On Linear Cryptanalysis with Many Linear Approximations

B. Gérard and J.-P. Tillich

INRIA, project-team SECRET

12th IMACC - 15th December 2009

Outline

Outline

Notation

A block cipher E

- is parametrized by a key $\mathbf{k}$;
- enciphers a fixed length plaintext $\mathbf{p}$ to a ciphertext $\mathbf{c}$.

$$\begin{array}{rccc} E_{\mathbf{k}}: & \mathcal{P} & \longmapsto & \mathcal{C} \\ & \mathbf{p} & \longrightarrow & \mathbf{c} \end{array}$$

We will denote

- vectors using bold letters $\mathbf{x}$;
- random variables using capital letters X.

Linear Approximation

For $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{F}_2^m$, we denote by $< \cdot, \cdot >$ the scalar product

$$< \mathbf{a}, \mathbf{b} > \stackrel{\text{def}}{=} \bigoplus_{i=1}^{m} a_i \cdot b_i.$$

A linear approximation of a block cipher E is defined by a set $(\pi, \gamma, \kappa, b, \varepsilon)$ that fulfills

$$\Pr\left[< \mathbf{P}, \pi > \oplus < E_{\mathbf{K}}(\mathbf{P}), \gamma > \oplus < \mathbf{K}, \kappa >= b\right] = \frac{1}{2} + \varepsilon.$$

Vectors π, γ and κ are respectively named plaintext/ciphertext/key mask and ε is named bias of the approximation.

Linear Cryptanalysis: a Statistical Attack

A statistical attack aims at recovering the key $\mathbf{k}^*$ used to encipher a set of N samples $(\mathbf{p_i}, \mathbf{c_i})_{1 \leq i \leq N}$ intercepted by the attacker.

1. **Distillation phase**:
   Extracts some statistic Σ from the N available samples.
2. **Analysis phase**:
   Sorts the list of candidates in regard to there likelihood using Σ.
3. **Search phase**:
   Does an exhaustive search among this list to recover the key.

Linear Cryptanalysis: a Statistical Attack

- Actually the *Search Phase* is done up to the ℓ-th candidate where ℓ determines the Time Complexity of the attack.
- The rank of the correct key in the list of candidates is denoted by r^*.

Data Complexity	Time Complexity	Success Probability
N	$= r^*$ $\leq \ell$	$\Pr[R^* \leq \ell]$

Type 1 Attack

$$\Pr\left[< \mathbf{P}, \pi > \oplus < \mathbf{C}, \gamma > \oplus < \mathbf{K}, \kappa >= b\right] = \frac{1}{2} + \varepsilon.$$

Algorithm 1: Algorithm for type 1 attack

Data: N samples $(\mathbf{p_i}, \mathbf{c_i})_{1 \leq i \leq N}$ enciphered using $\mathbf{k}^*$.

Result: A guess for $< \mathbf{k}^*, \kappa >$.

$T \longleftarrow 0$

for $1 \leq i \leq N$ **do**

$\quad T \longleftarrow < \mathbf{p_i}, \pi > \oplus < \mathbf{c_i}, \gamma >$

end

return $\begin{cases} b \oplus 1 & \textbf{if } T > N/2; \\ b & \textbf{otherwise}. \end{cases}$

An Idea : Using Many Approximations

Aims

- Recovering more bits of $\mathbf{k}^*$.
- Reducing the data complexity.

We use n linear approximations $(\pi_i, \gamma_i, \kappa_i, b_i, \varepsilon_i)_{1 \leq i \leq n}$.

Hence, we can recover $\tilde{\mathbf{k}^*} = (< \kappa_1, \mathbf{k}^* >, \ldots, < \kappa_n, \mathbf{k}^* >)$

Let d be the number of bits of information $\tilde{\mathbf{k}^*}$ contains about $\mathbf{k}^*$, i.e.

- the dimension of the linear subspace spanned by κ's;
- the entropy of the corresponding random variable $\tilde{\mathbf{K}}$.

Outline

Using a structured set of approximations (Hermelin et al.)

1. Choose d linearly independent approximations $(\pi_i, \gamma_i, \kappa_i, b_i, \varepsilon_i)$ named *base-approximations*.
2. Compute the correlations (biases) of the 2^d approximations in the subspace spanned by these d base-approximations.
3. Base-correlations $\longrightarrow$ probability distribution of d-dimensional random variables $\mathbf{S}|\tilde{\mathbf{K}} = \mathbf{k}$ where

$$\mathbf{S} = (< \mathbf{P}, \pi_1 > \oplus < \mathbf{C}, \gamma_1 >, \ldots, < \mathbf{P}, \pi_d > \oplus < \mathbf{C}, \gamma_d >)\,.$$

4. Sort candidates $\mathbf{k}$ with regard to there likelihood $\Pr\left[\Sigma|\tilde{\mathbf{K}} = \mathbf{k}\right]$.

With $\Sigma = (\mathbf{s_1}, \ldots, \mathbf{s_N})$.

Using a structured set of approximations (Hermelin et al.)

Positive points

- No independence assumptions on approximations.
- The work of Baignères, Junod and Vaudenay can be used.

Negative points

- Needs a structured set of approximations.
- Some assumptions lead to pessimistic theoretical results.

- Theoretical framework based on boolean functions.

Using any set of approximations

- Choose n linear approximations $(\pi_i, \gamma_i, \kappa_i, b_i, \varepsilon_i)_{1 \leq i \leq n}$.
- Recall: d is the dimension of the subspace spanned by the κ's.
- Hence, $\tilde{\mathbf{K}} = (< \mathbf{K}, \kappa_1 >, \ldots, < \mathbf{K}, \kappa_n >)$ is a n-bits vector containing d bits of information about $\mathbf{K}$.

