Distribution Cryptanalysis

Kaisa Nyberg

Department of Information and Computer Science
Aalto University School of Science
kaisa.nyberg@aalto.fi

June 11, 2013

Introduction

Distribution Cryptanalysis

- Baignères, Junod, and Vaudenay, Asiacrypt 2004 developed distinguishing techniques based on χ^2.
- Maximov developed computational techniques for computing distributions over ciphers round by round, see e.g. the paper by Englund and Maximov at Indocrypt 2005
- Hermelin et al. 2008, developed a technique called Multidimensional Linear Cryptanalysis to compute estimates of distributions using strong linear approximations.
- Collard and Standaert 2009 introduced an heuristic cryptanalysis technique called Statistical Saturation Attack (SSA)
- Leander Eurocrypt 2011 showed that there is a mathematical link between SSA and Multidimensional LC

Using Multiple Linear Approximations

- My first lecture presented classical linear cryptanalysis based on a single linear approximation $u \cdot x + w \cdot E_k(x)$ and we learnt how to establish a good estimate of $c_x(u \cdot x + w \cdot E_k(x))^2$ by collecting as many trails from u to w as we can.
- Already Matsui in 1994 studied the possibility of using multiple linear approximations (more than one u and w) simultaneusly.
- Biryukov at al. developed statistical framework under the assumption that the linear approximations are statistically independent.
- Multidimensional linear cryptanalysis removes the assumption of independence [Hermelin et al. 2008]. The resulting statistical model leads to distribution cryptanalysis
- We start by introducing criterion of statistical independence of binary random variables.

Piling-up Lemma

Piling-Up Lemma

Definition. Let $\mathbf{T}$ be a binary-valued random variable with $p = \mathrm{P}[\mathbf{T} = 0]$. The quantity $c = 2p - 1$ is called the correlation of $\mathbf{T}$.

Theorem. Suppose we have k binary-valued random variables $\mathbf{T}_j$, and let c_j be the correlation of $\mathbf{T}_j$, $j = 1, 2, \ldots, k$. Then $\mathbf{T}_j$, $j = 1, 2, \ldots, k$, is a set of independent random variables if and only if for all subsets J of $\{1, 2, \ldots, k\}$, correlation of the binary random variable

$$\mathbf{T}_J = \bigoplus_{j \in J} \mathbf{T}_j$$

is equal to

$$\prod_{j \in J} c_j$$

The "only if" part of this theorem is known to cryptographers as Piling-up lemma.

Proof of Piling-Up Lemma

Proof. We will give the proof for $k = 2$ and denote $\mathbf{T}_1 + \mathbf{T}_2$ by $\mathbf{T}$. The general case follows by induction. By independency assumption

$$\begin{aligned} & \mathrm{P}[\mathbf{T} = 0] = \mathrm{P}[\mathbf{T}_1 = 0]\mathrm{P}[\mathbf{T}_2 = 0] + \mathrm{P}[\mathbf{T}_1 = 1]\mathrm{P}[\mathbf{T}_2 = 1] \\ & = \mathrm{P}[\mathbf{T}_1 = 0]\mathrm{P}[\mathbf{T}_2 = 0] + (1 - \mathrm{P}[\mathbf{T}_1 = 0])(1 - \mathrm{P}[\mathbf{T}_2 = 0]) \\ & = 2\mathrm{P}[\mathbf{T}_1 = 0]\mathrm{P}[\mathbf{T}_2 = 0] - \mathrm{P}[\mathbf{T}_1 = 0] - \mathrm{P}[\mathbf{T}_2 = 0] + 1 \end{aligned}$$

From this we get

$$\begin{aligned} & 2\mathrm{P}[\mathbf{T} = 0] - 1 \\ & = 4(\mathrm{P}[\mathbf{T}_1 = 0]\mathrm{P}[\mathbf{T}_2 = 0] - 2\mathrm{P}[\mathbf{T}_1 = 0] - 2\mathrm{P}[\mathbf{T}_2 = 0] + 1) \\ & = (2\mathrm{P}[\mathbf{T}_1 = 0] - 1)(2\mathrm{P}[\mathbf{T}_2 = 0] - 1) = c_1 c_2. \end{aligned}$$

