The Efficiency of Optimal Sampling in the Random S-box Model

Stankovski, Paul; Brynielsson, Lennart; Hell, Martin

Published in:
[Host publication title missing]

2014

The Efficiency of Optimal Sampling in the Random S-box Model

Paul Stankovski*, Lennart Brynielsson† and Martin Hell*
*Department of Electrical and Information Technology, Lund University, Lund, Sweden
Email: {paul,martin}@eit.lth.se
†MUST/TSA, Swedish Armed Forces, 107 86 Stockholm, Sweden
Email: lennart.brynielsson@mil.se

***Abstract*—In this paper we show a closed caption formula for the efficiency of the optimal sampling technique in the random S-box model. This formula is derived by analyzing the given model and sampling technique using statistical techniques. We further generalize the original random S-box model in two ways; allowing multiple-bit entries, xor of several random S-box outputs. For all cases we show the corresponding closed caption efficiency formula.**

Using these new formulas, it is now possible to instantaneously give accurate analytical estimates of the output quality of random S-boxes. This can be of great practical importance in, for example, analysis and design of cryptographic primitives based on such building blocks.

I. Introduction

In essence, the point of the random S-box model is to isolate a stream cipher building block for analysis. A random S-box can be thought of as a table with (pseudo-)random entries, or a boolean function chosen at random in some way. The random S-box model itself can be thought of as a very simple stream cipher (described in Section II-A).

This model is not only of theoretical interest. An optimal sampling technique for the random S-box model was shown in [4], and this sampling technique was used to produce the best known distinguisher for the eSTREAM [2] portfolio stream cipher HC-128 [6], see [3], [5].

In [4], it is specified how to efficiently perform optimal sampling in this model. However, neither explicit nor implicit formulas or expressions for the efficiency of the sampling technique are given. In this paper we remedy that situation by delivering explicit closed caption formulas, not only for the efficiency of optimal sampling in the original random S-box setting, but also for generalized versions of it.

We will use statistical machinery to derive these closed caption formulas, which are deceivingly simple in all generalizations of the random S-box model. The uniformity of the various formulas and their simplicity are clear indications that something fundamental is at work here.

The new formulas that we present are also of practical interest. They can be used for cryptanalysis of cryptographic primitives that utilize random S-boxes as building blocks, or by algorithm designers for assessing the security of such primitives in a better way than what was possible before.

The final publication is available at IEEE.

II. Preliminaries

The random S-box model and its generalizations are described in Section II-A. Hypothesis testing and the optimal sampling technique in the original random S-box model are reviewed in Sections II-B and II-C, respectively.

A. The Random S-box Model

Consider an S-box (or table) of size n with single-bit entries, and let its entries be initialized with random bits. That is, each single-bit entry is chosen uniformly at random from $\mathbb{B} = \{0, 1\}$. After initialization, output is produced from the S-box in the following way.

At each time instance, one of the n table entries are selected by drawing a table index uniformly at random from the index interval $[0, n-1]$. The single-bit entry in the given table slot is then used as an output bit. Output bits are continually produced in this way, but after every ℓ^{th} output bit, the S-box is reinitialized with new random entries. Each instantiation of the S-box is thus used for a duration of ℓ time instances, and the ℓ-bit output block produced during this process is called a *chunk*.

The random S-box model described above is the original model from [4]. One generalization of the original random S-box model is to take k different S-boxes and combine their outputs using bitwise addition modulo 2. The resulting chunk will still be ℓ bits long in this case, but each bit is the modulo 2 sum of k single S-box output bits. All S-boxes are individually reinitialized after each chunk has been produced.

Another generalization is to allow m-bit entries in the table, so that each table entry is drawn uniformly at random from $\mathbb{B}^m$. In this model, m bits are output at every time instance, for a total of ℓm bits in each chunk.

And of course, both generalization may be combined for an m-bit addition model, in which the addition is taken modulo 2^m over the corresponding m-bit S-box outputs.

To differentiate between the four different random S-box models presented here, we will use the following notation. Let model S, model SA, model M and model MA be shorthand notations for the original single-bit (S), single-bit with addition (SA), m-bit (M) and the conglomerate m-bit with addition (MA) versions of the model, respectively, as described above.

