Statistical Cryptanalysis of Block Ciphers

Pascal Junod

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Aussois (France), February 2$^{\text{nd}}$, 2005

Outline

Cryptanalysis of Block Ciphers

- Most existing "generic" attacks against block ciphers are of statistical nature.
  - Differential cryptanalysis (and variants) [Biham-Shamir, 1990,...]
  - Linear cryptanalysis [Matsui, 1993]
  - Davies and Murphy's attack [Davies-Murphy, 1995]
  - χ^2 cryptanalysis [Vaudenay, 1996]
  - Partitioning cryptanalysis [Harpes-Massey, 1997]
  - Stochastic cryptanalysis [Minier-Gilbert, 2000]
- Focus is often put on the "deviant" property itself.

In this Talk

Focus

In this talk, we are mostly interested in how it is possible to optimally exploit these deviant properties.

Outline

Linear Cryptanalysis

- Matsui's attacks against DES (1993)
- $\longrightarrow$ First observations by Shamir/Franklin (1985)
- $\longrightarrow$ Tardy-Corfdir and Gilbert's attack against FEAL (1991)
- First successful experimental attack against DES (Matsui, 1994)

Best Known Linear Approximation of 15-round DES

- The best known linear approximation on 15-round DES is

$$\mathbf{x}_{\mathsf{l}\{7,13,24\}} \oplus \mathbf{x}_{\mathsf{r}\{15,19\}} \oplus \mathbf{y}_{\mathsf{l}\{2,7,13,24\}} \oplus \mathbf{y}_{\mathsf{r}\{16\}} =$$
$$\mathbf{k}^{(1)}_{\{24,28\}} \oplus \mathbf{k}^{(3)}_{\{25\}} \oplus \mathbf{k}^{(4)}_{\{3\}} \oplus \mathbf{k}^{(5)}_{\{25\}} \oplus \mathbf{k}^{(7)}_{\{25\}} \oplus \mathbf{k}^{(8)}_{\{3\}} \oplus \mathbf{k}^{(9)}_{\{25\}} \oplus \mathbf{k}^{(11)}_{\{25\}} \oplus$$
$$\mathbf{k}^{(12)}_{\{3\}} \oplus \mathbf{k}^{(13)}_{\{25\}} \oplus \mathbf{k}^{(15)}_{\{25\}}$$

  where $\mathbf{k}^{(i)}_{\{\mathcal{B}\}}$ denotes the set $\mathcal{B}$ of the i-th round subkey. The above linear approximation holds with probability $\frac{1}{2} - 1.19 \cdot 2^{-22}$.
- We can write the linear approximation as $\mathbf{a} \cdot \mathbf{x} \oplus \mathbf{b} \cdot \mathbf{y} = \mathbf{c} \cdot \mathbf{k}$.

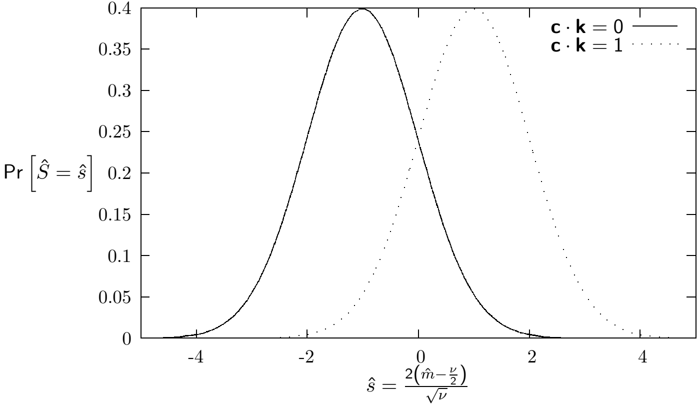
Information Extraction About the Key (1)