Piling-Up Lemma and Independence

Example [Stinson] Let $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ be independent random variables with correlations $c_1 = c_2 = c_3 = 1/2$. Denote

$$\begin{aligned}
\mathbf{T}_{12} &= \mathbf{T}_1 + \mathbf{T}_2 \text{ with correlation } c_{12} = c_1 c_2 = \frac{1}{4},\\
\mathbf{T}_{23} &= \mathbf{T}_2 + \mathbf{T}_3 \text{ with correlation } c_{23} = c_2 c_3 = \frac{1}{4},\\
\mathbf{T}_{13} &= \mathbf{T}_1 + \mathbf{T}_3 \text{ with correlation } c_{13} = c_1 c_3 = \frac{1}{4}.
\end{aligned}$$

Then we can prove that $\mathbf{T}_{12}$ and $\mathbf{T}_{23}$ cannot be independent. If they would be independent, then by the Piling-up lemma the bias of $\mathbf{T}_{13} = \mathbf{T}_{12} + \mathbf{T}_{23}$ would be equal to $\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$ which is not the case.

To prove the converse of the Piling-up lemma, we introduce the Walsh-Hadamard transform, which allows us to establish a relationship between correlations and probability distributions of multidimensinal binary random variables.

Walsh-Hadamard Transform

Definition Suppose $f : \{0,1\}^n \to \mathbb{R}$ is any real-valued function of bit strings of length n. The Walsh-Hadamard transform transforms f to a function $F : \{0,1\}^n \to \mathbb{R}$ defined as

$$F(w) = \sum_{x \in \{0,1\}^n} f(x)(-1)^{w \cdot x},\ w \in \{0,1\}^n,$$

where the sum is taken over $\mathbb{R}$.

Similarly as the Walsh transform, the Walsh-Hadamard transform can also be inverted. It is its own inverse (involution) up to a constant multiplier:

$$f(x) \quad = \quad 2^{-n} \sum_{w \in \{0,1\}^n} F(w)(-1)^{w \cdot x} \text{, for all } x \in \{0,1\}^n.$$

Probability Distribution and Correlation of $(\mathbf{T}_1, \mathbf{T}_2)$

Suppose $\mathbf{Z} = (\mathbf{T}_1, \mathbf{T}_2)$ is a pair of binary random variables, $a = (a_1, a_2)$ be a pair of bits and c_a be the correlation of $a \cdot \mathbf{Z} = a_1 \mathbf{T}_1 + a_2 \mathbf{T}_2$.

Lemma

$$c_a = \sum_{(t_1, t_2)} \mathrm{P}[\mathbf{Z} = (t_1, t_2)](-1)^{a_1 t_1 + a_2 t_2}$$

Proof. Denote $t = (t_1, t_2)$ and $a \cdot t = a_1 t_1 + a_2 t_2$. Then

$$\begin{aligned} c_a &= 2\mathrm{P}[a \cdot \mathbf{Z} = 0] - 1 = \mathrm{P}[a \cdot \mathbf{Z} = 0] - \mathrm{P}[a \cdot \mathbf{Z} = 1] \\ &= \sum_{t, a \cdot t = 0} \mathrm{P}[\mathbf{Z} = t] - \sum_{t, a \cdot t = 1} \mathrm{P}[\mathbf{Z} = t] = \sum_{t} \mathrm{P}[\mathbf{Z} = t](-1)^{a \cdot t}. \end{aligned}$$

Probability Distribution and Correlation of $(\mathbf{T}_1, \mathbf{T}_2)$

- We saw that $c_a = F(a)$ is the Walsh-Hadamard transform of the real-valued function $f(t) = \mathrm{P}[\mathbf{Z} = t]$.
- Using the inverse Walsh-Hadamard transform we get the following

$$\mathrm{P}[\mathbf{Z} = t] = \frac{1}{4} \sum_{(a_1, a_2)} c_a (-1)^{a_1 t_1 + a_2 t_2} = \frac{1}{4} \sum_{a} c_a (-1)^{a \cdot t}.$$

Proof of the Converse of the Piling-Up Lemma, $k = 2$

Claim. If the correlation of $\mathbf{T}_1 + \mathbf{T}_2$ is equal to $c_1 c_2$ then $\mathbf{T}_1$ and $\mathbf{T}_2$ are independent.