The random S-box has been used as a building block in stream ciphers for generation of pseudorandom keystream. In

this context it is clearly reasonable to measure the quality of the output in terms of the efficiency of the optimal distinguisher for the given output. This concept is central, and the necessary distinguishing tools will now be detailed in Sections II-B and II-C.

B. Hypothesis Testing

A hypothesis test is used at the core of a distinguisher in order to tell which of two probability distributions that is the most likely output sequence source.

Let the empirical probability distribution as defined by the sampling be denoted P^*. Let the corresponding (theoretical) probability distribution of the S-box be denoted P_1, and let P_2 denote its uniform probability distribution. The Neyman-Pearson lemma, see e.g., [1], provides the optimal hypothesis test.

Lemma 1 (Neyman-Pearson): Let $X_1, X_2, \ldots, X_t$ be independent and identically distributed random variables according to P^*. Consider the decision problem corresponding to the hypotheses $P^* = P_1$ vs. $P^* = P_2$. For $Q \geq 0$ define a region

$$\mathcal{A}_t(Q) = \left\{ \frac{P_1(x_1, x_2, \ldots, x_t)}{P_2(x_1, x_2, \ldots, x_t)} > Q \right\}.$$

Let $\alpha_t = P_1^t(\mathcal{A}_t^c(Q))$ and $\beta_t = P_2^t(\mathcal{A}_t(Q))$ be the error probabilities corresponding to the decision region $\mathcal{A}_t$. Let $\mathcal{B}_t$ be any other decision region with associated error probabilities α^* and β^*. If $\alpha^* \leq \alpha$, then $\beta^* \geq \beta$.

If we want to minimize the (unweighted) sum of the error probabilities, we set $Q = 1$. In other words, we decide $P^* = P_1$ if

$$\frac{P_1(x_1, \ldots, x_t)}{P_2(x_1, \ldots, x_t)} > 1 \underset{\text{indep.}}{\Leftrightarrow} \sum_{i=1}^{t} \log \frac{P_1(x_i)}{P_2(x_i)} > 0, \tag{1}$$

and $P^* = P_2$ otherwise. The equivalence in (1) is valid when the samples $x_1, \ldots, x_t$ are independent.

Let us now assess the efficiency of the hypothesis test. We need to introduce relative entropy, which can be thought of as a distance measure between probability distributions.

Definition 1 (Relative entropy): The relative entropy between two probability distributions P_1 and P_2 over the same domain $\mathcal{X}$ is defined as

$$D\left(P_1 \| P_2\right) = \sum_{x \in \mathcal{X}} P_1(x) \log \frac{P_1(x)}{P_2(x)}. \tag{2}$$

There are a few aliases for relative entropy in the literature; information divergence, Kullback-Leibler divergence, information gain and redundancy.

The Neyman-Pearson hypothesis test models independent and identically distributed samples drawn from a probability distribution P^*. There are two possible hypotheses, the null hypothesis H_0 and the alternate hypothesis H_1;

$$H_0 : P^* = P_1,$$
$$H_1 : P^* = P_2.$$

Two types of errors are possible in this hypothesis test.

Type I error:	Reject H_0 when it is true (prob. α).
Type II error:	Accept H_0 when H_1 is true (prob. β).

No universal expressions for α and β exist, so the performance of the test in the general case is not known. However, *asymptotic* expressions for these error probabilities do exist. The interplay between the asymptotic error probabilities and the relative entropy is described by Stein's lemma, which roughly states that β decreases so that

$$\lim_{t \to \infty} \frac{\log \beta}{t} = -D(P_1 \| P_2),$$

if the error probability α is fixed. Note that the magnitude of α does not affect the exponential rate at which β decreases. Asymptotically we can therefore write

$$\beta \approx 2^{-tD(P_1 \| P_2)},$$

so that the error probabilities of the hypothesis test start to decrease exponentially when the number of samples approaches

$$t = \frac{1}{D(P_1 \| P_2)}. \tag{3}$$

In a practical scenario, one would be required to use a small multiple of the number t as the number of samples needed by the distinguisher, but the number t as defined in Equation (3) can be seen as a baseline requirement for the number of samples that a distinguisher needs.

