# The Analysis of Simultaneous Differences in Differential Cryptanalysis

S. Murphy

Technical Report
RHUL–MA–2012–13
13 May 2011


Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX, England

http://www.rhul.ac.uk/mathematics/techreports

**Abstract.** This paper considers an extension of standard differential cryptanalysis in which a number of output differences arising from a single input difference are considered, and gives a statistical treatment of this situation.

# 1 Introduction

The basic idea of standard differential cryptanalysis [1, 2] is to use the observation that a particular input difference $\Delta$ of a pair of plaintexts gives a particular output difference $\Delta_1$ several encryption rounds later with some relatively large probability $p_1$ to obtain key information. The key information is essentially obtained by carrying out trial partial decryptions of the last few rounds of the pair of corresponding ciphertexts under various values of the last round subkeys and choosing those that give the best empirical agreement with the probability $p_1$ for the difference $\Delta_1$.

We discuss a simple extension of standard differential cryptanalysis [1, 2] in which a particular input difference $\Delta$ can give rise to $l$ possible round differences $\Delta_1, \ldots, \Delta_l$ after several rounds with reasonable probability, so we have

$$\begin{array}{cccc} \text{Differences} & \Delta \longrightarrow \Delta_1 & \text{with probability} & p_1, \\ \vdots & \vdots & \vdots & \vdots \\ \text{Differences} & \Delta \longrightarrow \Delta_l & \text{with probability} & p_l. \end{array}$$

Thus input difference $\Delta$ is not mapped to any of the specified output differences $\Delta_1, \ldots, \Delta_l$ with probability $p_0 = 1 - (p_1 + \ldots + p_l)$. We discuss the optimal method for finding key information in this situation

# 2 A Statistical Analysis

The basic observation in differential cryptanalysis is a pair of plaintexts with a given difference $\Delta$ and a corresponding pair of ciphertexts, and we use $z$ to denote such a generic observation. Suppose now that we have $N$ such observations $z_1, \ldots, z_N$, that is $N$ plaintext pairs with difference $\Delta$ and their corresponding ciphertexts, and we let $\mathbf{z} = (z_1, \ldots, z_N)$ be the vector denoting the entire data.

In order to perform this cryptanalysis, we have to carry out trial decryptions over a few rounds using different possible key values $k$ to see how often each of the possible round differences $\Delta_1, \ldots, \Delta_l$ occurs with each possible such key value $k$. Accordingly, we let

$$\begin{array}{ll} X_i(k, \mathbf{z}) & \text{be the number of times round difference } \Delta_i \text{ occurs} \\ & \text{with data } \mathbf{z} \text{ and trial decyrpion with key value } k, \end{array}$$

that is a count for round difference $\Delta_i$ $(i = 1, \ldots, l)$ with plaintext-ciphertext data $\mathbf{z}$ under key value $k$. For completeness, we define

$$X_0(k, \mathbf{z}) = N - (X_1(k, \mathbf{z}) + \ldots + X_i(k, \mathbf{z}))\,,$$

so $X_0(k, \mathbf{z})$ is count of the number of times that none of the specified round differences $\Delta_1, \ldots, \Delta_l$ occur. We let

$$\begin{array}{rcl} \mathbf{X}(k, \mathbf{z}) & = & (X_0(k, \mathbf{z}), X_1(k, \mathbf{z}), \ldots, X_l(k, \mathbf{z}))^T \\ \text{and} \quad \mathbf{X}'(k, \mathbf{z}) & = & (X_1(k, \mathbf{z}), \ldots, X_l(k, \mathbf{z}))^T \end{array}$$

be the vector of all counts and the deleted vector of counts for the $l$ specified differences respectively. Similarly, we respectively let $\mathbf{p} = (p_0, p_1, \ldots, p_l)^T$ and $\mathbf{p}' = (p_1, \ldots, p_l)^T$ be the vector of all probabilities and the deleted vector of probabilities for the $l$ specified differences.