- **Input**: an oracle $\mathbf{\Omega}$, a data complexity ν, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, ε.
- **Output**: a guess about $\mathbf{c} \cdot \mathbf{k}$
- Initialize a counter $\hat{m}$ to 0.
- **For** $i \leftarrow 1$ **to** $i = \nu$
  - Generate a plaintext $\mathbf{x}_i$ uniformly at random and independently of the other queries. Submit $\mathbf{x}_i$ to $\mathbf{\Omega}$ and get $\mathbf{y}_i = \mathsf{f}_{\mathbf{k}}(\mathbf{x}_i)$.
  - **If** $\mathbf{a} \cdot \mathbf{x}_i \oplus \mathbf{b} \cdot \mathbf{y}_i = 0$
  - Increment $\hat{m}$.
  - **End If**
- **End For**

Information Extraction About the Key (2)

- **If** $\varepsilon > 0$
  - **If** $\hat{m} > \frac{\nu}{2}$
  - Output "$\mathbf{c} \cdot \mathbf{k} = 0$".
  - **else**
  - Output "$\mathbf{c} \cdot \mathbf{k} = 1$".
  - **End If**
- **Else**
  - **If** $\hat{m} > \frac{\nu}{2}$
  - Output "$\mathbf{c} \cdot \mathbf{k} = 1$".
  - **else**
  - Output "$\mathbf{c} \cdot \mathbf{k} = 0$".
  - **End If**
- **End If**

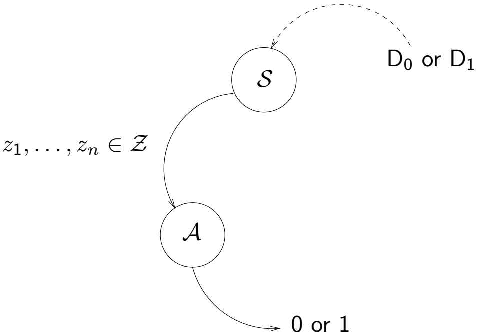
Distinguishing Two Probability Distributions



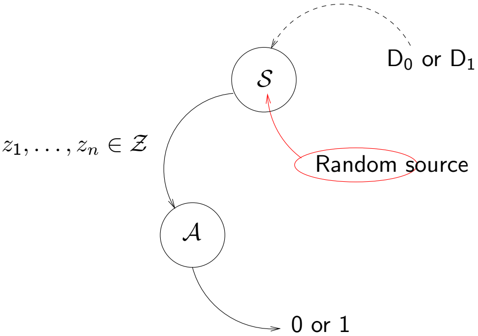
Information Extraction About the Key (3)

- In the order of ε^{-2} plaintext-ciphertext pairs are sufficient to get the bit $\mathbf{c} \cdot \mathbf{k}$ with high success probability.
- Are ε^{-2} plaintext-ciphertext pairs necessary ?
- Do we fully exploit the statistical information we have at disposal?

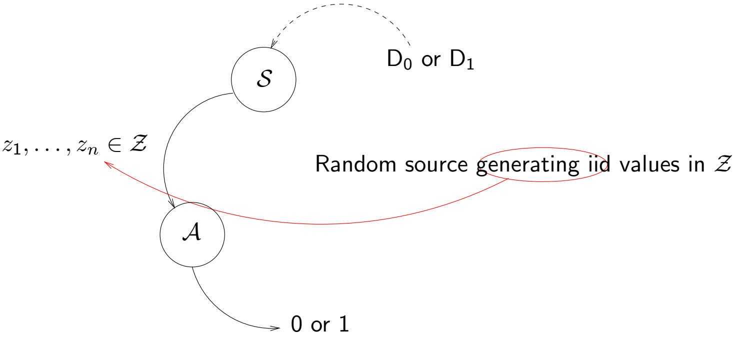
Statistical Hypothesis Tests (1)



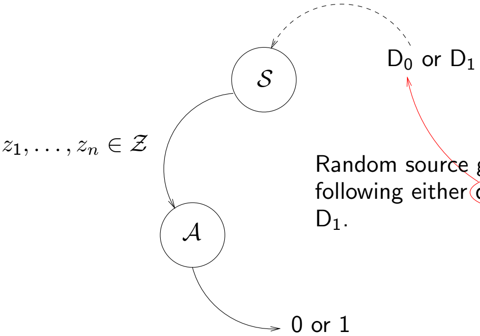
Statistical Hypothesis Tests (1)