Note that sample requirement is fully determined by the divergence (relative entropy) between the two probability distributions P_1 and P_2.

If P_1 and P_2 are $\mathcal{N}(\mu_1, \sigma_1)$ and $\mathcal{N}(\mu_2, \sigma_2)$, respectively, then

$$D(P_1 \| P_2) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} \geq 0. \tag{4}$$

We will also be using the corresponding result for the multivariate normal distribution. In this case we let

$$\mu = \left[\mathrm{E}\left(X_1\right), \mathrm{E}\left(X_2\right), \cdots, \mathrm{E}\left(X_m\right)\right]$$

denote an m-dimensional mean vector, and let

$$\mathbf{C} = \left[\mathrm{Cov}(X_i, X_j)\right], \quad i, j = 1, 2, \cdots, m$$

denote a non-singular $m \times m$-dimensional covariance matrix. If P_1 and P_2 are $\mathcal{N}_m(\mu_1, \mathbf{C}_1)$ and $\mathcal{N}_m(\mu_2, \mathbf{C}_2)$, respectively, then

$$D(P_1 \| P_2) = \frac{1}{2}\Big(tr\left(\mathbf{C}_2^{-1}\mathbf{C}_1\right) + \mu_\Delta^T \mathbf{C}_2^{-1} \mu_\Delta - m - \log\det\left(\mathbf{C}_2^{-1}\mathbf{C}_1\right)\Big), \tag{5}$$

where $\mu_\Delta = \mu_2 - \mu_1$.

C. Optimal Sampling for Model S

An optimal sampling technique for distinguishing the output sequence in the single-bit model S from a truly random sequence was described in [4].

Let s_i denote a single-bit observation from the given S-box at time i. Taking entire ℓ-bit chunk vectors $(s_1, s_2, \ldots, s_\ell)$ as samples is obviously optimal in an information theoretical

Algorithm I – Weight Distribution (**wd**)

Input: S-box size n, vector length ℓ, current depth d, current probability p, probability distribution container $dist$ of length $\ell+1$, weight w, number of opened table entries with zeros a_0, number of opened table entries with ones a_1.
Output: probability distribution $dist$.
Initial recursion parameters: $dist$ zeroized, $(d, p, w, a_0, a_1) = (0, 1, 0, 0, 0)$.

if $(d == \ell)$ { $dist[w]$ += p; **return**; }
if $(a_0 > 0)$ $\mathbf{wd}(dist, n, \ell, d+1, p \cdot \frac{a_0}{n}, w, a_0, a_1)$; /* old 0 */
if $(a_1 > 0)$ $\mathbf{wd}(dist, n, \ell, d+1, p \cdot \frac{a_1}{n}, w+1, a_0, a_1)$; /* old 1 */
if $(a_0 + a_1 < n)$ { /* table not exhausted */
  $\mathbf{wd}(dist, n, \ell, d+1, p \cdot \frac{n-(a_0+a_1)}{2n}, w, a_0+1, a_1)$; /* new 0 */
  $\mathbf{wd}(dist, n, \ell, d+1, p \cdot \frac{n-(a_0+a_1)}{2n}, w+1, a_0, a_1+1)$; /* new 1 */
}

sense. However, the weight sampling technique (WS), in which we take chunk weights $\|(s_1, s_2, \ldots, s_\ell)\|_1 = \sum_{i=1}^{\ell} s_i$ as samples, is information theoretically equivalent and computationally more efficient (see [4]). The corresponding weight distributions P_1 and P_2 have domains of size $\ell+1$.

For the uniform probability distribution P_2, every vector is equally likely. The resulting chunk weight probability distribution is therefore combinatorially determined by

$$P_2(w) = \binom{\ell}{w} 2^{-\ell} \tag{6}$$

for all possible chunk weights $0 \le w \le \ell$.

P_1 can be calculated according to Algorithm I, which is stated recursively for simplicity, but can also be implemented in a dynamic programming fashion.