The count vector $\mathbf{X}(k, \mathbf{z}) \sim \text{Mult}(N, \mathbf{p})$, that is $\mathbf{X}$ has a multinomial distribution with mean vector $\mathbf{E}(\mathbf{X}(k, \mathbf{z})) = N\mathbf{p}$. The variance and covariance of the components of $\mathbf{X}$ are given by

$$\begin{array}{rcll} \text{Var}(X_i(k, \mathbf{z})) & = & Np_i(1 - p_i) & \\ \text{and} \quad \text{Cov}(X_i(k, \mathbf{z}), X_j(k, \mathbf{z})) & = & -Np_ip_j & [i \neq j]. \end{array}$$

Thus if we define the $(l + 1) \times (l + 1)$ matrix $\Sigma_{\mathbf{p}}$ by

$$\Sigma_{\mathbf{p}} = D(\mathbf{p}) - \mathbf{p}^T\mathbf{p},$$

where $D(\mathbf{p})$ denotes the diagonal matrix with diagonal elements $p_0, p_1, \ldots, p_l$, then $\mathbf{X}(k, \mathbf{z})$ has covariance matrix

$$N\Sigma_{\mathbf{p}} = N\left(D(\mathbf{p}) - \mathbf{p}^T\mathbf{p}\right).$$

# 3 A Small Sample Version

The likelihood function of the parameter key value $k$ given the data $\mathbf{z}$ is

$$L(k, \mathbf{z}) = \prod_{j=1}^{N}\left(\prod_{i=0}^{l} p_i^{x_i(k, z_j)}\right) = \prod_{i=0}^{l} p_i^{\sum_{j=1}^{N} x_i(k, z_j)} = \prod_{i=0}^{l} p_i^{x_i(k, \mathbf{z})}.$$

This means that the log-likelihood function $\mathcal{L}(k, \mathbf{z})$ is given by

$$\mathcal{L}(k, \mathbf{z}) = \sum_{i=0}^{l} x_i(k, \mathbf{z}) \log p_i.$$

If we let $q_i = \log p_i$ $(i = 0, 1, \ldots, l)$ so we can define the vector $\mathbf{q} = (q_0, q_1, \ldots, q_l)$, then the maximum likelihood estimate of the parameter key value $k$ given the data $\mathbf{z}$ is that value of $k$ which maximises

$$\Phi(k, \mathbf{z}) = \mathbf{q}^T X_i(k, \mathbf{z}) = \sum_{i=0}^{l} X_i(k, \mathbf{z}) \log p_i.$$

In many cases under consideration, one value of the probability, say $p_0$, dominates the other values $p_1, \ldots, p_l$, so $p_1 + \ldots + p_l$ is very small. In such situations, we have

$$\begin{aligned}
\mathcal{L}(k, \mathbf{z}) &= x_0(k, \mathbf{z}) \log p_0 + \textstyle\sum_{i=1}^{l} x_i(k, \mathbf{z}) \log p_i \\
&= x_0(k, \mathbf{z}) \log \left(1 - \textstyle\sum_{i=1}^{l} p_i\right) + \textstyle\sum_{i=1}^{l} x_i(k, \mathbf{z}) \log p_i \\
&\approx -x_0(k, \mathbf{z}) \textstyle\sum_{i=1}^{l} p_i + \textstyle\sum_{i=1}^{l} x_i(k, \mathbf{z}) \log p_i \\
&= -\left(N - \textstyle\sum_{i=1}^{l} x_i(k, \mathbf{z})\right) \textstyle\sum_{i=1}^{l} p_i + \textstyle\sum_{i=1}^{l} x_i(k, \mathbf{z}) \log p_i \\
&= \textstyle\sum_{i=1}^{l} x_i(k, \mathbf{z}) \left(\log p_i + \textstyle\sum_{j=1}^{l} p_j\right) - N \textstyle\sum_{i=1}^{l} p_i.
\end{aligned}$$