Proof. For $a = (a_1, a_2) \in \{0, 1\}^2$, we use c_a to denote the correlation of $a \cdot \mathbf{Z} = a_1 \mathbf{T}_1 + a_2 \mathbf{T}_2$. Then

$$
\begin{aligned}
\mathbf{P}[\mathbf{T}_1 = t_1, \mathbf{T}_2 = t_2] &= \frac{1}{4} \sum_a c_a (-1)^{a_1 t_1 + a_2 t_2} \\
&= \frac{1}{4} (c_{(0,0)} + c_{(1,0)} (-1)^{t_1} + c_{(0,1)} (-1)^{t_2} + c_{(1,1)} (-1)^{t_1 + t_2}) \\
&= \frac{1}{4} (1 + c_1 (-1)^{t_1} + c_2 (-1)^{t_2} + c_1 c_2 (-1)^{t_1} (-1)^{t_2}) \\
&= \frac{1}{4} (c_1 (-1)^{t_1} + 1)(c_2 (-1)^{t_2} + 1) \\
&= \mathrm{P}[\mathbf{T}_1 = t_1] \mathrm{P}[\mathbf{T}_2 = t_2]
\end{aligned}
$$

Multidimensional Linear Cryptanalysis

Correlation and Distribution of Values of Functions

$f : \mathbb{F}_2^n \to \mathbb{F}_2^m$ vectorial Boolean function. For $\eta \in \mathbb{F}_2^m$ we denote

$$p_\eta = 2^{-n} \# \{ x \in \mathbb{F}_2^n \mid f(x) = \eta \},$$

and call the sequence p_η, $\eta \in \mathbb{F}_2^m$, the distribution of f.

Theorem The correlations of masked vectorial Boolean function can be computed as Walsh-Hadamard transform of the distribution of the function:

$$c_x(a \cdot f(x)) = 2^{-n} \sum_{x \in \mathbb{F}_2^n} (-1)^{a \cdot f(x)} = \sum_{\eta \in \mathbb{F}_2^m} p_\eta (-1)^{a \cdot \eta}$$

And conversely,

$$p_\eta = 2^{-m} \sum_{a \in \mathbb{F}_2^m} (-1)^{a \cdot \eta} c_x(a \cdot f(x))$$

for all $\eta \in \mathbb{F}_2^m$.

Multidimensional Linear Cryptanalysis

Definition Let U and W be linear subspaces in $\mathbb{F}_2^n$. Then the set of linear approximations

$$u \cdot x + w \cdot E_k(x),\ u \in U,\ w \in W,$$

is called multidimensional linear approximation of E_k.

In practice, the input space is split into two parts $\mathbb{F}_2^n = \mathbb{F}_2^s \times \mathbb{F}_2^t$ and the output space is split into two parts $\mathbb{F}_2^n = \mathbb{F}_2^q \times \mathbb{F}_2^r$, and WLOG we assume that

$$U = \mathbb{F}_2^s \times \{0\} \text{ and } W = \mathbb{F}_2^q \times \{0\}.$$

Assume that we have the correlations of the linear approximations

$$c(u, w) = c_x(u \cdot x + w \cdot E_k(x)),\ u \in U,\ w \in W.$$

Then we can compute the distribution of values (x_s, y_q), where

$$x = (x_s, x_t) \in \mathbb{F}_2^s \times \mathbb{F}_2^t,\ \text{ and } E_k(x) = y = (y_q, y_r) \in \mathbb{F}_2^q \times \mathbb{F}_2^r.$$

Computing the Distribution

Theorem Using the notation introduced above

$$p_{(\xi_s,\eta_q)} = 2^{-(s+q)} \sum_{u\in U, w\in W} (-1)^{u\cdot\xi + w\cdot\eta} c(u,w),$$

for all $(\xi_s, \eta_q) \in \mathbb{F}_2^s \times \mathbb{F}_2^q$.