While the above describes an optimal sampling technique for model S, no general formula for its efficiency is known to date. Judging by the complexity of the explicit construction of the probability distribution P_1 in Algorithm I, it may surprise the reader to find that such a closed caption formula not only exists, but that it is also simple.

III. Statistical Analysis of Model S

In model S, an S-box B of size n is initialized by drawing each of the n single-bit entries uniformly at random from $\{0,1\}$. Let Z denote the number of one bits in B. Then $Z \in \mathrm{Bin}\left(n, \frac{1}{2}\right)$, for which we have

$$\mathrm{E}\left[Z\left(n-Z\right)\right] = (n-1)\mathrm{Var}\left[Z\right]. \tag{7}$$

Let Y denote the number of ones in a model S ℓ-bit chunk, and conditioned on $Z = z$, we have $Y \in \mathrm{Bin}\left(\ell, \frac{z}{n}\right)$. Using the laws of total expectation and variance together with Equation (7) we get

$$\mathrm{E}\left[Y\right] = \mathrm{E}\Big[\mathrm{E}\left[Y|Z\right]\Big] = \mathrm{E}\left[\frac{\ell Z}{n}\right] = \frac{\ell}{2}$$

and

$$\begin{aligned}\mathrm{Var}\left[Y\right] &= \mathrm{E}\Big[\mathrm{Var}\left[Y|Z\right]\Big] + \mathrm{Var}\Big[\mathrm{E}\left[Y|Z\right]\Big] \\ &= \mathrm{E}\left[\frac{\ell Z\left(n-Z\right)}{n^2}\right] + \mathrm{Var}\left[\frac{\ell Z}{n}\right] \\ &= \frac{\ell}{4}\underbrace{\left(1+\frac{\ell-1}{n}\right)}_{=\gamma} = \frac{\gamma\ell}{4}.\end{aligned}$$

One can see that $\mathrm{E}\left[Y\right]$ has the same value as in the uniform case in which every chunk bit is chosen uniformly at random from $\{0,1\}$. However, one can also see that $\mathrm{Var}\left[Y\right]$ is enlarged by a factor of $\gamma = 1 + \frac{\ell-1}{n}$. If we assume that the two probability distributions are approximately normal, then we can apply Equation (4) using $\mu_1 = \mu_2 = \frac{\ell}{2}$, $\sigma_1^2 = \frac{\gamma\ell}{4}$ and $\sigma_2^2 = \frac{\ell}{4}$ to calculate their divergence according to

$$D\left(P_1 \| P_2\right) = \frac{1}{2}\left(\gamma - 1 - \log\gamma\right) \tag{8}$$

$$\approx \frac{1}{4}\left(\gamma - 1 - (\gamma-1) + \frac{(\gamma-1)^2}{2}\right) \tag{9}$$

$$= \left(\frac{\ell-1}{2n}\right)^2$$

nats, which should be divided by $\ln 2$ for bits.

In Section VII, simulations will show that Equation (9) and its subsequent generalizations are indeed accurate for practical applications.

IV. Statistical Analysis of Model SA

A chunk in model SA is formed by addition (xor) of k independent model S chunks. Consider first the case $k = 2$, which adds two independently generated model S chunks.

An observation may be made here. Model SA chunks may be viewed as the output of a *larger* S-box of size n^2, formed by modular addition of the entries of the two corresponding model S S-boxes of size n. Note that the single-bit entries obtained in this way are pairwise independent. The number of ones Z in the large model SA S-box of size n^2 is the integer sum of all n^2 entries. That is, Z is a sum of n^2 uncorrelated and balanced bits. From this it follows that

$$\begin{aligned}&\mathrm{E}\left[Z\right] = \frac{n^2}{2}, \\ &\mathrm{Var}\left[Z\right] = \frac{n^2}{4} \quad \text{and} \\ &\mathrm{E}\left[Z\left(n^2-Z\right)\right] = \left(n^2-1\right)\mathrm{Var}\left[Z\right].\end{aligned}$$

Letting Y denote the number of ones in a model SA chunk, we now get

$$\begin{aligned}&\mathrm{E}\left[Y\right] = \frac{\ell}{2} \quad \text{and} \\ &\mathrm{Var}\left[Y\right] = \frac{\ell}{4}\left(1+\frac{\ell-1}{n^2}\right)\end{aligned}$$

when $k = 2$.