Statistical Hypothesis Tests (1)

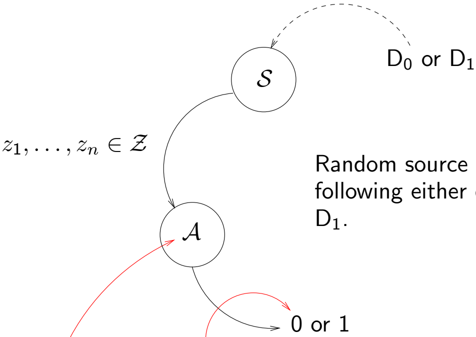


Statistical Hypothesis Tests (1)



Random source generating iid values in $\mathcal{Z}$ following either distribution D_0 or D_1.

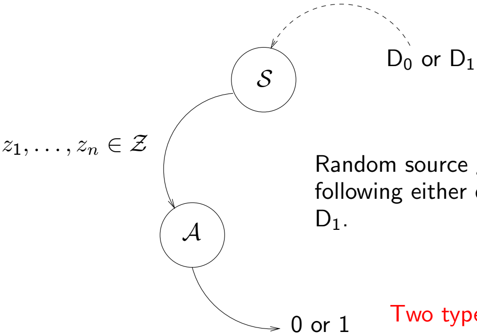
Statistical Hypothesis Tests (1)



Random source generating iid values in $\mathcal{Z}$ following either distribution D_0 or D_1.

The distinguisher decides whether $D = D_0$ or $D = D_1$

Statistical Hypothesis Tests (1)

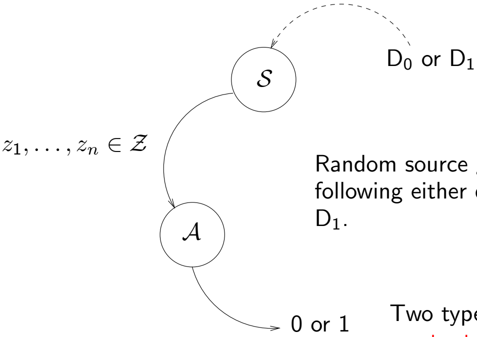


Random source generating iid values in $\mathcal{Z}$ following either distribution D_0 or D_1.

Two types of mistake:

The distinguisher decides whether $\mathsf{D} = \mathsf{D}_0$ or $\mathsf{D} = \mathsf{D}_1$

Statistical Hypothesis Tests (1)



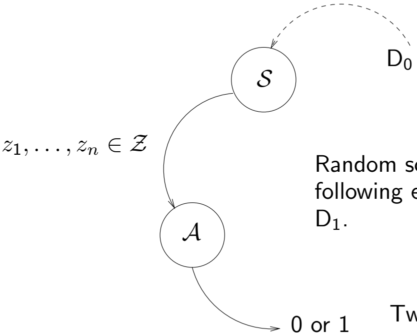
Random source generating iid values in $\mathcal{Z}$ following either distribution D_0 or D_1.

The distinguisher decides whether $D = D_0$ or $D = D_1$

Two types of mistake:
- output 0 when $D = D_1$ (probability α)
- output 1 when $D = D_0$ (probability β)

Statistical Hypothesis Tests (1)

Optimal distinguisher $\Leftrightarrow \mathsf{P_e} = \frac{1}{2}(\alpha + \beta)$ minimum

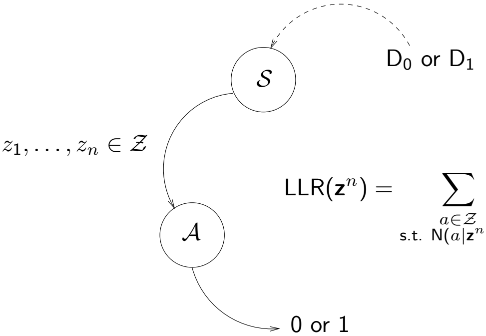


Random source generating iid values in $\mathcal{Z}$ following either distribution D$_0$ or D$_1$.

