How Far Can We Go Beyond Linear Cryptanalysis?

T. Baignères P. Junod S. Vaudenay

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

ASIACRYPT 2004

Outline

Introduction

Original Motivation

To give a generalization of linear cryptanalysis.

Result

The paper turns out to propose a very general statistical framework which can be used to construct and study optimal distinguishers, and to generalize the fundamental concepts behind linear cryptanalysis.

Previous Work

The original linear cryptanalysis was proposed by Matsui at EUROCRYPT'93. Since then, several generalizations have been proposed.

- Kaliski and Robshaw used multiple linear approximations,
- Vaudenay proposed the χ^2 attack, where a cipher can simply be considered as a black box,
- Harpes, Kramer, and Massey replaced linear expressions with I/O sums,
- Harpes and Massey considered partition pairs of the input and output spaces of the cipher,
- More recently, Junod and Vaudenay considered linear cryptanalysis in a purely statistical framework.

Previous Work

The original linear cryptanalysis was proposed by Matsui at EUROCRYPT'93. Since then, several generalizations have been proposed.

- Kaliski and Robshaw used multiple linear approximations,
- Vaudenay proposed the χ^2 attack, where a cipher can simply be considered as a black box,
- Harpes, Kramer, and Massey replaced linear expressions with I/O sums,
- Harpes and Massey considered partition pairs of the input and output spaces of the cipher,
- More recently, Junod and Vaudenay considered linear cryptanalysis in a purely statistical framework.

Previous Work

The original linear cryptanalysis was proposed by Matsui at EUROCRYPT'93. Since then, several generalizations have been proposed.

- Kaliski and Robshaw used multiple linear approximations,
- Vaudenay proposed the χ^2 attack, where a cipher can simply be considered as a black box,
- Harpes, Kramer, and Massey replaced linear expressions with I/O sums,
- Harpes and Massey considered partition pairs of the input and output spaces of the cipher,
- More recently, Junod and Vaudenay considered linear cryptanalysis in a purely statistical framework.

Previous Work

The original linear cryptanalysis was proposed by Matsui at EUROCRYPT'93. Since then, several generalizations have been proposed.

- Kaliski and Robshaw used multiple linear approximations,
- Vaudenay proposed the χ^2 attack, where a cipher can simply be considered as a black box,
- Harpes, Kramer, and Massey replaced linear expressions with I/O sums,
- Harpes and Massey considered partition pairs of the input and output spaces of the cipher,
- More recently, Junod and Vaudenay considered linear cryptanalysis in a purely statistical framework.

Previous Work

The original linear cryptanalysis was proposed by Matsui at EUROCRYPT'93. Since then, several generalizations have been proposed.

- Kaliski and Robshaw used multiple linear approximations,
- Vaudenay proposed the χ^2 attack, where a cipher can simply be considered as a black box,
- Harpes, Kramer, and Massey replaced linear expressions with I/O sums,
- Harpes and Massey considered partition pairs of the input and output spaces of the cipher,
- More recently, Junod and Vaudenay considered linear cryptanalysis in a purely statistical framework.

Previous Work

The original linear cryptanalysis was proposed by Matsui at EUROCRYPT'93. Since then, several generalizations have been proposed.

- Kaliski and Robshaw used multiple linear approximations,
- Vaudenay proposed the χ^2 attack, where a cipher can simply be considered as a black box,
- Harpes, Kramer, and Massey replaced linear expressions with I/O sums,
- Harpes and Massey considered partition pairs of the input and output spaces of the cipher,
- More recently, Junod and Vaudenay considered linear cryptanalysis in a purely statistical framework.

Previous Work

and at CRYPTO'04 ...

- Biryukov, De Cannière, and Quisquater used multiple linear approximations in order to reduce attack complexities against DES,
- and Courtois showed how a cipher that was designed to resist LC could be broken by his bi-linear cryptanalysis.

Previous Work

and at CRYPTO'04 ...

- Biryukov, De Cannière, and Quisquater used multiple linear approximations in order to reduce attack complexities against DES,
- and Courtois showed how a cipher that was designed to resist LC could be broken by his bi-linear cryptanalysis.

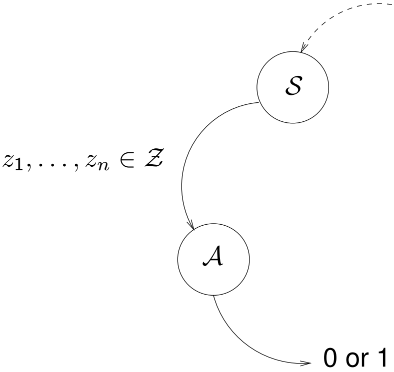
Previous Work

and at CRYPTO'04 ...

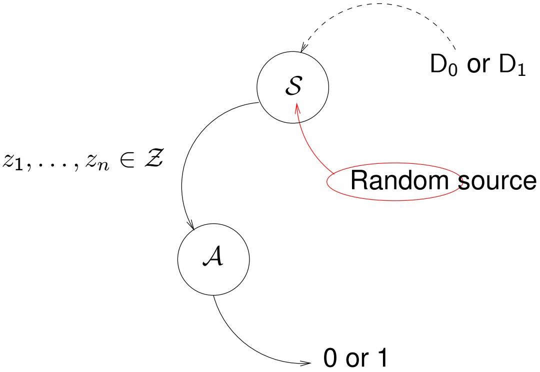
- Biryukov, De Cannière, and Quisquater used multiple linear approximations in order to reduce attack complexities against DES,
- and Courtois showed how a cipher that was designed to resist LC could be broken by his bi-linear cryptanalysis.