The above observation also applies in the more general setting of an arbitrary but fixed number k of S-boxes. Applying Equation (4) once more in the same way as in Section III, we get

$$D\left(P_1 \| P_2\right) = \left(\frac{\ell - 1}{2n^k}\right)^2 \tag{10}$$

for the model SA case. This is, again, expressed in nats, so division by $\ln 2$ is appropriate for bits.

At this point it is possible to verify the sanity of the derived formulas. By direct comparison to the values in Table 3 in [5], it is clear that the expression in Equation (10) is very reasonable.

V. Statistical Analysis of Model M

Now assume that each S-box slot is initialized by selecting a value in $[0, M-1]$ uniformly at random, and that ℓ slots are then selected uniformly at random (with repetition) for chunk output. The value M can be thought of as an m-bit number; $M = 2^m$.

Let Z_u denote the number of table slots that contain the value $u \in [0, M-1]$, so that $Z_u \in \mathrm{Bin}\left(n, \frac{1}{M}\right)$. Also, let Y_u denote the number of times that the value u appears in the chunk. Conditioned on $Z_u = z$, we have $Y_u \in \mathrm{Bin}\left(\ell, \frac{z}{M}\right)$. Similarly to the calculations in Section III, we get

$$\mathrm{E}\left[Y_u\right] = \mathrm{E}\Big[\mathrm{E}\left[Y_u | Z_u\right]\Big] = \mathrm{E}\left[\frac{\ell Z_u}{n}\right] = \frac{\ell}{M}$$

and

$$\begin{aligned}
\mathrm{Var}\left[Y_u\right] &= \mathrm{E}\Big[\mathrm{Var}\left[Y_u | Z_u\right]\Big] + \mathrm{Var}\Big[\mathrm{E}\left[Y_u | Z_u\right]\Big] \\
&= \mathrm{E}\left[\frac{\ell Z_u \left(n - Z_u\right)}{n^2}\right] + \mathrm{Var}\left[\frac{\ell Z_u}{n}\right] \\
&= \ell \frac{M-1}{M^2} \underbrace{\left(1 + \frac{\ell - 1}{n}\right)}_{=\gamma}.
\end{aligned}$$

Compared to the uniform case, variable Y_u has the same expected value, but its variance is enlarged by a factor of γ.

Now consider the covariance matrix

$$\mathbf{C}_1 = \left[\mathrm{Cov}\left(Y_u, Y_v\right)\right], \quad u, v = 0, \cdots, M-1.$$

By symmetry, all covariances outside the diagonal must be equal and $\sum Y_u = \ell$ is constant, so all values in the covariance matrix must sum to zero. The covariance matrix $\mathbf{C}_1$ must then be the same as for the multivariate normal case, but multiplied by a factor of γ, so that $\mathbf{C}_1 = \gamma \mathbf{C}_2$.

The divergence between two multidimensional normal distributions with the same mean is given by applying $\mu_\Delta = \mathbf{0}$ to Equation (5). However, when approximating a multinomial distribution with a normal one, the covariance matrix becomes singular since the sum of the variables is constant. Normal approximation is still possible by a reducing the dimensions of $\mathbf{C}_1$ and $\mathbf{C}_2$ by one by removing one row and one column[1].

[1]For the case $M = 2$, this corresponds to counting only ones.

Let $\mathbf{C}_1^{'}$ and $\mathbf{C}_2^{'}$ denote the covariance matrices with reduced dimensions.

