On Metrics to Quantify the Inter-Device Uniqueness of Physically Unclonable Functions

Linus Feiten, Matthias Sauer, and Bernd Becker
Chair of Computer Architecture, University of Freiburg
Georges-Köhler-Allee 51, 79110 Freiburg, Germany
{feiten,sauerm}@informatik.uni-freiburg.de

Abstract. Physically Unclonable Functions (PUFs) have been an emerging topic in hardware security and trust in recent years, and many different kinds of PUFs have been presented in the literature. An important criterion is always the diversity of PUF responses for different devices, called inter-device uniqueness. A very popular uniqueness metric consists of calculating the pairwise hamming distance between the response bit-strings of all devices, assuming that all response bits are uncorrelated. Such correlations, however, should be regarded when a statement about inter-device uniqueness is made. We therefore propose a novel correlation metric to fulfil this requirement. Furthermore, we show that the hamming distance metric is actually redundant when at the same time the also popular bit-aliasing metric is applied.

Keywords: PUF, Physically Unclonable Function, Uniqueness, Metric, Hamming Distance, Bit Aliasing, Correlation

1 Introduction

Physically unclonable functions (PUFs) are an emerging topic in hardware security. They provide an alternative to storing cryptographic keys in non-volatile memory that might be susceptible to manipulation or key extraction. A PUF, on the other hand, generates a unique signature for each device it is implemented on, based on the device’s physical characteristics. Hence, any physical tampering attack to extract the signature may already lead to changes in these characteristics and thereby alter the signature.

Many different ways to implement PUFs on integrated circuits have been proposed in recent years; e.g. arbiter PUFs [10], butterfly PUFs [9], ring oscillator (RO)-PUFs [14,16,17], TERO-PUFs [2] or SRAM-PUFs [6,1]. For all these PUFs it is important to quantify their ability to uniquely identify a device. I.e. the generated signature should be unique for each device and knowing the signature of one device should not allow for any conclusions about other devices’ signatures.

Some very popular metrics to quantify this inter-device uniqueness are presented in [12] and have since been used in several publications, e.g. [4,7,13,5,8]. Section 2 recapitulates these metrics providing a novel analysis of their extreme
values. Section 3 then shows that the so-called hamming distance metric is actually redundant in the face of the bit-aliasing metric, that is mostly applied in addition to the former. Section 4 shows that neither of these two metrics is able to recognise correlations between signature bits, resulting in misleading judgements of inter-device uniqueness. To overcome this, a new metric is suggested that takes correlated bits into account. Section 5 concludes the paper.

2 Traditional Uniqueness Metrics

Let $m$ be the number of different devices and let $n$ be the total number of PUF response bits generated per device. While a single response bit $r_{i,j}$ with $1 \leq i \leq m$ and $1 \leq j \leq n$ can have different outcomes on a single device (poor intra-device reliability), we define the signature bit $\text{sig}_{i,j}$ to be the bit’s outcome irrespective of poor reliability. The signature bit can thus be considered the outcome after an error correction has been applied. When analysing inter-device uniqueness, only the signature bits outcomes need to be regarded.

2.1 Inter-device hamming distance

When $\text{Sig}_i = (\text{sig}_{i,1}, \text{sig}_{i,2}, \ldots, \text{sig}_{i,n})$ is the $n$-bit signature of a device $i$, [12] defines the Hamming Distance (HD) metric as

$$HD_{\text{metric}} = \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} \frac{HD(\text{Sig}_u, \text{Sig}_v)}{n}$$

where

$$HD(\text{Sig}_u, \text{Sig}_v) = \sum_{j=1}^{n} (\text{sig}_{u,j} \oplus \text{sig}_{v,j})$$

Thus, with $HD(\text{Sig}_u, \text{Sig}_v) = 0$ for $\text{Sig}_u$ being bitwise equal to $\text{Sig}_v$, and $HD(\text{Sig}_u, \text{Sig}_v) = n$ for $\text{Sig}_u$ being the bitwise inverse of $\text{Sig}_v$, the minimum outcome of $HD_{\text{metric}} = 0.0$ is achieved when all devices generate the same signature (worst inter-device uniqueness). The maximal outcome of $HD_{\text{metric}}$ converges to 0.5 with growing number of devices $m$. To understand this, we first realise that $HD_{\text{metric}}$ can be calculated by regarding each signature bit $j$ separately:
On Metrics to Quantify the Inter-Device Uniqueness of PUFs

\[ HD_{\text{metric}} = \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} \frac{HD(Sig_u, Sig_v)}{n} \]

\[ = \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} \frac{1}{n} \sum_{j=1}^{n} (sig_{u,j} \oplus sig_{v,j}) \]

\[ = \frac{1}{n} \sum_{j=1}^{n} \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (sig_{u,j} \oplus sig_{v,j}) \]