Outline

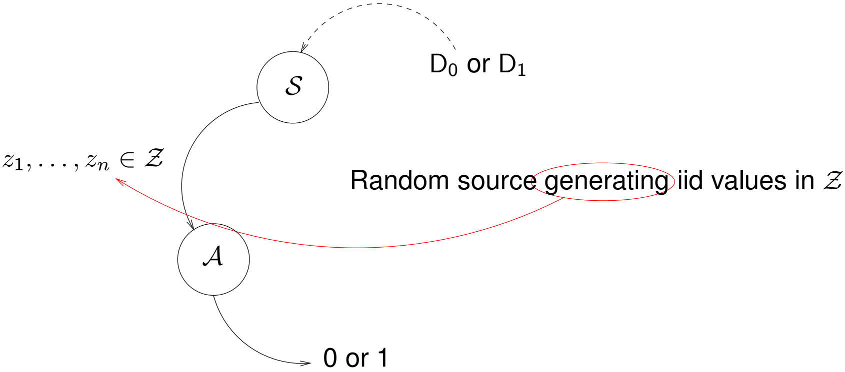
General case (1)



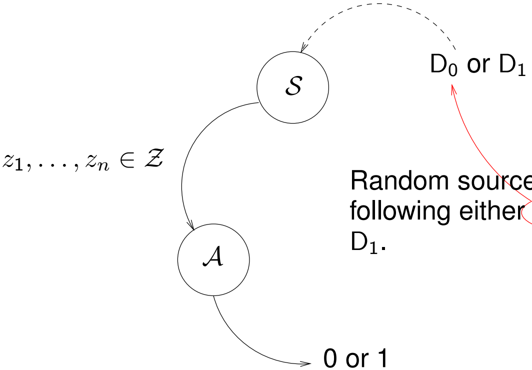
General case (1)



General case (1)

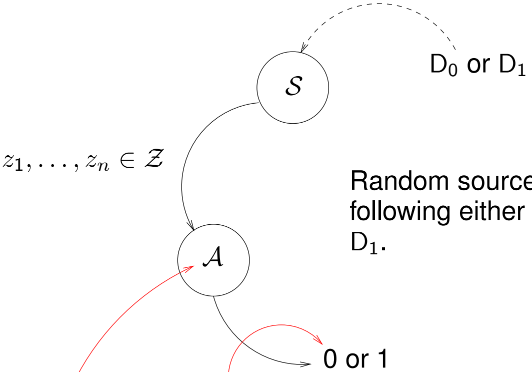


General case (1)



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

General case (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 $D = D_1$

General case (1)

D_0 or D_1

$\mathcal{S}$

$z_1, \ldots, z_n \in \mathcal{Z}$

$\mathcal{A}$

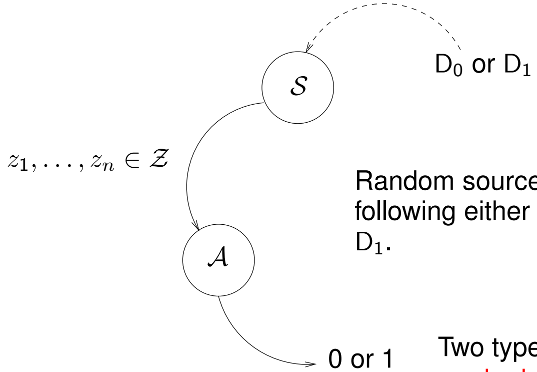
0 or 1

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

Two types of mistake:

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

General case (1)



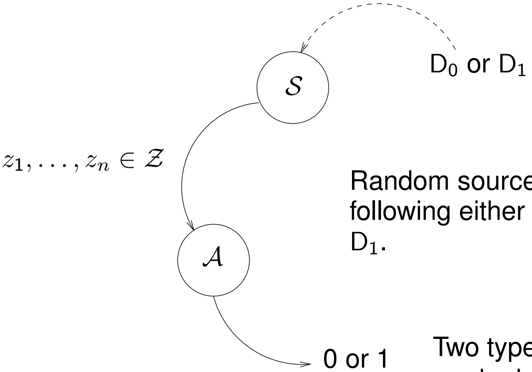
Random source generating iid values in $\mathcal{Z}$ following either distribution D$_0$ or D$_1$.

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

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

General case (1)

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

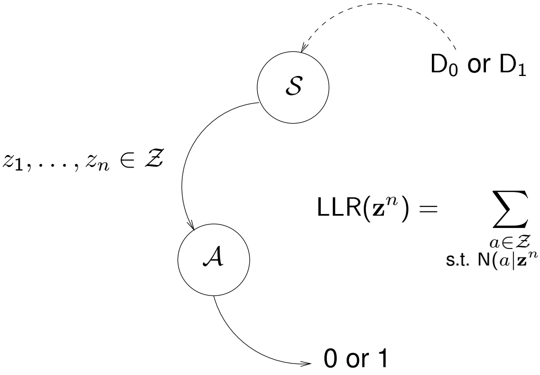


Random source generating iid values in $\mathcal{Z}$ following either distribution D_0 or D_1.