Now applying $\mathbf{C}_1^{'} = \gamma \mathbf{C}_2^{'}$ and $\mu_\Delta = \mathbf{0}$ to Equation (5), we have

$$\begin{aligned}
D(P_1 \| P_2) &= \frac{1}{2}\Big(tr\left(\gamma \mathbf{I}^{'}\right) - (M-1) - \log\det\left(\gamma \mathbf{I}^{'}\right)\Big) \\
&= \frac{M-1}{2}\Big(\gamma - 1 - \log\gamma\Big) \qquad (11) \\
&\approx \frac{M-1}{2}\Big(\gamma - 1 - (\gamma - 1) + \frac{(\gamma-1)^2}{2}\Big) \qquad (12) \\
&= \left(\frac{\ell - 1}{2n}\right)^2 (M-1), \qquad (13)
\end{aligned}$$

which gives us the general efficiency formula for model M.

VI. Statistical Analysis of Model MA

It is also possible to combine several model M chunks into one model MA chunk. The calculations here are analogous to those in Section IV, extending the addition operator from single-bit addition modulo 2 (single-bit xor) to any m-bit addition operator that has a corresponding subtraction operator, such as addition modulo M or bitwise m-bit xor.

Combining k model M chunks, the divergence becomes

$$D(P_1 \| P_2) = \left(\frac{\ell - 1}{2n^k}\right)^2 (M-1). \tag{14}$$

To be divided by $\ln 2$ for conversion from nats to bits.

Note that Equation (14) reduces to Equation (10) for the binary case $M = 2$.

VII. Simulation results

Simulations have been performed to verify the validity of the analytically derived formulas. Following the notation in Section II-B, simulations were performed as follows.

The theoretical probability distribution P_1 of the chunks was derived using Algorithm II, which updates Algorithm I to take multiple-bit table entries into account. For notation, here, we let $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ denote the i^{th} unit vector in the natural way. Algorithm II is, again, presented recursively for simplicity, but it is possible to employ a dynamic programming approach for efficiency. This is what we have done to produce our simulation results.

The corresponding uniform distribution P_2 can be combinatorially determined by generalizing the chunk weight expression in Equation (6) to

$$P_2\left(\mathbf{w}\right) = \binom{\ell}{w_0, \ldots, w_{M-1}} M^{-\ell}.$$

The divergence $D\left(P_1 \| P_2\right)$ for various chunk lengths ℓ is plotted in Figures 1 and 2. The theoretical model is represented by the solid curve, and the approximation curve defined by Equation (13) is overlaid (dotted curve).

A few typical values of M and n were selected. Figures 1 and 2 depict $M = 2$ and 256 (1- and 8-bit table values), respectively. When comparing Figures 1 and 2, note that the chunk length axes differ.

Algorithm II – M-Weight Distribution ($\mathbf{mwd}$)

Input: S-box size n, maximum entry size M, vector length ℓ, current depth d, current probability p, probability distribution container $dist$ of length $\ell + 1$, weight vector $\mathbf{w} = (w_0, \ldots, w_{M-1})$ where w_i denotes the number of times that value i appears in the chunk, vector $\mathbf{a} = (a_0, \ldots, a_{M-1})$ where a_i denotes the number of opened table entries with value i.
Output: probability distribution $dist$.
Initial recursion parameters: $dist$ zeroized, $(d, p, \mathbf{w}, \mathbf{a}) = (0, 1, (0, \ldots, 0), (0, \ldots, 0))$.

```
if (d == ℓ) { dist[w] += p; return; }
for (i = 0; i < M; i++) {
    if (a_i > 0) { /* old value i */
        mwd(dist, n, M, ℓ, d + 1, p · a_i/n, w + e_i, a);
    }
}
if (||a||_1 < n) { /* table not exhausted */
    for (i = 0; i < M; i++) { /* new value i */
        mwd(dist, n, M, ℓ, d + 1, p · (n − ||a||_1)/(Mn), w + e_i, a + e_i);
    }
}
```

One may further note that the approximation given by Equation (11) only has one source of error, namely the (multivariate) normal approximation. The dotted curve representing Equation (13) has one additional error source, that of the Taylor expansion in Equation (12), which converges only when $\ell \leq n$.

The data show that the accuracy of the approximation formula increases as the table size n grows. This can also be seen analytically as the error term in the Taylor approximation in Equation (12) diminishes as $n \to \infty$.