Thus, $\tilde{\mathbf{K}}$ can be seen as a codeword of a linear code of length n and dimension d.

Then, the *Analysis Phase* boils down to a **maximum likelihood soft decision list decoding** problem on a linear code.

When d is small enough ($\approx 30 - 40$) this can be done using a Walsh-Hadamard transform but algorithms exist for a bigger d.

Using any set of approximations

Positive point

- Any set of approximations can be used.

Negative point

- The stochastic independence of approximations is assumed.

Biryukov et al.
Results based on an estimate of $\mathbb{E}(R^*)$ where R^* is the random variable corresponding to the rank of $\tilde{k}^*$ in the list of candidates.

On the relevance of studying $\mathbb{E}(R^*)$

Studying $\mathbb{E}(R^*)$ may not be relevant in this setting.

In some extremely rare situations, R^* will take a very large value.

This implies pessimistic theoretical results if considering $\mathbb{E}(R^*)$.

- Known phenomenon in Coding Theory.
- Mentioned by Junod in 2001 about Matsui's attack theoretical and empirical complexities.

Solution proposed

Using an **entropy based approach** is well suited for this problem.

Outline

Model used

Memoryless canal with Additive White Gaussian Noise.

$$(< \mathbf{k}^*, \kappa_1 >, \ldots, < \mathbf{k}^*, \kappa_n >) \longrightarrow (Y_1, \ldots, Y_n) = \mathbf{Y}$$

- **Memoryless**: stochastic independence of approximations.
- **Additive White Gaussian Noise**: Central Limit Theorem.

$$Y_j = (-1)^{<\mathbf{k}^*, \kappa_j>} + GN_j.$$

Where the Gaussian noises GN_j are normally distributed :

$$GN_j \sim \mathcal{N}\left(0, \frac{1}{4N\varepsilon_j^2}\right).$$

Some words about Entropy

- The entropy quantify the uncertainty of a random variable.

Let X (resp. Y) be a discrete random variable that takes values in $\mathcal{X}$ (resp. $\mathcal{Y}$).

$$
\begin{aligned}
H(X) &\stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} \Pr[X = x] \log_2(\Pr[X = x]). \\
H(X|Y) &\stackrel{\text{def}}{=} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \Pr[X = x] \log_2(\Pr[X = x | Y = y]). \\
I(X;Y) &\stackrel{\text{def}}{=} H(X) - H(X|Y).
\end{aligned}
$$

The generalization for continuous variables is straightforward.

A Bound on Conditional Entropy

$$I(\tilde{\mathbf{K}};\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\tilde{\mathbf{K}}) \leq \sum_{1}^{n} H(Y_j) - H(\mathbf{Y}|\tilde{\mathbf{K}}).$$

Chain rule for entropy

$$H(Y_1,\ldots,Y_n|\tilde{\mathbf{K}}) = H(Y_1|\tilde{\mathbf{K}}) + \cdots + H(Y_n|Y_{n-1},\ldots,Y_1,\tilde{\mathbf{K}}).$$

This together with the stochastic independence assumption leads to

$$H(\mathbf{Y}|\tilde{\mathbf{K}}) = \sum_{j=1}^{n} H(Y_j|\tilde{K}_j).$$

And finally

$$H(\tilde{\mathbf{K}}|\mathbf{Y}) \geq d - \sum_{j=1}^{n} I(\tilde{K}_j; Y_j).$$

Comparison with Biryukov et al.

The required Data Complexity to have $\Pr[R^* = 1] \approx 1$ is

$$N \approx \frac{d \ln(2)}{\sum_{i=1}^{n} \varepsilon_i^2} (1 + o(1)).$$

In the paper: proof that taking $\sum_{j=1}^{n} I(\tilde{K}_j; Y_j) = d + \delta$ leads to $\Pr[R^* = 1] \approx 1 - o\left(\frac{1}{\delta^2 n} + 2^{-\delta n/2}\right)$.

The required Data Complexity to have $\sum_{j=1}^{n} I(\tilde{K}_j; Y_j) \approx d$ is

$$N \approx \frac{d \ln(2)}{2 \cdot \sum_{i=1}^{n} \varepsilon_i^2} (1 + o(1)).$$

Comparison with Biryukov et al.

The estimate from Biryukov et al. is twice bigger!

✓ This confirms the fact that $\mathbb{E}(R^*)$ is not relevant here.

Remark 1:
For a given N, keeping $2^{H(\tilde{\mathbf{K}}|\mathbf{Y})}$ candidates leads to an attack with a good probability of success. Thus, the previous estimate is an estimate of the number of candidates to try.

Remark 2:
In the paper: formulae for 3 different types of linear attack but this point of view is general and could be used for other cryptanalyses.

Some results

Linear cryptanalysis of the DES performed knowing 2^{39} pairs and using 32968 approximations. The bound on the conditional entropy estimates the number of tried candidates to 2^{40}.

- ✓ Only 3 experiments over 19 failed that is $P_S \approx 0.84$.
- ✓ This attack is better than Matsui's attack using a distinguisher as soon as less than 2^{42} pairs are known.

Using the estimate for $H(\mathbf{K}|\mathbf{Y})$, we found that Matsui's attack is successful with a high probability when at most 2^{41} DES encryptions are done.

- ✓ This is exactly what Junod wrote in 2001 concerning his experiments.

Outline

Conclusion

- We proposed an information theoretical framework for analysing linear cryptanalysis using many approximations **that is non pessimistic**.
- We performed a type 1 attack on DES using many approximations **that challenges the other linear attacks on DES**.

Further work

- Deep study of soft decision list decoding algorithm for random linear codes.
- Improving the theoretical results of Hermelin et al..
- Combining information from many structured sets of approximations to minimize the restriction on the choice of approximations.