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

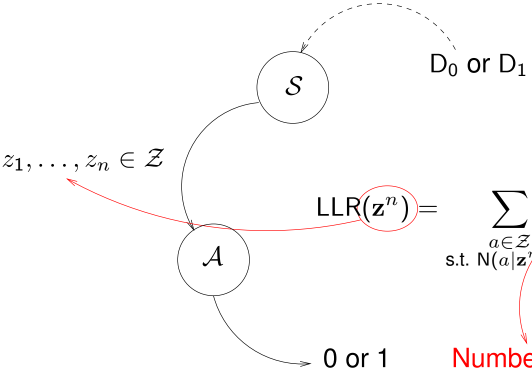
Two types of mistake:
- output 0 when $\mathsf{D} = \mathsf{D}_1$ (probability α)
- output 1 when $\mathsf{D} = \mathsf{D}_0$ (probability β)

General case (2)



$$\mathrm{LLR}(\mathbf{z}^n) = \sum_{\substack{a \in \mathcal{Z} \\ \text{s.t. } \mathrm{N}(a|\mathbf{z}^n) > 0}} \mathrm{N}(a \mid \mathbf{z}^n) \log \frac{\Pr_{D_0}[a]}{\Pr_{D_1}[a]}$$

General case (2)

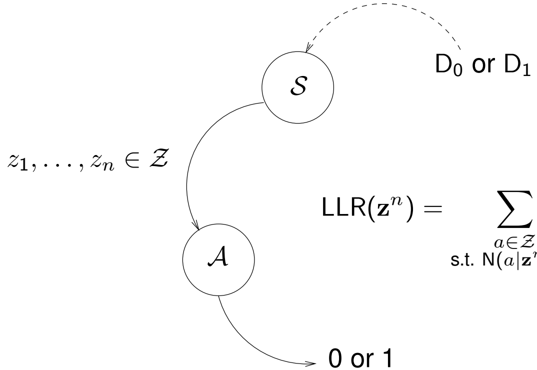


Probability that $\mathcal{S}$ sent a when $\mathsf{D} = \mathsf{D}_0$

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

Number of times $\mathcal{S}$ sent a

General case (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)

General case (3)

For a given $\mathsf{P_e}$, how many queries does the distinguisher need?

Theorem

Considering that

- $Z_1, \ldots, Z_n$ are iid, following distribution $\mathsf{D} \in \{\mathsf{D}_0, \mathsf{D}_1\}$,
- D_0 is close to D_1, i.e., $\Pr_{\mathsf{D}_0}[z] - \Pr_{\mathsf{D}_1}[z] = \epsilon_z \ll 1$,

$$n = \frac{d}{\sum_{z \in \mathcal{Z}} \frac{\epsilon_z^2}{p_z}} \qquad \text{with} \qquad \mathsf{P_e} \approx 1 - \Phi\left(\frac{\sqrt{d}}{2}\right) .$$

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2} du .$$

General case (3)

For a given $\mathsf{P_e}$, how many queries does the distinguisher need?

Theorem

Considering that

- $Z_1, \ldots, Z_n$ are iid, following distribution $\mathsf{D} \in \{\mathsf{D}_0, \mathsf{D}_1\}$,
- D_0 is close to D_1, i.e., $\Pr_{\mathsf{D}_0}[z] - \Pr_{\mathsf{D}_1}[z] = \epsilon_z \ll 1$,

$$n = \frac{d}{\sum_{z \in \mathcal{Z}} \frac{\epsilon_z^2}{p_z}} \qquad \text{with} \qquad \mathsf{P_e} \approx 1 - \Phi\left(\frac{\sqrt{d}}{2}\right) .$$

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2}\, du .$$

General case (3)

For a given $\mathsf{P_e}$, how many queries does the distinguisher need?

Theorem

Considering that

- $Z_1, \ldots, Z_n$ are iid, following distribution $\mathsf{D} \in \{\mathsf{D}_0, \mathsf{D}_1\}$,
- D_0 is close to D_1, i.e., $\Pr_{\mathsf{D}_0}[z] - \Pr_{\mathsf{D}_1}[z] = \epsilon_z \ll 1$,

$$n = \frac{d}{\sum_{z \in \mathcal{Z}} \frac{\epsilon_z^2}{p_z}} \quad \text{with} \quad \mathsf{P_e} \approx 1 - \Phi\left(\frac{\sqrt{d}}{2}\right) .$$

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2} du .$$

General case (3)

For a given $\mathsf{P_e}$, how many queries does the distinguisher need?

Theorem

Considering that

- $Z_1, \ldots, Z_n$ are iid, following distribution $\mathsf{D} \in \{\mathsf{D}_0, \mathsf{D}_1\}$,
- D_0 is close to D_1, i.e., $\Pr_{\mathsf{D}_0}[z] - \Pr_{\mathsf{D}_1}[z] = \epsilon_z \ll 1$,

$$n = \frac{d}{\sum_{z \in \mathcal{Z}} \frac{\epsilon_z^2}{p_z}} \qquad \text{with} \qquad \mathsf{P_e} \approx 1 - \Phi\left(\frac{\sqrt{d}}{2}\right) \quad .$$

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2} du \quad .$$

Outline

One source following a uniform distribution

Squared Euclidean Imbalance (SEI)

If D_1 is the uniform distribution (i.e., $\Pr_{\mathsf{D}_1}[z] = p_z = \frac{1}{|\mathcal{Z}|}$), we define the Squared Euclidean Imbalance (SEI):

$$\Delta(\mathsf{D}_0) = |\mathcal{Z}| \sum_{z \in \mathcal{Z}} \epsilon_z^2 \ .$$

Corollary

$$n = \frac{d}{\Delta(\mathsf{D}_0)} \qquad \text{with} \qquad \mathsf{P_e} \approx 1 - \Phi\left(\frac{\sqrt{d}}{2}\right) \ .$$

$\Rightarrow$ The complexity of distinguishing D_0 from D_1 can be measured by means of the SEI.