The maximum of \( HD_{\text{metric}} \) is thus achieved, when \( \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (sig_{u,j} \oplus sig_{v,j}) \) is maximal for each bit \( j \). This is the case, when exactly \( \frac{m}{2} \) devices generate \( sig_{i,j} = 0 \) and the other \( \frac{m}{2} \) devices generate \( sig_{i,j} = 1 \) (without loss of generality let \( m \mid 2 \)). The sum is incremented by 1 for each XOR of a “0” bit with a “1” bit. As there are \( \frac{m}{2} \) many “0” bits and \( \frac{m}{2} \) many “1” bits, the sum’s maximal outcome is:

\[ \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (sig_{u,j} \oplus sig_{v,j}) = \frac{m}{2} \cdot \frac{m}{2} = \frac{m^2}{4} \]

Inserting this into the formula of \( HD_{\text{metric}} \) we get:

\[ \frac{1}{n} \sum_{j=1}^{n} \frac{2}{m(m-1)} \cdot \frac{m^2}{4} = \frac{1}{2} \cdot \frac{m}{m-1} \]

Thus, the more devices are involved, the closer the maximal \( HD_{\text{metric}} \) outcome is to 0.5 – indicating optimal inter-device uniqueness. For smaller device populations, however, greater \( HD_{\text{metric}} \) outcomes must be achieved to demonstrate the same quality of uniqueness. Table 1 shows the maximal outcomes for different numbers of devices.

<table>
<thead>
<tr>
<th>Number of devices ( m )</th>
<th>Maximal ( HD_{\text{metric}} ) outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.0000</td>
</tr>
<tr>
<td>4</td>
<td>0.6667</td>
</tr>
<tr>
<td>8</td>
<td>0.5714</td>
</tr>
<tr>
<td>14</td>
<td>0.5385</td>
</tr>
<tr>
<td>20</td>
<td>0.5263</td>
</tr>
<tr>
<td>40</td>
<td>0.5128</td>
</tr>
<tr>
<td>60</td>
<td>0.5085</td>
</tr>
<tr>
<td>100</td>
<td>0.5051</td>
</tr>
</tbody>
</table>
2.2 Bit-Aliasing

Another metric introduced in [12] is bit-aliasing. Here, each signature bit \( j \) is analysed individually to detect whether its outcome is biased toward either 0 or 1 over all \( m \) devices:

\[
BA_j = \frac{1}{m} \sum_{i=1}^{m} s_{i,j}
\]

An unbiased bit should have a \( BA_j \) outcome close to 0.5, which is the case when half of the devices generate the bit as 0 and the other half generate it as 1. An outcome of 0.0 (or 1.0) means that the bit was 0 (or 1) on all devices, which is an indicator for poor inter-device uniqueness.

3 Redundancy of Hamming Distance Metric in the Face of Bit-Aliasing Metric

In this work, we show that \( HD_{\text{metric}} \) is in fact equivalent to the \( BA_j \). I.e. if the \( BA_j \) value of some or all bits indicates poor uniqueness, \( HD_{\text{metric}} \) indicates poor uniqueness as well, and vice versa. There is no case in which one metric indicates something that is not also indicated by the other.

This is why \( HD_{\text{metric}} \) can be considered redundant, especially as \( BA_j \) is more expressive identifying single signature bits with poor uniqueness and indicating their bias toward 0 or 1, whereas \( HD_{\text{metric}} \) only produces a single number between 0 and 0.5 for all bits. If such a single number is desired, it could also be derived from the \( BA_j \) values as follows. First, a normalised version of \( BA_j \) is calculated that does not differentiate any more between biases toward 0 or 1:

\[
BA_{j}^{\text{norm}} = 0.5 - |0.5 - BA_j|
\]

Then, the average over all signature bits is calculated:

\[
BA_{\text{metric}} = \frac{1}{n} \sum_{j=1}^{n} BA_{j}^{\text{norm}}
\]