Proof.

$$\begin{aligned}
p_{(\xi_s,\eta_q)} &= \sum_{\xi_t,\eta_r} p(\xi,\eta) \\
&= \sum_{\xi_t,\eta_r} 2^{-2n} \sum_{a,b} (-1)^{a\cdot\xi + b\cdot\eta} c(a,b) \\
&= \sum_{\xi_t,\eta_r} 2^{-2n} \sum_{a,b} (-1)^{a_s\cdot\xi_s + a_t\cdot\xi_t + b_q\cdot\eta_q + b_r\cdot\eta_r} c(a,b) \\
&= 2^{-(s+q)} \sum_{a_s,b_q} (-1)^{a_s\cdot\xi_s + b_q\cdot\eta_q} c((a_s,0),(b_q,0)),
\end{aligned}$$

from where we see the result.

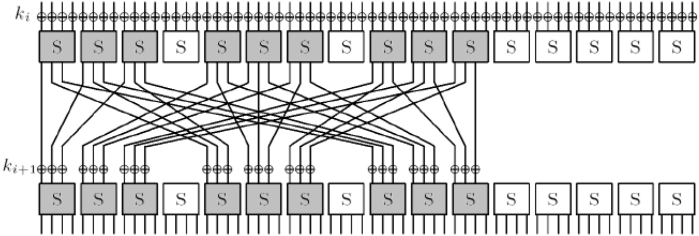
Multidimensional Linear Cryptanalysis in Practice

- Find U and W such that there exists several linear approximations $u \cdot x + w \cdot E_k(x)$, $u \in U$, $w \in W$, with large correlations $c(u, w)$. Linear approximations with significant smaller correlations cn be omitted.
- Compute probabilities $p(\xi_s, \eta_q)$ from the correlations as shown above.
- The strength of the multidimensional linear approximations depends on the nonuniformity of the distribution $p_{(\xi_s, \eta_q)}$, $(\xi_s, \eta_q) \in \mathbb{F}_2^s \times \mathbb{F}_2^q$
- Nonuniformity of $p_{(\xi_s, \eta_q)}$ is measured in terms of capacity:

$$
\begin{aligned}
C &= \sum_{\xi_s, \eta_q} \left(p_{(\xi_s, \eta_q)} - 2^{-(s+q)} \right)^2 \\
&= \sum_{(u,w) \in U \times W \setminus \{(0,0)\}} c(u, w)^2
\end{aligned}
$$

Mathematical Link between SSA and Multidimensional LC

SSA Trail



Multidimensional Linear Trail

The same multitrail was used be Joo Cho in his multidimensional linear attack on PRESENT in CT-RSA 2010. This is not an accidental coincidence.

To see this, let us recall the following correlations

$$f : \mathbb{F}_2^s \times \mathbb{F}_2^t \to \mathbb{F}_2^n$$

$$c_x(u \cdot x + v \cdot z + w \cdot f(x, z)) = 2^{-n}(-1)^{v \cdot z} \sum_{x \in \mathbb{F}_2^s} (-1)^{u \cdot x + w \cdot f(x,z)},$$

for any (fixed) $z \in \mathbb{F}_2^t,$ and

$$c_{x,z}(u \cdot x + v \cdot z + w \cdot f(x, z)) = 2^{-n} \sum_{x \in \mathbb{F}_2^s,\, z \in \mathbb{F}_2^t} (-1)^{u \cdot x + v \cdot z + w \cdot f(x,z)}$$

The Fundamental Theorem

Theorem

$$2^{-t} \sum_{z \in \mathbb{F}_2^t} c_x (u \cdot x + w \cdot f(x,z))^2 \quad = \quad \sum_{v \in \mathbb{F}_2^t} c_{x,z} (u \cdot x + v \cdot z + w \cdot f(x,z))^2$$

This result, in different contexts and notation, has previously appeared (at least) in:

A. Canteaut, C. Carlet, P. Charpin, C. Fontaine. On cryptographic properties of the cosets of r(1, m). IEEE Trans. IT 47(4), 14941513 (2001)

N. Linial, Y. Mansour and N. Nisan. Constant depth circuits, Fourier transform, and learnability. Journal of the ACM 40 (3), 607-620 (1993).