Link to χ^2 attacks

In a χ^2 cryptanalysis, the adversary does not need to know D_1, i.e., what exactly happens in the inner transformations of the cipher (which can therefore be considered as a *black box*).

- Complexity of a χ^2 attack $\rightarrow O(1/\Delta(\mathrm{D}_0))$
- Not worse (up to a constant term) than an optimal distinguisher.

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.

Link to χ^2 attacks

In a χ^2 cryptanalysis, the adversary does not need to know D_1, i.e., what exactly happens in the inner transformations of the cipher (which can therefore be considered as a *black box*).

- Complexity of a χ^2 attack $\rightarrow O(1/\Delta(\mathsf{D}_0))$
- Not worse (up to a constant term) than an optimal distinguisher.

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.

Link to χ^2 attacks

In a χ^2 cryptanalysis, the adversary does not need to know D_1, i.e., what exactly happens in the inner transformations of the cipher (which can therefore be considered as a *black box*).

- Complexity of a χ^2 attack $\rightarrow O(1/\Delta(\mathrm{D}_0))$
- Not worse (up to a constant term) than an optimal distinguisher.

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.

Link to χ^2 attacks

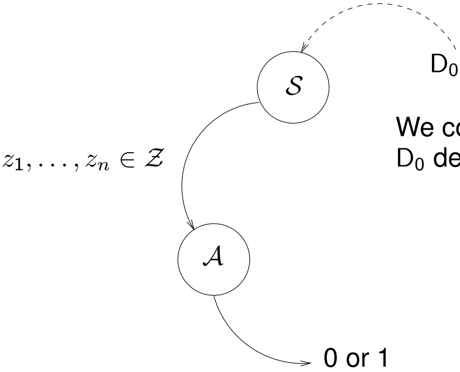
In a χ^2 cryptanalysis, the adversary does not need to know D_1, i.e., what exactly happens in the inner transformations of the cipher (which can therefore be considered as a *black box*).

- Complexity of a χ^2 attack $\rightarrow O(1/\Delta(\mathrm{D}_0))$
- Not worse (up to a constant term) than an optimal distinguisher.

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.

Outline

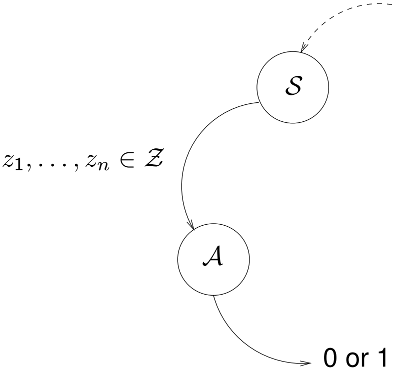
Source of random bit strings (1)



D_0 or D_1

We consider $\mathcal{Z} = \{0, 1\}^\ell$
D_0 defined by the set $\{\epsilon_z\}_{z \in \mathcal{Z}}$

Source of random bit strings (1)



D_0 or D_1

We consider $\mathcal{Z} = \{0,1\}^{\ell}$
D_0 defined by the set $\{\epsilon_z\}_{z \in \mathcal{Z}}$

Fourier transform of D_0 at point $u \in \mathcal{Z}$:

$$\widehat{\epsilon}_u = \sum_{z \in \mathcal{Z}} (-1)^{u \cdot z} \epsilon_z$$

Source of random bit strings (2)

Properties of the SEI (using the Fourier transform):

- $\Delta(\mathrm{D}_0) = \sum_{u \in \mathcal{Z}} \widehat{\epsilon}_u^{\,2}$
- When B is a random bit, recall the linear probability is $\mathrm{LP}(B) = (2\Pr[B = 0] - 1)^2$. Then,
  - $\Delta(\mathrm{D}_0) = \sum_{w \in \mathcal{Z} \setminus \{0\}} \mathrm{LP}(w \cdot Z)$
  - with $\mathrm{LP}^Z_{\max} = \max_{w \in \mathcal{Z} \setminus \{0\}} \mathrm{LP}(w \cdot Z)$,

$$\Delta(\mathrm{D}_0) \leq (2^\ell - 1)\mathrm{LP}^Z_{\max} .$$

Source of random bit strings (2)

Properties of the SEI (using the Fourier transform):

- $\Delta(\mathsf{D}_0) = \sum_{u \in \mathcal{Z}} \widehat{\epsilon}_u^{\,2}$
- When B is a random bit, recall the linear probability is $\mathsf{LP}(B) = (2\Pr[B = 0] - 1)^2$. Then,
  - $\Delta(\mathsf{D}_0) = \sum_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$
  - with $\mathsf{LP}^{Z}_{\max} = \max_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$,

$$\Delta(\mathsf{D}_0) \leq (2^{\ell} - 1)\mathsf{LP}^{Z}_{\max} .$$

Source of random bit strings (2)