Next, we show how \( HD_{\text{metric}} \) and \( BA_{\text{metric}} \) do actually differ in just a slight manner. First notice, that both metrics are in fact averages over all signature bits. This is obvious for the above definition of \( BA_{\text{metric}} \) and has been shown for \( HD_{\text{metric}} \) in 2.1 by rearranging the sums to:

\[
HD_{\text{metric}} = \frac{1}{n} \sum_{j=1}^{n} \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (s_{u,j} \oplus s_{v,j})
\]

Thus, both metrics do not consider correlations between signature bits, such that for an analysis of the metrics only one single signature bit \( j \) needs to be regarded.
Let $0 \leq \gamma \leq m$ be the number of devices for which the regarded signature bit’s outcome is 1; for the remaining $m - \gamma$ devices it is 0. For $\gamma = 0$ ($\forall i: \text{sig}_{i,j} = 0$) and $\gamma = m$ ($\forall i: \text{sig}_{i,j} = 1$), both metrics produce the outcome 0.0. For $\gamma = \frac{m}{2}$ (optimal distribution of 0s and 1s) both assume $\sim 0.5$. The slight difference is in how both metrics are mapping the $\gamma$ values in between.

Figure 1 shows the results of both metrics (y-axis) for the range of $\gamma$ values (x-axis). With $BA_j = \frac{1}{m} \sum_{i=1}^{m} \text{sig}_{i,j} = \frac{\gamma}{m}$, the plot for $BA_{\text{metric}}$ is:

$$BA_{\text{plot}}(\gamma) = 0.5 - |0.5 - \frac{\gamma}{m}|$$

![Figure 1](image_url)

**Fig. 1.** The results of $HD_{\text{metric}}$ and $BA_{\text{metric}}$ when applied to a single signature bit. $0 \leq \gamma \leq m$ is the number of devices for which this bit is 1.

To construct the plot for $HD_{\text{metric}}$, its definition must first be rewritten to:

$$\frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (\text{sig}_{u,j} \oplus \text{sig}_{v,j})$$

$$= \frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (\text{sig}_{u,j} + \text{sig}_{v,j} - 2(\text{sig}_{u,j} \cdot \text{sig}_{v,j}))$$

$$= \frac{2}{m(m-1)} \left( \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} \text{sig}_{u,j} + \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} \text{sig}_{v,j} - \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} 2(\text{sig}_{u,j} \cdot \text{sig}_{v,j}) \right)$$
\[
\begin{align*}
&= \frac{2}{m(m-1)} \left( \sum_{u=1}^{m-1} (m-u) \cdot \text{sig}_{u,j} + \sum_{v=2}^{m} (v-1) \cdot \text{sig}_{v,j} \\
&\quad - \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} 2(\text{sig}_{u,j} \cdot \text{sig}_{v,j}) \right) \\
&= \frac{2}{m(m-1)} \left( \sum_{w=1}^{m} (m-w) \cdot \text{sig}_{w,j} + \sum_{w=1}^{m} (w-1) \cdot \text{sig}_{w,j} \\
&\quad - \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} 2(\text{sig}_{u,j} \cdot \text{sig}_{v,j}) \right) \\
&= \frac{2}{m(m-1)} \left( \sum_{w=1}^{m} ((m-w) + (w-1)) \cdot \text{sig}_{w,j} \\
&\quad - \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} 2(\text{sig}_{u,j} \cdot \text{sig}_{v,j}) \right) \\
&= \frac{2}{m(m-1)} \sum_{w=1}^{m} (m-1) \cdot \text{sig}_{w,j} \\
&\quad - \frac{4}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (\text{sig}_{u,j} \cdot \text{sig}_{v,j}) \\
&= \frac{2}{m} \sum_{w=1}^{m} \text{sig}_{w,j} - \frac{4}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (\text{sig}_{u,j} \cdot \text{sig}_{v,j}) \\
\end{align*}
\]

With
\[
\sum_{w=1}^{m} \text{sig}_{w,j} = \gamma
\]

and
\[
\sum_{u=1}^{m-1} \sum_{v=u+1}^{m} (\text{sig}_{u,j} \cdot \text{sig}_{v,j}) = \frac{\gamma(\gamma-1)}{2}
\]

we get:
\[
HD_{\text{plot}}(\gamma) = 2 \cdot \left( \frac{\gamma}{m} - \frac{\gamma(\gamma-1)}{2m(m-1)} \right)
\]