K. Nyberg: Linear Approximations of Block Ciphers (1994) (see also the Linear Hull theorem in my first lecture)

Statistical Saturation Link

[Leander 2011]

$E_k : \mathbb{F}_2^s \times \mathbb{F}_2^t \rightarrow \mathbb{F}_2^q \times \mathbb{F}_2^r$

Straightforward application of the Fundamental Theorem gives

$$2^{-s} \sum_{x_s \in \mathbb{F}_2^s} \sum_{w \in W \setminus \{0\}} c_{x_t}(w \cdot E_k(x_s, x_t))^2 = \sum_{u \in U} \sum_{w \in W \setminus \{0\}} c_x(u \cdot x + w \cdot E_k(x))^2$$

The expression on the right hand side is the capacity of the multidimensional linear approximation

$$u \cdot x + w \cdot E_k(x), u \in U = \mathbb{F}_2^s \times \{0\},\ w \in W = \mathbb{F}_2^q \times \{0\}.$$

The expression on the left hand side is the averge capacity of the distribution of the values

$$y_q, \text{ where } y = (y_q, y_r) = E_k(x)$$

taken over all fixations $x_s \in \mathbb{F}_2^s$.

Attacks are Different in Practice

The mathematical link offers different ways for performing the attacks.

Running the known plaintext multidimensional linear attack takes 2^{s+q} memory.

Sampling for evaluation of the expression on the left can be done with 2^q memory using chosen plaintext.

Question: How much the behaviour for a fixed x_s differs from the average behaviour?

Distinguishing Distributions

The Best Distinguisher

- Given two probability distributions $p = (p_z)$ and $p' = (p'_z)$ the question is to decide whether a given sample distribution $q(N) = (q_z(N))$ obtained from a sample of size N, is drawn from p or p'.
- The optimal distinguisher is based on the LLR

$$\mathrm{LLR}(q(N)) = \sum_{z \in \mathrm{Supp}(q)} q_z(N) \log \frac{p_z}{p'_z}$$

- The distinguisher decides for p if $\mathrm{LLR}(q)$ is above a threshold, otherwise it decides for p'.
- The threshold determines the error probability as a function of the size N of the sample.
- The error probability depends of the Chernoff information $\mathrm{C}(p, p')$ between p and p'

Close to Uniform Distribution

- Let p' be the uniform distribution and p a close-to-uniform probability distribution over a set of cardinality M
- For close-to-uniform distributions, the Chernoff information between p and p' can be approximated using the squared Euclidean distance between the distributions or the sum of squared correlations over nontrivial linear approximations as:

$$\frac{M}{8 \ln 2} \sum_{z} (p_z - p'_z)^2 = \frac{1}{8 \ln 2} \sum_{w \neq 0} |c_z(w \cdot z)|^2$$

- We call the quantity

$$M \sum_{z} (p_z - p'_z)^2 = \sum_{w \neq 0} |c_z(w \cdot z)|^2$$

the capacity of p and denote it by $C(p)$.

Data Requirement for Optimal Distinguisher

- Baignères and Vaudenay (ICITS 2008) showed that, for close to uniform distributions, the data requirement for the LLR distinguisher can be given as:

$$N_{\mathrm{LLR}} \approx \frac{\lambda}{C(p)},$$

  where the constant λ depends only on the success probability.
- In practice, accurate estimates of the alternative p required for LLR computation is hard to obtain. But an estimate of its capacity may be available.
- Junod 2003: χ^2 test is asymptotically optimal distinguisher for distributions of binary variables.

Outline

- Problem: Determine data complexity of the χ^2 distinguisher that is reasonably accurate also for probability distributions with large support of size *M*.
- Solution: use χ^2 cryptanalysis by Vaudenay (ACM CCS 1995). It is based on using
- We derive a bound for data complexity and demonstrate its accuracy by using distributions of support size 10^8.
- We can also predict the data complexity of the SSA.

Distinguishing Test

- Distinguishing probability distributions over a large set of values of size M
  - Uniform distribution
  - Non-uniform distribution p with known capacity

$$C(p) = M \sum_{\eta=1}^{M} (p(\eta) - \frac{1}{M})^2.$$

- Problem. Determine the data complexity estimates of the χ^2 distinguisher.
- Solution. Use statistic

$$T = NM \sum_{\eta=1}^{M} (q(\eta) - \frac{1}{M})^2,$$

where q is the distribution obtained from the data.