Properties of the SEI (using the Fourier transform):

- $\Delta(\mathsf{D}_0) = \sum_{u \in \mathcal{Z}} \widehat{\epsilon}_u^{\,2}$
- When B is a random bit, recall the linear probability is $\mathsf{LP}(B) = (2\Pr[B = 0] - 1)^2$. Then,
- $\Delta(\mathsf{D}_0) = \sum_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$
- with $\mathsf{LP}^Z_{\max} = \max_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$,

$$\Delta(\mathsf{D}_0) \leq (2^\ell - 1)\mathsf{LP}^Z_{\max} .$$

Source of random bit strings (2)

Properties of the SEI (using the Fourier transform):

- $\Delta(\mathsf{D}_0) = \sum_{u \in \mathcal{Z}} \widehat{\epsilon}_u^{\,2}$
- When B is a random bit, recall the linear probability is $\mathsf{LP}(B) = (2\Pr[B = 0] - 1)^2$. Then,
  - $\Delta(\mathsf{D}_0) = \sum_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$
  - with $\mathsf{LP}^Z_{\max} = \max_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$,

$$\Delta(\mathsf{D}_0) \le (2^\ell - 1)\mathsf{LP}^Z_{\max} .$$

Source of random bit strings (2)

Properties of the SEI (using the Fourier transform):

- $\Delta(\mathsf{D}_0) = \sum_{u \in \mathcal{Z}} \widehat{\epsilon}_u^{\,2}$
- When B is a random bit, recall the linear probability is $\mathsf{LP}(B) = (2\Pr[B = 0] - 1)^2$. Then,
  - $\Delta(\mathsf{D}_0) = \sum_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$
  - with $\mathsf{LP}^Z_{\max} = \max_{w \in \mathcal{Z} \setminus \{0\}} \mathsf{LP}(w \cdot Z)$,

$$\Delta(\mathsf{D}_0) \leq (2^\ell - 1)\mathsf{LP}^Z_{\max} .$$

Outline

Statistical distinguishers

We know how to distinguish distributions in $\{0,1\}^\ell$ of *small cardinality* (i.e., ℓ is small).

What if the source generates variables in $\{0,1\}^L$ where L is *large*?

Solution:

- *reduce the sample space* by means of a projection:

$$h : \{0,1\}^L \longrightarrow \mathcal{Z} \quad .$$

- $Z = h(S) \in \mathcal{Z}$ follows either D_0 or D_1.

But how should we choose the projection h?!? (This may be where cryptanalysis becomes an art !)

Statistical distinguishers

We know how to distinguish distributions in $\{0,1\}^\ell$ of *small cardinality* (i.e., ℓ is small).

What if the source generates variables in $\{0,1\}^L$ where L is *large*?

Solution:

- *reduce the sample space* by means of a projection:

$$h : \{0,1\}^L \longrightarrow \mathcal{Z} \quad .$$

- $Z = h(S) \in \mathcal{Z}$ follows either D_0 or D_1.

But how should we choose the projection h?!? (This may be where cryptanalysis becomes an art !)

Statistical distinguishers

We know how to distinguish distributions in $\{0,1\}^\ell$ of *small cardinality* (i.e., ℓ is small).

What if the source generates variables in $\{0,1\}^L$ where L is *large*?

Solution:

- *reduce the sample space* by means of a projection:

$$h : \{0,1\}^L \longrightarrow \mathcal{Z} \quad .$$

- $Z = h(S) \in \mathcal{Z}$ follows either D_0 or D_1.

But how should we choose the projection h?!? (This may be where cryptanalysis becomes an art !)

Statistical distinguishers

We know how to distinguish distributions in $\{0,1\}^\ell$ of *small cardinality* (i.e., ℓ is small).

What if the source generates variables in $\{0,1\}^L$ where L is *large*?

Solution:

- *reduce the sample space* by means of a projection:

$$h : \{0,1\}^L \longrightarrow \mathcal{Z} \quad .$$

- $Z = h(S) \in \mathcal{Z}$ follows either D_0 or D_1.

But how should we choose the projection h?!? (This may be where cryptanalysis becomes an art !)

First example of a statistical distinguisher

For some non-zero $a \in \{0,1\}^L$

$$\begin{array}{rccc} h: & \{0,1\}^L & \longrightarrow & \mathcal{Z} = \{0,1\} \\ & S & \longmapsto & h(S) = a \cdot S \end{array} .$$

This is a linear distinguisher.

We note that $\Delta(h(S)) = \mathrm{LP}(a \cdot S) \leq \mathrm{LP}^S_{\max}$.

Modern ciphers have a bounded $\mathrm{LP}^S_{\max}$
$\Rightarrow$ protected against *linear cryptanalysis*.

First example of a statistical distinguisher

For some non-zero $a \in \{0,1\}^L$

$$\begin{array}{rccc} h: & \{0,1\}^L & \longrightarrow & \mathcal{Z} = \{0,1\} \\ & S & \longmapsto & h(S) = a \cdot S \end{array} .$$

This is a linear distinguisher.

We note that $\Delta(h(S)) = \mathrm{LP}(a \cdot S) \leq \mathrm{LP}^S_{\mathrm{max}}$.

Modern ciphers have a bounded $\mathrm{LP}^S_{\mathrm{max}}$
$\Rightarrow$ protected against *linear cryptanalysis*.