A more detailed analysis of the data, not visible in the graph, shows that the Taylor approximation dominates the resulting error for small values of M, while the normal approximation dominates it for larger M.

The approximation suffers from inaccuracies when it comes to very short chunk lengths. At a chunk length of two—the worst case, the estimated divergence halves the actual divergence. This initial approximation behavior can be explained with refined analyses involving Walsh- and Fourier transforms, ultimately providing even better approximation formulas, but such analysis is out of scope for this paper.

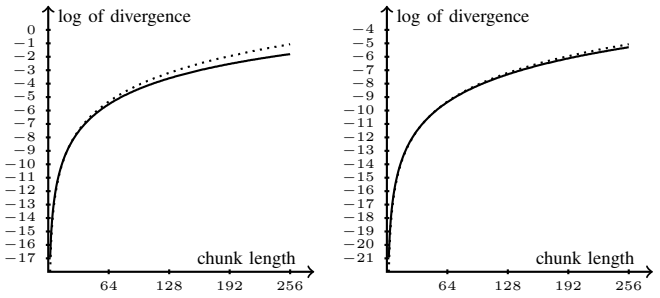

Fig. 1. Actual divergence for $M = 2$ with table size $n = 256$ (left) and $n = 1024$ (right) according to theoretical model (solid), approximation according to Equation (13) (dotted).

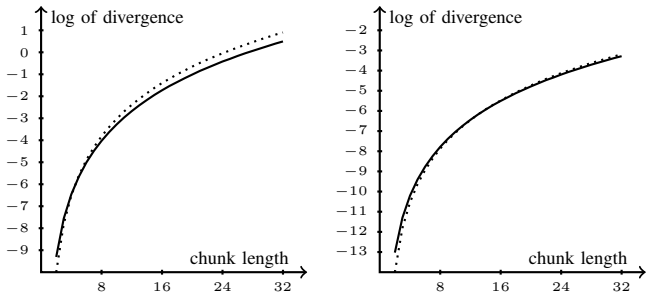

Fig. 2. Actual divergence for $M = 256$ with table size $n = 256$ (left) and $n = 1024$ (right) according to theoretical model (solid), approximation according to Equation (13) (dotted).

Furthermore, for larger chunk lengths, the approximation formula tends to show an upper limit for the actual divergence. This is particularly useful in practical scenarios where a designer of cryptographic algorithms wants to prove a limit to the usefulness of statistical analysis against a random S-box building block.

VIII. Concluding remarks

A closed caption formula for the efficiency of optimal sampling in the random S-box model was shown. Also, the random S-box model was generalized to include both multiple-bit S-box entries and the summation of several chunks. For these random S-box model variants, separately and together, we showed the corresponding efficiency formulas.

Note that the *optimal* sampling technique was analyzed here. The consequence of this is that we can now, for the first time, quantify to which degree a given random S-box is susceptible to statistical cryptanalysis. This, in turn, enables designers of cryptographic algorithms to provide proofs of non-susceptibility. That is, an algorithm designer can use our results to tune the parameters of one or several concurrent random S-boxes to resist statistical analysis (distinguishers).

References

[1] T. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley series in Telecommunication. Wiley, 1991.

[2] ECRYPT. eSTREAM: ECRYPT Stream Cipher Project, IST-2002-507932. Available at: *http://www.ecrypt.eu.org/stream/*. Last accessed on January 14, 2011.

[3] P. Stankovski. *Cryptanalysis of Selected Stream Ciphers*. PhD thesis, Lund University, June 2013.

[4] P. Stankovski and M. Hell. An Optimal Sampling Technique for Distinguishing Random S-boxes. In *ISIT12*, pages 846–850, July 2012.

[5] P. Stankovski, S. Ruj, M. Hell, and T. Johansson. Improved Distinguishers for HC-128. *Designs, Codes and Cryptography*, pages 1–16, 2012. *http://dx.doi.org/10.1007/s10623-011-9550-9*.

[6] H. Wu. The Stream Cipher HC-128. In *New Stream Cipher Designs*, volume 4986 of *Lecture Notes in Computer Science*, pages 39–47. Springer-Verlag, 2008.