Computing the Success Probability in Linear Cryptanalysis

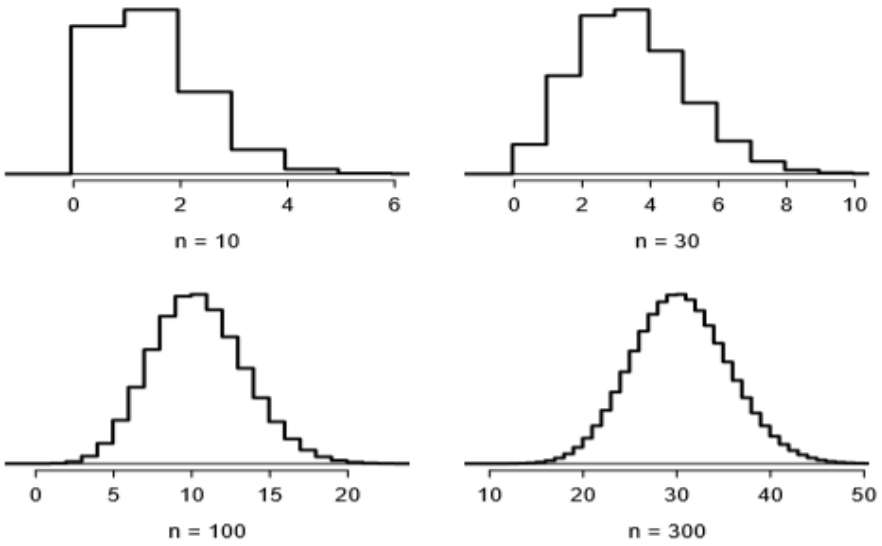
Prof. Dr. Ali Aydın Selçuk
TOBB-ETÜ

Outline

- Binomial counters & normal approximation
- Key ranking approach
- Order statistics & the “advantage”
- Hypothesis testing approach
- Neyman-Pearson and optimal tests
- Correlation matrices & linear hulls
- Multidimensional linear cryptanalysis

Binomial Counters & Normal Approximation

- Let $X = \sum_{i=1}^{N} x_i$, where each x_i is an i.i.d. binary variable with $\Pr(1) = p$, $\Pr(1) = 1 - p$.
  X follows a binomial distribution $\mathrm{Bin}(N, p)$.
  We have $\mathrm{E}(X) = pN$, $\mathrm{Var}(N) = p(1 - p)N$.
- Binomial distribution can be approximated by the normal distribution. Good if $p(1 - p)N \geq 4$. Best if $p \approx ½$ and N large.
- E.g., $p = 0.10$:

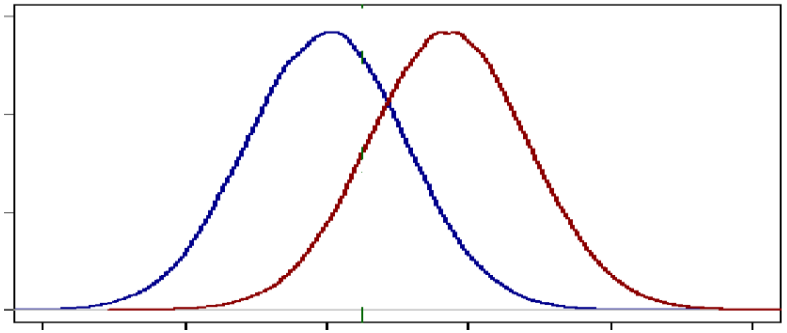


Linear Cryptanalysis, Alg.2

- Find a binary linear relation (LA) between the input, output, and key bits of the first $r - 1$ rounds with $p \neq ½$. ("bias" $\varepsilon = |p - ½|$)
- Collect a large number (N) of plaintext-ciphertext.
- Try all possible values for the last round subkey with the plaintext sample, decrypt the final round, and evaluate the linear equation each time.
- T_k = # of plaintexts satisfying the eq. with key k
- Take the key value that maximizes the "sample bias" $|T_k/N - ½|$.

Linear Cryptanalysis, Alg.2

- The right key will give a bias around $(p - 1/2)$, whereas wrong keys will give a bias around 0.