First example of a statistical distinguisher

For some non-zero $a \in \{0,1\}^L$

$$\begin{array}{rccc} h: & \{0,1\}^L & \longrightarrow & \mathcal{Z} = \{0,1\} \\ & S & \longmapsto & h(S) = a \cdot S \end{array} .$$

This is a linear distinguisher.

We note that $\Delta(h(S)) = \mathrm{LP}(a \cdot S) \leq \mathrm{LP}^S_{\mathrm{max}}$.

Modern ciphers have a bounded $\mathrm{LP}^S_{\mathrm{max}}$
$\Rightarrow$ protected against *linear cryptanalysis*.

Second example of a statistical distinguisher

$$\begin{array}{rccl} h: & \{0,1\}^L & \longrightarrow & \mathcal{Z} = \{0,1\}^\ell \\ & S & \longmapsto & h(S) \ . \end{array}$$

where h is GF (2)-linear and $\ell \neq 1$ is small.

Theorem

$$\Delta(h(S)) \leq (2^\ell - 1)\mathrm{LP}^S_{\max} \ .$$

Ciphers protected against linear cryptanalysis (bounded $\mathrm{LP}^S_{\max}$)
$\Rightarrow$ somewhat protected against several generalizations!

Second example of a statistical distinguisher

$$\begin{array}{rccc} h: & \{0,1\}^L & \longrightarrow & \mathcal{Z} = \{0,1\}^\ell \\ & S & \longmapsto & h(S) \quad . \end{array}$$

where h is GF (2)-linear and $\ell \neq 1$ is small.

Theorem

$$\Delta(h(S)) \leq (2^\ell - 1)\mathrm{LP}^S_{\max} \ .$$

Ciphers protected against linear cryptanalysis (bounded $\mathrm{LP}^S_{\max}$)
$\Rightarrow$ somewhat protected against several generalizations!

Second example of a statistical distinguisher

$$\begin{array}{rccc} h: & \{0,1\}^L & \longrightarrow & \mathcal{Z} = \{0,1\}^\ell \\ & S & \longmapsto & h(S)\ . \end{array}$$

where h is GF (2)-linear and $\ell \neq 1$ is small.

Theorem

$$\Delta(h(S)) \leq (2^\ell - 1)\mathrm{LP}^S_{\max}\ .$$

Ciphers protected against linear cryptanalysis (bounded $\mathrm{LP}^S_{\max}$)
$\Rightarrow$ somewhat protected against several generalizations!

Bounded $\mathrm{LP}^{S}_{\max}$ and low advantage are not equivalent!

Is it possible to find a distinguisher

- with a high advantage,
- even though the value of $\mathrm{LP}^{S}_{\max}$ is small?

Practical examples exist. For example

- Jakobsen and Knudsen's interpolation attack (where quadratic functions are used),
- Courtois' bi-linear cryptanalysis.

In the paper we provide an example of a source

- impossible to break with a linear distinguisher,
- trivially broken by a (well-chosen) non-linear distinguisher.

Bounded $\mathrm{LP}^S_{\max}$ and low advantage are not equivalent!

Is it possible to find a distinguisher

- with a high advantage,
- even though the value of $\mathrm{LP}^S_{\max}$ is small?

Practical examples exist. For example

- Jakobsen and Knudsen's interpolation attack (where quadratic functions are used),
- Courtois' bi-linear cryptanalysis.

In the paper we provide an example of a source

- impossible to break with a linear distinguisher,
- trivially broken by a (well-chosen) non-linear distinguisher.

Bounded $\mathrm{LP}^{S}_{\max}$ and low advantage are not equivalent!

Is it possible to find a distinguisher

- with a high advantage,
- even though the value of $\mathrm{LP}^{S}_{\max}$ is small?

Practical examples exist. For example

- Jakobsen and Knudsen's interpolation attack (where quadratic functions are used),
- Courtois' bi-linear cryptanalysis.

In the paper we provide an example of a source

- impossible to break with a linear distinguisher,
- trivially broken by a (well-chosen) non-linear distinguisher.

Bounded $\mathrm{LP}^{S}_{\max}$ and low advantage are not equivalent!

Is it possible to find a distinguisher

- with a high advantage,
- even though the value of $\mathrm{LP}^{S}_{\max}$ is small?

Practical examples exist. For example

- Jakobsen and Knudsen's interpolation attack (where quadratic functions are used),
- Courtois' bi-linear cryptanalysis.

In the paper we provide an example of a source

- impossible to break with a linear distinguisher,
- trivially broken by a (well-chosen) non-linear distinguisher.

Bounded $\mathrm{LP}^{S}_{\mathrm{max}}$ and low advantage are not equivalent!

Is it possible to find a distinguisher

- with a high advantage,
- even though the value of $\mathrm{LP}^{S}_{\mathrm{max}}$ is small?

Practical examples exist. For example

- Jakobsen and Knudsen's interpolation attack (where quadratic functions are used),
- Courtois' bi-linear cryptanalysis.

In the paper we provide an example of a source

- impossible to break with a linear distinguisher,
- trivially broken by a (well-chosen) non-linear distinguisher.

Bounded $\mathrm{LP}^{S}_{\max}$ and low advantage are not equivalent!

Is it possible to find a distinguisher

- with a high advantage,
- even though the value of $\mathrm{LP}^{S}_{\max}$ is small?