If we define the vector $\mathbf{r}' = (r_1, \ldots, r_l)$ by $r_i = \log p_i + \sum_{j=1}^{l} p_j$ $(i = 1, \ldots, l)$, then the maximum likelihood estimate of the key value $k$ given the data $\mathbf{z}$ is essentially that value of $k$ which maximises

$$\mathbf{r}'^T X_i'(k, \mathbf{z}) = \sum_{i=1}^{l} X_i(k, \mathbf{z}) r_i.$$

In many cases, $r_i \approx \log p_i$, which gives the maximum likelihood estimate of the key value $k$ given the data $\mathbf{z}$ is essentially that value of $k$ which maximises

$$\Omega(k, \mathbf{z}) = \mathbf{q}'^T X_i'(k, \mathbf{z}) = \sum_{i=1}^{l} X_i(k, \mathbf{z}) \log p_i.$$

If required, it should be possible to derive approximate distributions for $\Phi'(k, \mathbf{z})$ when $k$ is the true key $k^*$ and when $k$ is not the true key. For example, if $p_1 = \ldots = p_l$, then $\frac{\Phi'(k, \mathbf{z})}{\log p_i}$ is approximately a Poisson random variable both when $k$ is the true key $k^*$ and otherwise. More generally, for a reasonable number $l$ of specified output differences, we would expect $\Omega(k, \mathbf{z})$ to have a normal distribution both for the true key value $k^*$ and for the other key values.

# 4 A Large Sample Version

For large $N$ and reasonable $\mathbf{p}$, this multinomial distribution $\mathbf{X}(k, \mathbf{z})$ is extremely well-approximated by an $(l+1)$-dimensional multivariate normal distribution, so we have

$$\mathbf{X}(k, \mathbf{z}) \sim \mathrm{N}\left(N\mathbf{p}; N\Sigma_{\mathbf{p}}\right) = \mathrm{N}\left(N\mathbf{p}; N\left(D(\mathbf{p}) - \mathbf{p}^T\mathbf{p}\right)\right).$$

If $\mathbf{1} = (1, \ldots, 1)^T$, then $\mathbf{1}^T\mathbf{X} = N$, so the random vector $\mathbf{X}$ is degenerate only having rank $l$. We therefore consider the full rank $l$-dimensional difference (deleted) count vector

$$\mathbf{X}'(k, \mathbf{z}) \sim \mathrm{N}\left(N\mathbf{p}'; N\Sigma_{\mathbf{p}'}\right) = \mathrm{N}\left(N\mathbf{p}'; N\left(D(\mathbf{p}') - \mathbf{p}'^T\mathbf{p}'\right)\right)$$

which has density function $f_{\mathbf{X}'}(\mathbf{z}; k)$ for data $\mathbf{z}$ and key value $k$ given by

$$\left|2\pi N\Sigma_{\mathbf{p}'}\right|^{-\frac{l}{2}} \exp\left(-\tfrac{1}{2}\ \left(\mathbf{x}'(k, \mathbf{z}) - N\mathbf{p}'\right)^T \left(N\Sigma_{\mathbf{p}'}\right)^{-1} \left(\mathbf{x}'(k, \mathbf{z}) - N\mathbf{p}'\right)\right).$$

This gives rise to a likelihood function $L(k; \mathbf{z})$ for the key parameter value $k$ given the data $\mathbf{z}$ satisfying

$$L(k; \mathbf{z}) \propto \exp\left(-\tfrac{1}{2}\ \left(\mathbf{x}'(k, \mathbf{z}) - N\mathbf{p}'\right)^T \left(N\Sigma_{\mathbf{p}'}\right)^{-1} \left(\mathbf{x}'(k, \mathbf{z}) - N\mathbf{p}'\right)\right),$$

which leads to a log-likelihood function $\mathcal{L}(k; \mathbf{z}) = \log L(k; \mathbf{z})$ given (up to an additive constant) by

$$\mathcal{L}(k; \mathbf{z}) = -\tfrac{1}{2}\ N^{-1} \left(\mathbf{x}'(k, \mathbf{z}) - N\mathbf{p}'\right)^T \Sigma_{\mathbf{p}'}^{-1} \left(\mathbf{x}'(k, \mathbf{z}) - N\mathbf{p}'\right).$$