- Need to determine the probability distribution of T in both cases.

Uniform Binomial Distribution

N number of data (sample size)

M cells with equal probabilities $\frac{1}{M}$

$X(\eta)$ number of data in cell η

$X(\eta) \sim \mathcal{B}(\frac{1}{M})$

For large N:

$$X(\eta) \sim \mathcal{N}(\frac{N}{M}, \frac{N}{M})$$

Nonuniform Binomial Distribution

N number of data (sample size)

M cells with different probabilities $p(\eta)$, $\eta = 1, 2, \ldots, M$

$Y(\eta)$ number of data in cell η

$Y(\eta) \sim \mathcal{B}(p(\eta))$

For N large:

$$Y(\eta) \quad \sim \quad \mathcal{N}(Np(\eta), Np(\eta)) \approx \mathcal{N}(Np(\eta), N/M)$$

Central and Noncentral χ^2 Distributions

Let $X_i = \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$. Then

$$T_0 = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2}$$

has central χ^2_{n-1}-distribution with $n-1$ degrees of freedom, and

$$T_1 = \sum_{i=1}^{n} \frac{(X_i)^2}{\sigma_i^2}$$

has noncentral $\chi^2_{n-1}(\delta)$-distribution with $n-1$ degrees of freedom and noncentrality parameter

$$\delta = \sum_{i=1}^{n} \frac{\mu_i^2}{\sigma_i^2}.$$

The expected values and variances are

$$\mu_{T_0} = n-1, \qquad \sigma^2_{T_0} = 2(n-1)$$
$$\mu_{T_1} = n-1+\delta, \; \sigma^2_{T_1} = 2(n-1+2\delta).$$

Probability Distributions of T

- If q is drawn from uniform distribution, then

$$T = T_0 = \sum_{\eta=1}^{M} \frac{(Nq(\eta) - N/M)^2}{N/M} \sim \chi^2_{M-1}.$$

- If q is drawn from nonuniform distribution p, then

$$T = T_1 = \sum_{\eta=1}^{M} \frac{(Nq(\eta) - N/M)^2}{N/M} \sim \chi^2_{M-1}(\delta),$$

where

$$\delta = \sum_{\eta=1}^{M} \frac{(Np(\eta) - N/M)^2}{N/M} = NC(p).$$

- Denote $C(p) = C$.

Normal Approximations of Distributions of T

- If q is drawn from uniform distribution, then

$$T = T_0 \sim \mathcal{N}(M, 2M).$$

- If q is drawn from nonuniform distribution with capacity C, then

$$T = T_1 \sim \mathcal{N}(M + NC, 2(M + 2NC)).$$

- We will see later that the relevant area of N will be around $\sqrt{M}/C$. Assuming

$$N < \frac{M}{2C},$$

we obtain that the variance of T_1 is upperbounded by $4M$.

The χ^2 Test

H$_0$: q is drawn from the uniform distribution
H$_1$: q is drawn from unknown nonuniform distribution with capacity C

H$_1$ is accepted if and only if

$$T \geq M + \tau, \text{ where } 0 < \tau < NC.$$

Threshold τ is set such that the probabilities

$$\begin{aligned} \alpha &= \Pr[T_0 \geq M + \tau] \\ \beta &= \Pr[T_1 < M + \tau] \end{aligned}$$

of Type 1 and 2 errors are equal. Then we obtain

$$\tau = \frac{NC}{1+\sqrt{2}}, \text{ and } \alpha = \beta = \Phi\left(-\frac{NC}{(2+\sqrt{2})\sqrt{M}}\right)$$