Practical examples exist. For example

- Jakobsen and Knudsen's interpolation attack (where quadratic functions are used),
- Courtois' bi-linear cryptanalysis.

In the paper we provide an example of a source

- impossible to break with a linear distinguisher,
- trivially broken by a (well-chosen) non-linear distinguisher.

Bounded $\mathrm{LP}^S_{\max}$ and low advantage are not equivalent!

Is it possible to find a distinguisher

- with a high advantage,
- even though the value of $\mathrm{LP}^S_{\max}$ is small?

Practical examples exist. For example

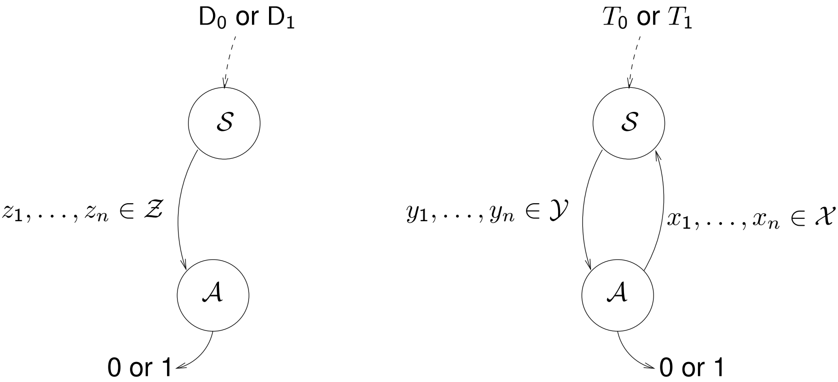
- Jakobsen and Knudsen's interpolation attack (where quadratic functions are used),
- Courtois' bi-linear cryptanalysis.

In the paper we provide an example of a source

- impossible to break with a linear distinguisher,
- trivially broken by a (well-chosen) non-linear distinguisher.

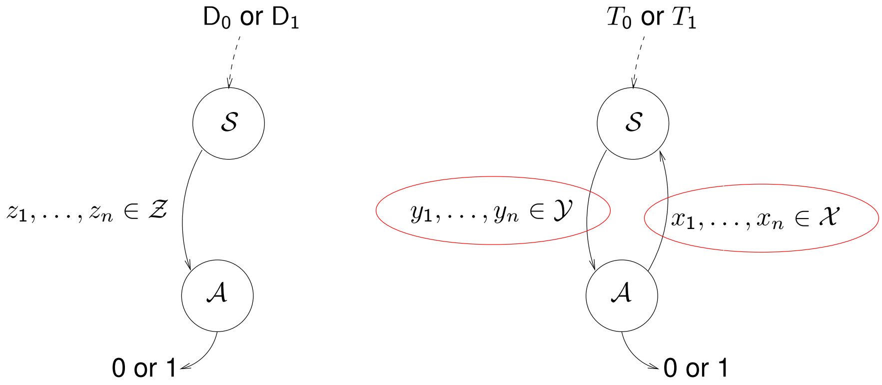
Outline

Beyond linear probabilities and linear expressions (1)



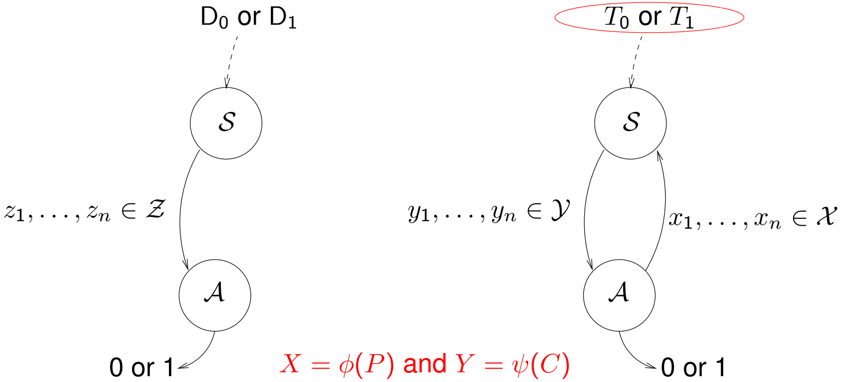
We know how to distinguish random sources.
$\rightarrow$ what about random oracles?

Beyond linear probabilities and linear expressions (1)



$Z \in \mathcal{Z}$ becomes a couple of random variables $(X, Y) \in \mathcal{X} \times \mathcal{Y}$.

Beyond linear probabilities and linear expressions (1)



$X = \phi(P)$ and $Y = \psi(C)$

known plaintext attack $\rightarrow P \sim$ uniform distrib. $\rightarrow X \sim$ uniform distrib.

Distribution of Y defined by a transition matrix:
$[T]_{x,y} = \Pr[Y = y \mid X = x]$

Beyond linear probabilities and linear expressions (2)

Transition Matrix

$$[T]_{x,y} = \Pr\left[Y = y \mid X = x\right] .$$

When $T = T_1$, Y is uniformly distributed.

Bias Matrix

$$B = T_0 - T_1 .$$

Link between bias matrix and SEI

$$\Delta(\mathsf{D}_0) = \frac{|\mathcal{Y}|}{|\mathcal{X}|} \parallel B \parallel_2^2 .$$

Beyond linear probabilities and linear expressions (2)

Transition Matrix

$$[T]_{x,y} = \Pr\left[Y = y \mid X = x\right] .$$

When $T = T_1$, Y is uniformly distributed.

