On an Approach to Compute (at least Almost) Exact Probabilities of Differential Paths

Max Gebhardt, Georg Illies and Werner Schindler

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Godesberger Allee 185–189
53175 Bonn, Germany
{Maximilian.Gebhardt,Georg.Illies,Werner.Schindler}@bsi.bund.de

Keywords: Differential path, probability, stochastic model.

Extended Abstract

Differential attacks constitute powerful tools in cryptanalysis which are applied in different areas, e.g. to find keys of block ciphers or stream ciphers or to find collisions of hash functions.

The efficiency of a differential attack is closely related to the probability that pairs of intermediate values follow a particular differential path. From the designer's point of view the efficiency of an attack determines its risk potential. It is clearly desirable to know the probabilities of differential paths as exact as possible, especially if the estimated path probability implies a workload which appears to be in a region between practical feasibility and infeasibility.

More formally, we are interested in probabilities

$$\mathrm{Prob}((X_n, X'_n) \in B_n, (X_{n-1}, X'_{n-1}) \in B_{n-1}, \ldots, (X_0, X'_0) \in B_0) \qquad (1)$$

where $X_0, \ldots, X_n, X'_0, \ldots, X'_n$ denote random variables that assume values on a finite set Ω (typically, $\Omega = \{0,1\}^v$) while the subsets $B_0, \ldots, B_n \subseteq \Omega \times \Omega$ characterize conditions that define the differential path. This joint probability can be expressed as a product of conditional probabilities

$$\mathrm{Prob}((X_i, X'_i) \in B_i \mid (X_{i-1}, X'_{i-1}) \in B_{i-1}, \ldots, (X_0, X'_0) \in B_0) \text{ for } i \in \{1, \ldots, n\}. \qquad (2)$$

These conditional probabilities usually cannot be computed exactly. One reason is that the random variables X_i and X'_i are strongly correlated, which complicates concrete calculations unless $|\Omega|$ is very small or the sets B_i are extremely simple. Note that the pairs of random variables $(X_i, X'_i), (X_{i-1}, X'_{i-1}), \ldots$ are usually not independent, at least not in a strict sense, which causes further difficulties. For these reasons the conditional probabilities (2) can usually only be roughly estimated. For hash collision paths, for instance, the subsets B_i typically define conditions on particular bits, and $2^{-(\#\text{affected bits})}$ serves as an approximator for the unknown conditional probability.

In this talk we focus on conditional probabilities of the following type:

$$\mathrm{Prob}((Y_3, Y'_3) := ((Y_1, Y'_1) + (f(Y_2), f(Y'_2)) \in B_3 \mid (Y_1, Y'_1) \in B_1, (Y_2, Y'_2) \in B_2) \qquad (3)$$

with $\Omega := \{0,1\}^v$, and $B_1, B_2, B_3 \subseteq \Omega \times \Omega$ while "+" stands for the componentwise addition modulo 2^v. The subsets B_i are either of the type $\Delta_a = \{(w, w') \in \Omega \times \Omega \mid w - w' \equiv a (\bmod 2^v)\}$ for a 2^v-difference a, or B_i defines conditions on certain bits, e.g. $w_{17} = w'_{17}$, $(w_4, w'_4) = (0, 1)$ etc. The function f induces a cyclical shift on the arguments. For small v the conditional probability (3) can be determined exhaustively wheras large v requires more sophisticated methods.

We show that the understanding of the conditional probabilities of type (3) can help (under suitable circumstances) to simplify conditional probabilities on the product space $\Omega \times \Omega$ with strongly correlated components (2) to conditional probabilities on Ω, which clearly is an enormous advantage. Another goal is to find sufficient conditions so that the conditional random variable $(Y_3, Y'_3) \mid B_3$ is independent of $(Y_1, Y'_1) \mid B_1$ and $(Y_2, Y'_2) \mid B_2$. Depending on the concrete situation this may reduce the relevant part of the 'pre-history' in (2), which additionally simplifies calculations. Hence it is worthwhile to study conditional probabilities of type (3). We point out that our approach can be adapted to other finite groups than $\{0,1\}^v$ (equipped with the addition modulo 2^v), which clearly enlarges the field of applications.

The effectiveness of our approach was confirmed by practical experiments with MD5 hash collision paths. Based on a stochastic model with mild assumptions on the mixing properties of the MD5 step function the theorems on conditional probabilities of type (3) could be applied. The 'theoretically' derived path probabilities matched with empirical results. Compared with the 'straightforward' approximators for the path probabilities (obtained by 'bit counting') we obtained non-negligible 'correction factors' between 1/12 and 5 (cf. [1]), which in turn imply 'correction factors' between 1/5 and 12 on the expected workload of the collision attack.

References

1. M. Gebhardt, G. Illies, W. Schindler: Precise Probabilities of Hash Collision Paths. Second Cryptographic Hash Workshop http://www.csrc.nist.gov/pki/HashWorkshop/2006/Papers/