Data Complexity

For success probability $P_S = 1 - \frac{\alpha+\beta}{2} = 1 - \alpha$ we get

$$\frac{NC}{(2+\sqrt{2})\sqrt{M}} \geq -\Phi^{-1}(1 - P_S) = \Phi^{-1}(P_S),$$

that is,

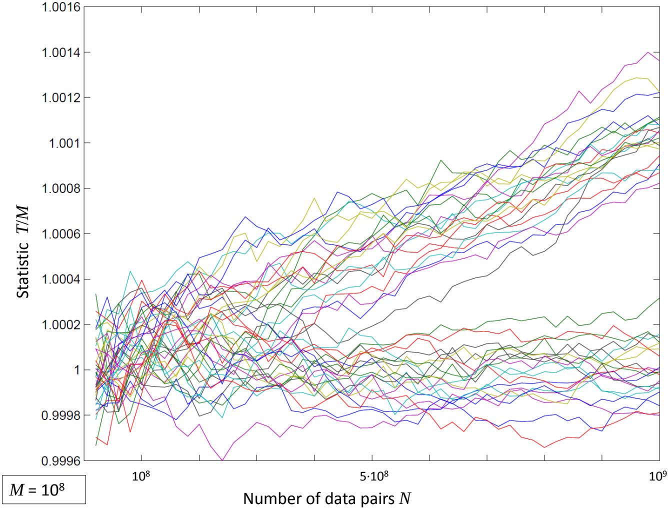
$$N \geq (2+\sqrt{2})\Phi^{-1}(P_S)\frac{\sqrt{M}}{C}.$$

For typical P_S, the multiplier of $\sqrt{M}/C$ is around 8. Then we must check that

$$8\frac{\sqrt{M}}{C} < \frac{M}{2C}$$

which holds for $M \geq 2^8$.

Experiment on a Large Distribution



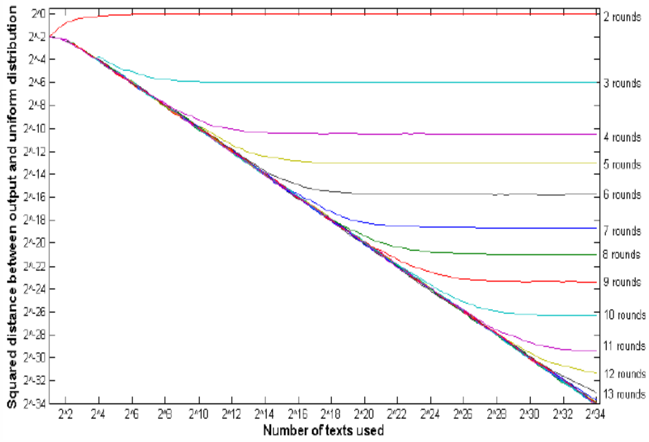
Experiment on a Large Distribution

- $M = 10^8$
- $C = 10^{-4}$
- x-axis = N
- y-axis:

$$y = \frac{T}{M} \approx \begin{cases} 1, & \text{for random,} \\ 1 + \frac{C}{M}N, & \text{for cipher.} \end{cases}$$

- So the slope of the upper bunch of lines should be equal to $C/M = 10^{-12}$. From the data the slope is $\approx 10^{-12}$.
- Distinguisher seems to work for $N \geq 5 \cdot 10^8 = 5\sqrt{M}/C$

SSA



[Collard and Standaert CT-RSA 2009]

SSA

- y-axis:

$$y = \log_2 \frac{T}{MN} = \log_2 T - \log_2 M - \log_2 N$$

- x-axis: $x = \log_2 N$; thus

$$y = \log_2 T - log_2 M - x$$

- For random curves $T \approx M$, and we get the line $y = -x$.
- For cipher curves $T \approx M + CN$ and

$$y = \log_2(\frac{1}{N} + \frac{C}{M}) \longrightarrow \log_2 \frac{C}{M} \quad \text{as } N \longrightarrow \infty.$$

- Given that $M = 2^8$ we can read average capacities of the distributions for small number of rounds from the picture.

Key Ranking in Distribution-based Algorithm 2

Definition[Selçuk, JoC 2009] A key recovery attack for an n-bit key achieves an advantage of a bits over exhaustive search, if the correct key is ranked among the top $r = 2^{n-a}$ out of all 2^n key candidates.

Assumption (Wrong-key Hypothesis) There are two different probability distributions p and p' such that for the right key κ_0, the data is drawn from p and for a wrong key $\kappa \neq \kappa_0$ the data is drawn from p'.