Bias Matrix

$$B = T_0 - T_1 .$$

Link between bias matrix and SEI

$$\Delta(\mathsf{D}_0) = \frac{|\mathcal{Y}|}{|\mathcal{X}|} \parallel B \parallel_2^2 .$$

Beyond linear probabilities and linear expressions (2)

Transition Matrix

$$[T]_{x,y} = \Pr\left[Y = y \mid X = x\right] \ .$$

When $T = T_1$, Y is uniformly distributed.

Bias Matrix

$$B = T_0 - T_1 \ .$$

Link between bias matrix and SEI

$$\Delta(\mathsf{D}_0) = \frac{|\mathcal{Y}|}{|\mathcal{X}|} \parallel B \parallel_2^2 \ .$$

Outline

Piling-up transition matrices

$$
\begin{array}{ccc}
P^{(1)} & \xrightarrow{\quad\phi\quad} & X \\
\mathsf{C}^{(1)} \big\downarrow & & \big\downarrow T^{(1)} \\
P^{(2)} & \xrightarrow{\quad\xi\quad} & W \\
\mathsf{C}^{(2)} \big\downarrow & & \big\downarrow T^{(2)} \\
P^{(3)} & \xrightarrow{\quad\psi\quad} & Y
\end{array}
$$

Piling-up transition matrices

$$
\begin{tikzcd}
P^{(1)} \arrow[r, "\phi"] \arrow[d, "\mathsf{C}^{(1)}"'] & X \arrow[d, "T^{(1)}"] \arrow[dd, bend left, "T"] \\
P^{(2)} \arrow[r, "\xi"] \arrow[d, "\mathsf{C}^{(2)}"'] & W \arrow[d, "T^{(2)}"] \\
P^{(3)} \arrow[r, "\psi"] & Y
\end{tikzcd}
$$

If $X \leftrightarrow W \leftrightarrow Y$ is a Markov chain

$$T = T^{(1)} \times T^{(2)}$$

Piling-up transition matrices

$$
\begin{tikzcd}
P^{(1)} \arrow[r, "\phi"] \arrow[d, "\mathsf{C}^{(1)}"'] & X \arrow[d, "T^{(1)}"] \arrow[dd, bend left, "T"] \\
P^{(2)} \arrow[r, "\xi"] \arrow[d, "\mathsf{C}^{(2)}"'] & W \arrow[d, "T^{(2)}"] \\
P^{(3)} \arrow[r, "\psi"] & Y
\end{tikzcd}
$$

If $X \leftrightarrow W \leftrightarrow Y$ is a Markov chain

$$T = T^{(1)} \times T^{(2)} \rightsquigarrow B = B^{(1)} \times B^{(2)} \rightsquigarrow \| B \|_2 \leq \| B^{(1)} \|_2 \times \| B^{(2)} \|_2$$

Outline

Key recovery attacks

The framework can be adapted to key recovery.

In the paper we show how to build an optimal key ranking procedure that recovers a k bits key provided that the number of samples n is s.t.

$$n \geq \frac{4k \log 2}{\Delta(\mathsf{D}_0)} .$$

This formula was used to estimate the complexity of attacks against E0 (don't miss this morning's last talk!!).

Key recovery attacks

The framework can be adapted to key recovery.

In the paper we show how to build an optimal key ranking procedure that recovers a k bits key provided that the number of samples n is s.t.

$$n \geq \frac{4k \log 2}{\Delta(\mathsf{D}_0)} .$$

This formula was used to estimate the complexity of attacks against E0 (don't miss this morning's last talk!!).

Key recovery attacks

The framework can be adapted to key recovery.

In the paper we show how to build an optimal key ranking procedure that recovers a k bits key provided that the number of samples n is s.t.

$$n \geq \frac{4k \log 2}{\Delta(\mathsf{D}_0)}.$$

This formula was used to estimate the complexity of attacks against E0 (don't miss this morning's last talk!!).

Conclusion

- We defined a rigorous statistical framework in order to interpret LC and its generalizations in a unified way.
- Modern block ciphers are proven resistant against LC.
- This resistance extends to linear generalizations of LC,
- …but definitely not to non-linear ones!

Thank you for your attention!

Conclusion

- We defined a rigorous statistical framework in order to interpret LC and its generalizations in a unified way.
- Modern block ciphers are proven resistant against LC.
- This resistance extends to linear generalizations of LC,
- . . . but definitely not to non-linear ones!

Thank you for your attention!

Conclusion

- We defined a rigorous statistical framework in order to interpret LC and its generalizations in a unified way.
- Modern block ciphers are proven resistant against LC.
- This resistance extends to linear generalizations of LC,
- . . . but definitely not to non-linear ones!

Thank you for your attention!

Conclusion

- We defined a rigorous statistical framework in order to interpret LC and its generalizations in a unified way.
- Modern block ciphers are proven resistant against LC.
- This resistance extends to linear generalizations of LC,
- . . . but definitely not to non-linear ones!

Thank you for your attention!

Conclusion

- We defined a rigorous statistical framework in order to interpret LC and its generalizations in a unified way.
- Modern block ciphers are proven resistant against LC.
- This resistance extends to linear generalizations of LC,
- . . . but definitely not to non-linear ones!

Thank you for your attention!