The distinguisher decides whether D $=$ D$_0$ or D $=$ D$_1$

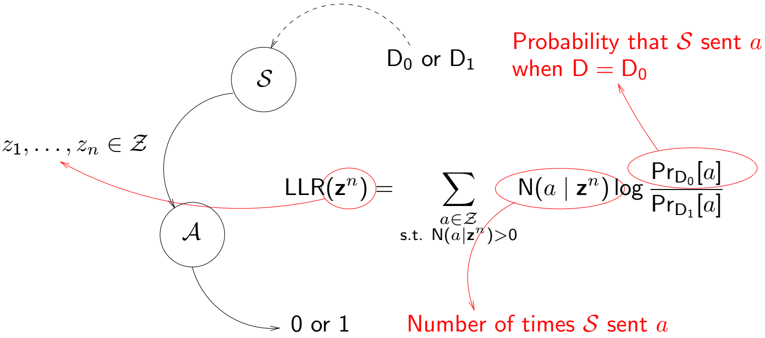
Two types of mistake:
- output 0 when D $=$ D$_1$ (probability α)
- output 1 when D $=$ D$_0$ (probability β)

Statistical Hypothesis Tests (2)

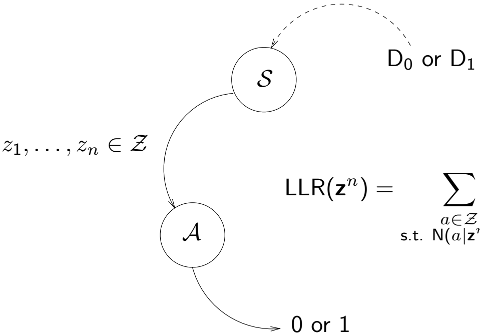


$$\mathsf{LLR}(\mathbf{z}^n) = \sum_{\substack{a \in \mathcal{Z} \\ \text{s.t. } \mathsf{N}(a|\mathbf{z}^n) > 0}} \mathsf{N}(a \mid \mathbf{z}^n) \log \frac{\Pr_{\mathsf{D}_0}[a]}{\Pr_{\mathsf{D}_1}[a]}$$

Statistical Hypothesis Tests (2)



Statistical Hypothesis Tests (2)



$$\mathsf{LLR}(\mathbf{z}^n) = \sum_{\substack{a \in \mathcal{Z} \\ \text{s.t. } \mathsf{N}(a|\mathbf{z}^n) > 0}} \mathsf{N}(a \mid \mathbf{z}^n) \log \frac{\Pr_{\mathsf{D}_0}[a]}{\Pr_{\mathsf{D}_1}[a]}$$

Optimal Rule:
choose 0 when $\mathsf{LLR}(\mathbf{z}^n) \geq 0$
choose 1 when $\mathsf{LLR}(\mathbf{z}^n) < 0$

This minimizes $\mathsf{P_e}$ $\Rightarrow$ optimal distinguisher
(aka Neyman-Pearson lemma)

Back to Linear Cryptanalysis

- We have to distinguish between two binomial laws, one with parameters ν and $p = \frac{1}{2} + \varepsilon$, the other with ν and $p = \frac{1}{2} - \varepsilon$, depending on the value of $\mathbf{c} \cdot \mathbf{k}$.

Theorem

For a fixed number ν of data queried to the oracle $\mathbf{\Omega}$, Matsui's First Algorithm is optimal in the sense that it maximizes the success probability over all algorithms based on the sample bit

$$\mathbf{a} \cdot \boldsymbol{X}_i \oplus \mathbf{b} \cdot \mathsf{f}_{\mathbf{k}}(\boldsymbol{X}_i).$$

Soft Decision About the Key

- Matsui's First Algorithm extract only one bit of information about the key.
- **Idea**: guess the subkey of the last round (or of the first round), partially decrypt (encrypt) the pair of plaintext-ciphertext, and check a biased linear approximation.
- Wrong subkey: equivalent to the encryption by one more round.
- Right subkey: we should observe a bias in the linear approximation.

Soft Decision About the Key(3)

- Matsui's Second Algorithm: consider the right subkey to be *the* one producing the largest experimental bias, and look for the remaining unknown key bits.
- Matsui's Third Algorithm: rank the subkey according to their experimental biases, and look for the remaining unknown key bits *until* the right one is found.

