verlag, 245-259.

Borst, J., Preneel, B., and Vandewalle, J. (1999). Linear cryptanalysis of RC5 and RC6. *Proceedings Fast Software Encryption*, LNCS **1636**, 16-30.

Contini, S., Rivest, R., Robshaw, M., and Yin, Y. (1998). The Security of the RC6 Block Cipher. v 1.0. Available at http://www.rsasecurity.com/rsalabs /rc6/.

Contini, S., Rivest, R., Robshaw, M., and Yin, Y. (1999). Improved analysis of some simplified variants of RC6. *Proceedings Fast Software Encryption*, LNCS **1636**, 1-15.

Daemen, J. (1995). Propagation and correlation. In *Cipher and Hash Function Design. Strategies Based on Linear and Differential Cryptanalysis,* chapter 5.Katholieke Universiteit Leuven, Available from the Rijndael page, http://www. esat.kuleuven.ac.be/~rijmen/rijndael/.

Daemen, J. and Rijmen, V. (1998). AES Proposal: Rijndael. Submitted to the Advanced Encryption Standard (AES) contest, 1998.

Daemen, J. and Rijmen, V. (2001). *The Design of Rijndael.* First Edition, Springer.

Even, S. and Mansour, Y. (1997). A construction of a cipher from a single Pseudorandom permutation. *J. Crypt.*, **10**, 151-161.

Gilbert, H., Handschuh, H., Joux, A. and Vaudenay, S. (2000). A statistical attack on RC6. *Proceedings Fast Software Encryption,* LNCS **1978**, 64-74.

Handschuh, H. and Gilbert, H. (1997). X^2 Cryptanalysis of the SEAL Encryption Algorithm. *Proceedings Fast Software Encryption,* LNCS **1267**, 1-12.

Knudsen, L. (1995). Truncated and higher order differentials. *Proceedings Fast Software Encryption,* LNCS **1008**, 196–210.

Knudsen, R. L. and Meier, W. (2000). Correlations in RC6 with a reduced number of Rounds. *Proceedings Fast Software Encryption,* 94-108.

Knudsen, L. and Meier, W. (2001). Correlations in RC6 with a reduced number of rounds. *Proceedings Fast Software Encryption,* LNCS **1978**, 94-108.

Lai, X. (1994). Higer order derivatives and differential cryptanalysis. *In Communication and Cryptography,* Kluwer Academic Publishers, 227–233.

Lidl, R. and Niederreiter, H. (1994). *Introduction to finite fields and their applications.* Revised Edition, Cambridge University Press.

Matsui, M. (1993). Linear cryptanalysis method for DES cipher. *In Advances in Cryptology—Eurocrypt '93,* LNCS **765**, 386–397.

Matsui, M. (1994). The first experimental cryptanalysis of the Data Encryption Standard. *In Advances in Cryptology—Crypto '94,* LNCS **839**, 1-11.

National Bureau of Standards (1977). Data Encryption Standard. U.S. Department of Commerce, FIPS pub. **46**.

National Institute of Standards and Technology (2001). Advanced Encryption Standard. *Federal Information Processing Standard,* FIPS-**197**.

Rivest,L.R, Robshaw,B.J.M., Sidney,R., Yin,L.Y. (2000). The RC6 block cipher. AES Candidate Conference 2000, 337-342.

Roy, B. and Palit, S. (2004). Some statistical attacks on stream cipher cryptosystem. *J. Ind. Statist. Assoc.*. To appear.

Shimoyama, T., Takenaka, M. and Koshiba, T. (2002). Multiple linear cryptanalysis of a reduced round RC6. *Proceedings Fast Software Encryption,* LNCS **2365**, 76-88.

Shannon, C. (1949). Communication theory of secrecy systems. *Bell Sys. Tech. Journal,* **28**, 656-715.

Shimoyama, T., Takeuchi, M. and Hayakawa, J. (2000). Correlation attack to the block cipher RC5 and simplified variants of RC6. 3rd AES Candidate Conference.

Stinson, R. D (1995). *Cryptography theory and practice.* First Edition, CRC press company.

Stinson, R. D (2001). *Cryptography theory and practice.* Second Edition, CR press company.

Vaudenay, S. (1998). Provable Security for block cipher by decorrelation. *Pr ceedings STACS'98*, LNCS **1373**.

Vaudenay, S. (1996). An experiment on DES statistical cryptanalysis. *Proce ings 3rd ACM Conference on Computer and Communications Securi* ACM Press, 139-147.

Wagner D. (1999). The boomerang attack. *Proceedings Fast Software Encry tion* , LNCS **1636**, 156–170.

Received : September 2003
Revised : November 2003