- The two distributions will be better separated as N increases and the variance decreases.
- About $N = c\varepsilon^{-2}$ plaintexts are needed, $2 \leq c \leq 16$, and ε is typically between 2^{-20} and 2^{-64}.

Key Ranking Approach

- Attack algorithm:
  - Rank key candidates according to their sample bias
  - Take the key that is ranked the highest
- In practice, getting the right key at the top may require too many plaintexts. It may be more practical to get it "among the highest".
- For instance, $16\varepsilon^{-2}$ vs. $2\varepsilon^{-2}$ plaintexts. (E.g., Matsui's DES attack, 1994.)

Order Statistics

- Let $x_1, x_2, \ldots, x_n$ be i.i.d. random variables. (E.g., we roll a dice 10 times.) What is the probability distribution of the maximum? Or, the third highest value?
- For relatively large n (e.g., $n > 100$), the distribution of the order statistics can be closely modeled by a normal distribution.

Success Probability of LC

- "Success": Getting the right key ranked among the top ℓ candidates.
- X_0: Sample bias of the right key
  $Y_{\ell-1}$: Sample bias of the $(\ell-1)$st highest wrong key
- Success probability: $P_S = \Pr(X_0 \geq Y_{\ell-1})$
- m: bit length of the attacked subkey
  $a = m - \lg \ell$ (the "advantage" of the attack)
- Plaintext complexity depends on a and P_S :

Success Probability of LC

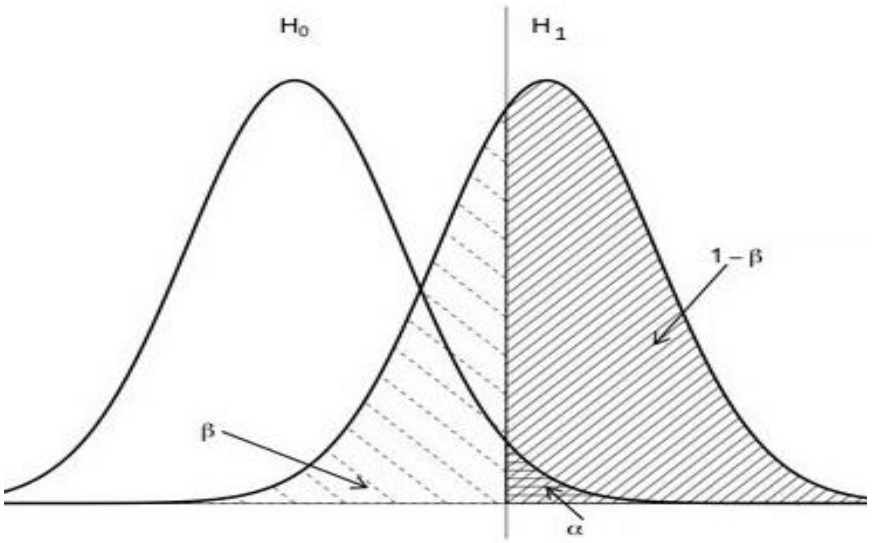
Using the normal approx. for binomial counters and for the order statistics: [Selçuk, 2007]

$$P_S = \Phi\left(2\sqrt{N}|p - 1/2| - \Phi^{-1}(1 - 2^{-a-1})\right)$$

$$N = \left(\frac{\Phi^{-1}(P_S) + \Phi^{-1}(1 - 2^{-a-1})}{2}\right)^2 \cdot |p - 1/2|^{-2}$$

Hypothesis Testing Approach

- We collect a sample from an unknown distribution D and try to decide whether H_0: D = D_0 against the alternative hypothesis H_1: D = D_1.
- Two types of error:
  - Non-detection:
    Pr(reject H_0 | H_0) = α
  - False alarm:
    Pr(accept H_0 | H_1) = β



Hypo.Test. for Key Recovery Attacks

- Test the subkey candidates with the plaintext sample. If a subkey generates a significant correlation, take it as a possible right key. [Siegenthaler, 1985]
- The exact same idea applies to LC, Alg.2.
- Somewhat equivalent to key ranking approach: False alarm prob. β here gives the length ℓ of the subkey list to be checked in key ranking.