An Application of Optimal Distinguishers

- Best attack exploits *two* linear approximations
- Observed that Matsui's way to combine the statistical information was not optimal.
- Introduced the concept of optimal key-ranking procedure (valid for any statistical cryptanalysis) based on *statistical hypothesis tests*.
- Experimentally confirmed: when applied to DES, it allows to gain a factor of about two regarding the computational complexity.
- Results published in [Junod-Vaudenay, FSE'03]

An Application of Optimal Distinguishers

- Best attack exploits *two* linear approximations
- Observed that Matsui's way to combine the statistical information was not optimal.
- Introduced the concept of optimal key-ranking procedure (valid for any statistical cryptanalysis) based on *statistical hypothesis tests*.
- Experimentally confirmed: when applied to DES, it allows to gain a factor of about two regarding the computational complexity.
- Results published in [Junod-Vaudenay, FSE'03]

An Application of Optimal Distinguishers

- Best attack exploits *two* linear approximations
- Observed that Matsui's way to combine the statistical information was not optimal.
- Introduced the concept of optimal key-ranking procedure (valid for any statistical cryptanalysis) based on *statistical hypothesis tests*.
- Experimentally confirmed: when applied to DES, it allows to gain a factor of about two regarding the computational complexity.
- Results published in [Junod-Vaudenay, FSE'03]

An Application of Optimal Distinguishers

- Best attack exploits *two* linear approximations
- Observed that Matsui's way to combine the statistical information was not optimal.
- Introduced the concept of optimal key-ranking procedure (valid for any statistical cryptanalysis) based on *statistical hypothesis tests*.
- Experimentally confirmed: when applied to DES, it allows to gain a factor of about two regarding the computational complexity.
- Results published in [Junod-Vaudenay, FSE'03]

An Application of Optimal Distinguishers

- Best attack exploits *two* linear approximations
- Observed that Matsui's way to combine the statistical information was not optimal.
- Introduced the concept of optimal key-ranking procedure (valid for any statistical cryptanalysis) based on *statistical hypothesis tests*.
- Experimentally confirmed: when applied to DES, it allows to gain a factor of about two regarding the computational complexity.
- Results published in [Junod-Vaudenay, FSE'03]

Outline

Luby-Rackoff Security Approach

- Luby and Rackoff (1988): construction of a pseudo-random permutation out of pseudo-random functions (construction based on a Feistel scheme).
- An oracle $\mathbf{\Omega}$ implementing either a permutation C or a uniformly distributed random permutation C^*.
- Central notion : computationally unbounded distinguisher δ^ν limited to ν queries to $\mathbf{\Omega}$.
- We are interested in the advantage of δ^ν:

$$\mathrm{Adv}_{\delta^\nu}(\mathsf{C}, \mathsf{C}^*) = \left| \Pr_{\mathsf{C}}\left[\delta^\nu(\mathbf{x}) = 1\right] - \Pr_{\mathsf{C}^*}\left[\delta^\nu(\mathbf{x}) = 1\right] \right|$$

Luby-Rackoff Security Approach (2)

- Security proof $\equiv$ finding a good upper bound on $\mathsf{Adv}_{\delta^\nu}(\mathsf{C}, \mathsf{C}^*)$
- Strong model (because of the infinite computational ressources of the adversary)
- We can weaken it by restricting ourselves to certain classes of attacks.
- Adaptive vs. non-adaptive attacks

Iterated Distinguisher of Order 1

- Notion introduced by Vaudenay in 1999
- Non-adaptive distinguisher keeping a single bit of information about each pair of data
- We are interested in the simplest case: distinguishing two random sources.

Iterated Distinguisher of Order 1

- Notion introduced by Vaudenay in 1999
- Non-adaptive distinguisher keeping a single bit of information about each pair of data
- We are interested in the simplest case: distinguishing two random sources.

Iterated Distinguisher of Order 1

- Notion introduced by Vaudenay in 1999
- Non-adaptive distinguisher keeping a single bit of information about each pair of data
- We are interested in the simplest case: distinguishing two random sources.