We therefore consider the function

$$\Psi(k, \mathbf{z}) = N^{-1} \left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right)^T \Sigma_{\mathbf{p}'}^{-1} \left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right).$$

as this is essentially the negative of the log-likelihood function. The log-likelihood and hence the likelihood is maximised by the value of the key parameter $k$ for which the quadratic form $\Psi(k, \mathbf{z})$ is minimised.

We now give a simple form for this quadratic form $\Psi(k, \mathbf{z})$. It can easily be verified by direct calculation that $\Sigma_{\mathbf{p}'}^{-1} = D\left(\mathbf{p}'\right)^{-1} + p_0^{-1}\mathbf{1}\mathbf{1}^T$, so we have

$$\begin{aligned}\Psi(k, \mathbf{z}) \ &= \ N^{-1} \left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right)^T D\left(\mathbf{p}'\right)^{-1} \left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right) \\ &\quad + N^{-1}p_0^{-1} \left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right)^T \mathbf{1}\mathbf{1}^T \left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right).\end{aligned}$$

We can evaluate the two parts of this sum separately. The first part gives

$$\left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right)^T D\left(\mathbf{p}'\right) \left(\mathbf{X}'(k, \mathbf{z}) - N\mathbf{p}'\right) = \sum_{i=1}^{l} \frac{\left(X_i(k, \mathbf{z}) - Np_i\right)^2}{Np_i}.$$

The second part can be easily evaluated by noting that

$$(\mathbf{X}'(k,\mathbf{z}) - N\mathbf{p}')^T\mathbf{1} = \sum_{i=1}^{l} X_i(k,\mathbf{z}) - Np_i = -\left(X_0(k,\mathbf{z}) - Np_0\right).$$

This means that the maximum likelihood estimate for the key value $k$ is that value of $k$ which minimises the quadratic form

$$\Psi(k,\mathbf{z}) = \sum_{i=0}^{l} \frac{(X_i(k,\mathbf{z}) - Np_i)^2}{Np_i}.$$

The above argument is essentially that used to derive the test statistic for the (Pearson) $\chi^2$ goodness-of-fit test for the multinomial distribution [3]. A widely-used expression for this goodness-of-fit test statistic allows us to express the quadratic form $\Psi(k,\mathbf{z})$ to be minimised as

$$\Psi(k,\mathbf{z}) = \sum_{i=0}^{l} \frac{(\text{Observed}(k,\mathbf{z})_i - \text{Expected}_i)^2}{\text{Expected}_i}.$$

The usual "rule-of-thumb" in applying a goodness-of-fit test is that each "bucket" should be expected to have at least five items in it, that is to say $\text{Expected}_i \geq 5$. One usual way of achieving this is to combine smaller buckets together so as to obtain buckets where at least five items occur. However, even when this "five-item" rule is not achieved, $\Psi(k,\mathbf{z})$ is still a good test statistic; it is just that its distribution is unclear as the normal approximation is less accurate. In the case though when the normal approximation is accurate, the value of $\Psi(k,\mathbf{z})$ under the true key value $k^*$, $\Psi(k^*,\mathbf{z})$, has a $\chi^2$ distribution with $l$ degrees of freedom. The distribution under other key values can be obtained using standard theory [3].

# References


[1] E. Biham and A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*. Springer–Verlag, 1993.

[2] E. Biham and A. Shamir. Differential Cryptanalysis of the DES-like Cryptosystems. *Journal of Cryptology*, 4:3–72, 1993.

[3] W.G. Cochran. The $\chi^2$ Test of Goodness of Fit. *Annals of Mathematical Statistics*, 23:315–345, 1952.