Likelihood Ratio Tests (Optimal)

- Likelihood function measures the "likelihood" of a hypothesis H according to the experiment outcome **z**: $L(\mathrm{H} \mid \mathbf{z}) = \Pr(\mathbf{z} \mid \mathrm{H})$.
- "Likelihood ratio": $LR(\mathbf{z}) = L(\mathrm{H}_0 \mid \mathbf{z}) / L(\mathrm{H}_1 \mid \mathbf{z})$
  "Log likelihood ratio" $LLR(\mathbf{z}) = \log LR(\mathbf{z})$
- Neyman-Pearson Theorem: The optimal rejection region that minimizes β for a given α is $\{\mathbf{z}: \mathrm{LR}(\mathbf{z}) \leq \tau\}$. (The τ value is determined from α and H_0.)
- The test that minimizes $\alpha+\beta$ is given by $\{\mathbf{z}: LR(\mathbf{z}) \leq 1\}$ (or, equivalently $\{\mathbf{z}: LLR(\mathbf{z}) \leq 0\}$).

Data Complexity of an Optimal Distinguisher

- Kullback-Liebler distance from distribution D_1 to D_0:

$$D(\mathrm{D}_0||\mathrm{D}_1) = \sum_z \Pr_{\mathrm{D}0}(z) \log \frac{\Pr_{\mathrm{D}0}(z)}{\Pr_{\mathrm{D}1}(z)}$$

(It is the expectation of *LLR* under D_0.)

- Data complexity of the *best distinguisher* (not just LC) that achieves an "error probability" $P_e = (\alpha+\beta)/2$ is

$$N = \frac{2\,\Phi^{-1}(P_e)}{D(\mathrm{D}_0||\mathrm{D}_1)} \qquad \text{(for LC: } D(\mathrm{D}_0||\mathrm{D}_1) \approx 4\varepsilon^2\text{)}$$

assuming the normal approx. for binomial counters. [Baignères, Junod, Vaudenay, 2004]

A Very Accurate Estimator

- Normal approximation for binomial counters works very well for LC, but less so for DC (Np is small).
- [Blondeau, Gérard, Tillich, 2011] used another approximation based on the Kullback-Leibler distance and obtained, for fixed $\alpha = 0.5$,

$$N = -\frac{\ln(2\sqrt{\pi}\beta)}{D(p_0||p_1)}$$

  which gives accurate results for a wide range of attacks.
- It can be computed directly and easily! (no Φ^{-1} or such)

Linear Hull Effect and Correlation Matrices

- Somewhat like differentials in DC, bias of an LA can be calculated better over multiple trails. (“Linear hull effect”)
- However, the biases should be added with a +/– sign, determined by the round keys. (“Correlation matrices”)
- Sum of the absolute biases over trails can only be used as an upper bound!
- The linear hull effect is better used to argue about the key dependence of the bias.

Key Dependence of Bias

- Traditionally, the bias has been assumed to be independent of the encryption key, which is not valid for certain ciphers.
- "Median bias" is helpful then: [Leander, 2011]
  - If the bias is calculated over one dominant trail, at least half of the keys will give a bias at least that high.
  - If the bias is calculated over many trails, at least one quarter of the keys will give a bias that is comparable or higher.

Multiple/Multidimensional LC

- When no single linear approximation is sufficiently significant (e.g., on PRESENT), several LAs with different input/output masks can be combined to attack.
- In this case, deviation from uniform randomness is checked by a "goodness of fit" test (χ^2 or LLR). The key that deviates most is taken.
- Both Alg.1 and Alg.2 have multidimensional extensions.
- Data complexity is proportional to $\left(\sum_i \varepsilon_i^2\right)^{-1}$.

Conclusions

- 20+ years after Matsui's original DES attack, we have a very good understanding of LC:
  - A very good understanding of the bias
  - Excellent formulas for success probability calculation
- However, in fact, we are not that far from Matsui's heuristic calculations:
  - Bias can mostly be approximated by the P.U. Lemma.
  - Data complexity is $c\varepsilon^{-2}$ for some small constant c.

Conclusions

- “Az gittik uz gittik dere tepe düz gitttik… Bir de dönüp baktı ki bir arpa boyu yol gitmişiz.” ☺

- Nevertheless, most of that work was applicable to DC and other attacks, whose success probabilities had previously been studied in a very ad hoc manner. These analysis made significant contributions on them.