Iterated Distinguisher of Order 1 (2)

Lemma (Vaudenay, 1999)

For any computationally unbounded distinguisher $\boldsymbol{\delta}^\nu$ limited to ν queries,

$$\mathsf{Adv}_{\boldsymbol{\delta}^\nu}(\mathsf{D}_0, \mathsf{D}_1) \leq 4|\varepsilon|\sqrt{\nu}$$

where D_0 is the uniform distribution on $\{0,1\}$ and D_1 is a probability distribution defined as $\mathsf{Pr}_{\mathsf{D}_1}[X = 0] = 1 - \mathsf{Pr}_{\mathsf{D}_1}[X = 1] = \frac{1}{2} + \varepsilon$.

Iterated Distinguisher of Order 1 (3)

- Interpretation of a distinguishing problem as a statistical hypotheses test

Lemma

Let $\pi_e = \frac{1}{2}(\alpha + \beta)$ *denote the overall probability of error of a distinguisher* $\boldsymbol{\delta}$*. Then,*

$$\mathsf{Adv}_{\boldsymbol{\delta}}(\mathsf{C}, \mathsf{C}^*) = 1 - 2\pi_e = 1 - (\alpha + \beta).$$

- Description of optimal distinguishers by means of the likelihood-ratio

Iterated Distinguisher of Order 1 (4)

Theorem

For any computationally unbounded optimal iterated distinguisher $\boldsymbol{\delta}^\nu$ of order 1 limited to ν queries,

$$1 - \frac{(\nu+1)}{2^{\nu\gamma-1}} \leq \mathsf{Adv}_{\boldsymbol{\delta}^\nu_{\mathsf{lin}}}(\mathsf{D}_0, \mathsf{D}_1) \leq 1 - \frac{1}{(\nu+1)\cdot 2^{\nu\gamma-1}}$$

where $\gamma = C(\mathsf{D}_0, \mathsf{D}_1)$ is the Chernoff information between D_0, the uniform distribution on $\{0,1\}$ and D_1, a probability distribution defined as $\Pr_{\mathsf{D}_1}[X=0] = 1 - \Pr_{\mathsf{D}_1}[X=1] = \frac{1}{2} + \varepsilon$ with

$$C(\mathsf{D}_0, \mathsf{D}_1) = -\min_{0\leq\lambda\leq 1} \log_2 \left(\sum_{x\in\mathcal{X}} \Pr_{X_0}[x]^\lambda \Pr_{X_1}[x]^{1-\lambda} \right).$$

Iterated Distinguisher of Order 1 (5)

- Proof of the asymptotic behaviour of an optimal distinguisher using (a slightly adapted version of) Chernoff's theorem
- Tighter bounds have been derived as well.
- Bounds have been adapted to linear and differential distinguishers.
- Results published in [Junod, Eurocrypt'03]

Iterated Distinguisher of Order 1 (5)

- Proof of the asymptotic behaviour of an optimal distinguisher using (a slightly adapted version of) Chernoff's theorem
- Tighter bounds have been derived as well.
- Bounds have been adapted to linear and differential distinguishers.
- Results published in [Junod, Eurocrypt'03]

Iterated Distinguisher of Order 1 (5)

- Proof of the asymptotic behaviour of an optimal distinguisher using (a slightly adapted version of) Chernoff's theorem
- Tighter bounds have been derived as well.
- Bounds have been adapted to linear and differential distinguishers.
- Results published in [Junod, Eurocrypt'03]

Iterated Distinguisher of Order 1 (5)

- Proof of the asymptotic behaviour of an optimal distinguisher using (a slightly adapted version of) Chernoff's theorem
- Tighter bounds have been derived as well.
- Bounds have been adapted to linear and differential distinguishers.
- Results published in [Junod, Eurocrypt'03]

Disgression

- Measures between discrete probability distributions: $||.||_1$, $||.||_2$, Chernoff exponent.
- $||.||_1$ is linked to the advantage.
- $||.||_2$ is linked to the number of necessary samples in a known-plaintext attack.
- Chernoff exponent is linked to the asymptotic behaviour of the advantage during a known-plaintext attack

Outline

Generalized Linear Cryptanalysis

- Idea : can we generalize classical linear cryptanalysis to linear approximations on bigger finite fields?
- Typically, by increasing the probability space cardinality, we may expect more distinguishing power...
- Instead of a linear approximation from GF(2) to GF(2), can we think about something from GF(2^ℓ) to GF($2^{\ell'}$) for $\ell, \ell' > 1$?