For simplicity, we restrict to the case where p' is the uniform distribution over M values.

p is a non-uniform distribution. Statistical analysis exploits known non-uniformity of p.

Ranking Statistics T

- Rearrange the keys κ according to their values $T(\kappa)$ in decreasing order of magnitude.
- Index the ordered T values as

$$T_0 \geq T_1 \geq \cdots \geq T_{2^n-1}$$

where T_i is called the ith order statistic.

- For fixed advantage a the right key κ_0 should be among the $r = 2^{n-a}$ highest ranking keys.

Theorem[Selçuk, JoC 2009] The statistic T_r for the wrong key in the r^{th} position is distributed as

$$T_r \sim \mathcal{N}(\mu_a, \sigma_a^2), \text{ where}$$

$$\mu_a = F_W^{-1}(1 - 2^{-a}) \text{ and } \sigma_a \approx \frac{2^{-(n+a)/2}}{f_W(\mu_a)}.$$

Here f_W and F_W are the density function and the cumulative density function of the statistic $T(\kappa)$ for a wrong key κ.

Success Probability

Assume that

$$T(\kappa_0) \sim \mathcal{N}(\mu_R, \sigma_R^2).$$

Then

$$P_S = \Pr(T(\kappa_0) - T_r > 0) = \Phi\left(\frac{\mu_R - \mu_a}{\sqrt{\sigma_R^2 + \sigma_a^2}}\right),$$

since $T(\kappa_0) - T_r \sim \mathcal{N}(\mu_R - \mu_a, \sigma_R^2 + \sigma_a^2)$.

χ^2 Statistics

- Assume that a good estimate of the capacity $C(p)$ of p is available.
- Compute statistic

$$T = NM \sum_{\eta=1}^{M} (q(\eta) - \frac{1}{M})^2,$$

where q is the distribution obtained from the data.

- For the correct key

$$T = T_1 \sim \chi^2_{M-1}(NC(p)) \approx \mathcal{N}(M + NC, 2(M + 2NC)).$$

- For the wrong key

$$T = T_0 \sim \chi^2_{M-1} \approx \mathcal{N}(M, 2M).$$

Estimates

$$
\begin{aligned}
\mu_R &= M + NC(p) \\
\sigma_R^2 &= 2(M + 2NC(p))
\end{aligned}
$$

$$
\begin{aligned}
\mu_a &= b\sqrt{2M} + M \\
\sigma_a^2 &= \frac{2M}{2^{n+a}\phi(b)^2},
\end{aligned}
$$

where $b = \Phi^{-1}(1 - 2^{-a})$.

- Estimate $\sigma_a^2 < M$.
- Restrict to the case $NC(p) < M/4$. This is not essential restriction, since finally $NC(p)$ will be close to a small constant multiple of $\sqrt{M}$.
- Obtain $\sqrt{\sigma_a^2 + \sigma_R^2} < 2\sqrt{M}$.

Data Complexity

By solving data complexity from the formula for success probability, we obtain an upperbound

$$N_{\chi^2} = \frac{\left(b + \sqrt{2}\Phi^{-1}(P_S)\right)\sqrt{2M}}{C(p)}$$

Compare with the FSE 2009 formula:

$$N_{\chi^2} = \frac{2\sqrt{M}b + 4(\Phi^{-1}(2P_S - 1))^2}{C(p)},$$

where it is assumed that b, that is, advantage a is large, and that P_S is large.

Summary

- For close-to-uniform distribution p (with support of any size), an upperbound to the data requirement of the LLR distinguisher can be given as:

$$N_{\mathrm{LLR}} = \frac{\lambda}{C(p)},$$

  where the constant λ depends only on the success probability.
- For close-to-uniform distribution p with support of cardinality M, the data requirement of the χ^2 distinguisher can be given as:

$$N_{\chi^2} = \frac{\lambda' \sqrt{M}}{C(p)},$$

  where

$$\begin{aligned} \lambda' &= \sqrt{2}\Phi^{-1}(1 - 2^{-a}) + 2\Phi^{-1}(P_S) \quad \text{(key recovery)} \\ \lambda' &= (\sqrt{2} + 2)\Phi^{-1}(P_S) \quad \text{(distinguisher)} \end{aligned}$$