Scribe: Cryptography and Network Security (Week 7 Class 3)

Rutwik Pandit (20CS60R13)

Class 26 : Fri 16th October 2020

1 Introduction

A linear cryptanalysis involves constructing a linear approximation function mapping cipher text as a linear approximate function of the plain text and the key.It is a known plaintext attack.A linear function will consist in XOR of plaintext bits and Key bits.The underlying assumption when we do linear cryptanalysis is that our encryption is not truly a random process and that if we can find biases in the input output relation , then we can find partial portion of the key much more easily and reduce the computation complexity required for the rest of key search.
For eg , lets say we had a 16 bit key , brute searching would need 2^{16} combinations. Now if we use linear cryptanalysis to find 8 of those 16 bits , our search space now becomes only 2^8 combinations , which means we might be able to brute force on the rest of the 8 bits.

2 S Box Linear Cryptanalysis

In our guinea pig cipher we used a combination of S boxes and permutations , since the S boxes are non linear , we try to approximate a linear equation foe the S box.
In our example we had a 4 bit - 4 bit S box , thus we had a total of 256 different linear combinations.For each combination we check ho many times the equation holds false ie the output is 0.Now in a perfect world for each linear equation the probability that it will be zero is exactly 0.5.But we dont live in a perfect world , thus we find that many combinations give us probabilities different from 0.5 , this will give us our biases ϵ.The value of ϵ is given as the probability of that event - 0.5.The higher the value of ϵ , the easier it becomes to crack the cipher.In a way a high value of ϵ means that the cipher is leaking information , which we can use to determine the key.

3 Linear Approximation Table

	b															
a	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
1	8	8	6	6	8	8	6	14	10	10	8	8	10	10	8	8
2	8	8	6	6	8	8	6	6	8	8	10	10	8	8	2	10
3	8	8	8	8	8	8	8	8	10	2	6	6	10	10	6	6
4	8	10	8	6	6	4	6	8	8	6	8	10	10	4	10	8
5	8	6	6	8	6	8	12	10	6	8	4	10	8	6	6	8
6	8	10	6	12	10	8	8	10	8	6	10	12	6	8	8	6
7	8	6	8	10	10	4	10	8	6	8	10	8	12	10	8	10
8	8	8	8	8	8	8	8	8	6	10	10	6	10	6	6	2
9	8	8	6	6	8	8	6	6	4	8	6	10	8	12	10	6
A	8	12	6	10	4	8	10	6	10	10	8	8	10	10	8	8
B	8	12	8	4	12	8	12	8	8	8	8	8	8	8	8	8
C	8	6	12	6	6	8	10	8	10	8	10	12	8	10	8	6
D	8	10	10	8	6	12	8	10	4	6	10	8	10	8	8	10
E	8	10	10	8	6	4	8	10	6	8	8	6	4	10	6	8
F	8	6	4	6	6	8	10	8	8	6	12	6	6	8	10	8

The linear approximation table is a tabular form of the 256 different linear equations , in the form 16 columns for the coefficients of output and 16 rows for coefficients of input.
Each entry in the table gives us the number of times the value of the equation is 0. The probability of the output being 0 can be calculated by dividing each element by 16. The bias of each element in the table can be calculated by substracting 8 from each element and dividing by 16.Ideally we want to chose linear equations with high absolute value of bias

4 Tracing the bits

Once we find a good linear equation with a high bias , we trace how this affects the next layer in the cipher , ideally we want to disturb as small number of s boxes as possible.Then using the pilling up lemma , we find the overall bias in the output.The final equation must have terms only in the form of input bits , penultimate output bits and key bits.

5 The Attack

We have a linear equation whose output we know is biased.If our key guess matches the real key ,we will see this bias in the output.If on the other hand

we don't match with the input key , we will get an almost bias free output that is the probabilty of 0 will be 0.5 .Thus if our bias is substantial , then we can easily determine the key with enough amount of plain text , cipher text pairs. In our example , we will need to guess 16 guesses for the 4 key bits.But after we get the correct bits for these , the rest i.e only 4 more bits will require again only 16 guesses.Thus with a single linear cryptanalysis attack , we can find the key in 32 guesses , where for a brute force search we would have needed 256 guesses.

6 Exercise Problem

Suppose that X1, X2 and X3 are independent discrete random variables defined on the set 0, 1. Let ϵ_i denote the bias of Xi , for i = 1, 2, 3. Prove that if X1 $\oplus$ X2 is independent of X2 $\oplus$ X3, then either ϵ_1*ϵ_3=0 or ϵ_2=$\pm$1/2

Solution:
The bias of X1 $\oplus$ X2 = 2*ϵ_1*ϵ_2 , and the bias of X2 $\oplus$ X3 = 2*ϵ_2*ϵ_3. Now if X1 $\oplus$ X2 is independent from X2 $\oplus$ X3 , then its overall bias would be the same as bias of X1 $\oplus$ X3 = 2*ϵ_1*ϵ_3.
Thus bias of (X1 $\oplus$ X2) $\oplus$ (X2 $\oplus$ X3) =bias of X1 $\oplus$ X3
2*((2*ϵ_1*ϵ_2)*(2*ϵ_2*ϵ_3))= 2*ϵ_1*ϵ_3.
Thus Either ϵ_1*ϵ_3=0 or 4*ϵ_2*ϵ_2=1
Thus Either ϵ_1,ϵ_3=0 or ϵ_2=$\pm$1/2

References

Lecture video.