Generalized Linear Cryptanalysis

- Idea : can we generalize classical linear cryptanalysis to linear approximations on bigger finite fields?
- Typically, by increasing the probability space cardinality, we may expect more distinguishing power...
- Instead of a linear approximation from GF(2) to GF(2), can we think about something from GF(2^{ℓ}) to GF($2^{\ell'}$) for $\ell, \ell' > 1$?

Generalized Linear Cryptanalysis

- Idea : can we generalize classical linear cryptanalysis to linear approximations on bigger finite fields?
- Typically, by increasing the probability space cardinality, we may expect more distinguishing power...
- Instead of a linear approximation from GF(2) to GF(2), can we think about something from GF(2^{ℓ}) to GF($2^{\ell'}$) for $\ell, \ell' > 1$?

Generalized Linear Cryptanalysis (2)

- Paper [Baignères-Junod-Vaudenay, Asiacrypt'04]
  - Definition of optimal distinguishers on discrete spaces of any cardinality.
  - Computation of the necessary amount of samples
  - Ciphers protected against classical linear cryptanalysis are somewhat protected against GF(2)-linear approximations.

Generalized Linear Cryptanalysis (3)

- Let D_0 and D_1 be two discrete probability distributions sharing the same support. We assume that

$$\forall z \in \mathcal{Z} \qquad \Pr_{\mathsf{D}_0}[z] = \pi_z \text{ and } \Pr_{\mathsf{D}_1}[z] = \pi_z + \varepsilon_z \text{ with } |\varepsilon_z| \ll \pi_z.$$

- Measure of "bias": Let $\varepsilon_z = \Pr_{\mathsf{D}_1}[z] - \frac{1}{|\mathcal{Z}|}$. The Squared Euclidean Imbalance (SEI) $\Delta(\mathsf{D}_1)$ of a distribution D_1 of support $\mathcal{Z}$ from the uniform distribution is defined by

$$\Delta(\mathsf{D}_1) = |\mathcal{Z}| \sum_{z \in \mathcal{Z}} \varepsilon_z^2.$$

Outline

Link to χ^2 attacks

In a χ^2 cryptanalysis, the adversary does not need to know D_0, i.e., what exactly happens in the inner transformations of the cipher (which can therefore be considered as a *black box*).

- $$\hat{\chi}^2 = \sum_{i=1}^{m} \frac{\left(\hat{x}_i - np_i(\bar{\theta})\right)^2}{np_i(\bar{\theta})}$$

Link to χ^2 attacks

In a χ^2 cryptanalysis, the adversary does not need to know D_0, i.e., what exactly happens in the inner transformations of the cipher (which can therefore be considered as a *black box*).

- $$\hat{\chi}^2 = \sum_{i=1}^{m} \frac{\left(\hat{x}_i - np_i(\bar{\theta})\right)^2}{np_i(\bar{\theta})}$$

Link to χ^2 attacks (2)

- Complexity of a χ^2 attack $\longrightarrow O(1/\Delta(\mathsf{D}_1))$
- Not worse (up to a constant term) than an optimal distinguisher.

Link to χ^2 attacks (2)

- Complexity of a χ^2 attack $\longrightarrow O(1/\Delta(\mathsf{D}_1))$
- Not worse (up to a constant term) than an optimal distinguisher.

Link to χ^2 attacks (3)

Observation

When one does not know precisely what happens in the attacked cipher, the best practical alternative to an optimal distinguisher seems to be the χ^2 attack.

Summary

- Very old and simple results in statistics still not fully exploited in 2004 in the crypto field.
- Theoretically, one could always describe an optimal distinguisher (but we still have to compute the underlying probability distributions...)
- More applications?

Merci !