As Figure 1 shows, the only difference between $\text{HD}_{\text{metric}}$ and $\text{BA}_{\text{metric}}$ is that $\text{HD}_{\text{metric}}$ penalises deviations from $\gamma = \frac{m}{2}$ less than $\text{BA}_{\text{metric}}$. In fact, the value of $\text{HD}_{\text{metric}}$ can even be directly calculated from the $\text{BA}_{j}$ values by transforming the $\text{BA}_{\text{plot}}$ values of each signature bit $j$ into the corresponding $\text{HD}_{\text{plot}}$ value, and then taking the average of all bits:
For each bit $j$ the value of $BA_j$ is transformed into $BA^\text{norm}_j$ as defined above:

$$BA^\text{norm}_j = 0.5 - |0.5 - BA_j|$$

Then the inverse function of $BA_{\text{plot}}$ calculates the corresponding $\gamma_j$ for that $BA^\text{norm}_j$ (as the plots are symmetric, it does not matter that we only get values $\gamma_j \leq \frac{m}{2}$):

$$\gamma_j = m \cdot BA^\text{norm}_j$$

This $\gamma_j$ is then inserted into $HD_{\text{plot}}$ for each bit $j$ and the average of the results is calculated:

$$HD_{\text{metric}} = \frac{1}{n} \sum_{j=1}^{n} \left( 2 \cdot \left( \frac{\gamma_j}{m} - \frac{\gamma_j(\gamma_j - 1)}{m(m-1)} \right) \right)$$

### 4 New Correlation Sensitive Metric (CSM)

As mentioned in the previous section, the traditional uniqueness metrics only analyse one signature bit at a time and – if applicable – calculate the average over all bits. Thus, correlations between signature bits are not registered. The undesirable effect of that can be illustrated by the following example. Let $\frac{m}{2}$ devices have the signature 11...1100...00 and the other $\frac{m}{2}$ devices have the signature 00...0011...11. Both $HD_{\text{metric}}$ and $BA_j$ would produce their optimal values of $\sim 0.5$, while the signatures are not actually unique.

Hence, good outcomes of $HD_{\text{metric}}$ or $BA_j$ are only necessary conditions for inter-device uniqueness but not sufficient – unless the a priori assumption is made that the signature bits are uncorrelated. In our view, however, a metric quantifying inter-device uniqueness should not rely on such an assumption, because unforeseen correlations might always be there. This is why we are presenting a new alternative metric.

#### 4.1 Definition of New Metric

To identify correlated signature bits, we suggest to compare the outcomes of the $n$ signature bits of each device $i$ with one another, resulting in $\frac{n(n-1)}{2}$ pairings:

$$\text{cor}_{1,2}^i \quad \text{cor}_{1,3}^i \quad \ldots \quad \text{cor}_{i,n}^i$$

$$\text{cor}_{2,3}^i \quad \ldots \quad \text{cor}_{i,n}^i$$

$$\ldots$$

$$\text{cor}_{n-1,n}^i$$

with

$$\text{cor}_{j,k}^i = \begin{cases} 
1, & \text{if } \text{sig}_{i,j} = \text{sig}_{i,k} \\
-1, & \text{otherwise}
\end{cases}$$
The $cor_{j,k}^i$ values for each pairing are then summed up over all $m$ devices:

$$cor_{j,k} = \frac{1}{m} \sum_{i=1}^{m} cor_{j,k}^i$$

such that $cor_{j,k} = 1.0$ means a positive correlation ($sg_i,j = sg_i,k$ on all devices) and $cor_{j,k} = -1.0$ means a negative correlation ($sg_i,j \neq sg_i,k$ on all devices). $cor_{j,k} \approx 0.0$, on the other hand, indicates the absence of systemic correlations between the respective bits $j$ and $k$.

There are $2m+1$ many possible outcomes for each of the $\frac{n(n-1)}{2}$ many $cor_{j,k}$ values:

$$-1.0 = \frac{-m}{m}, \frac{-m+1}{m}, \ldots, \frac{-1}{m}, \frac{1}{m}, \ldots, \frac{m-1}{m}, \frac{m}{m} = 1.0$$

To evaluate these outcomes, we suggest a plot as shown in Figure 2. The x-value of a dot represents how often the y-value occurred as $cor_{j,k}$. To estimate the PUF uniqueness quality, a plot gained from empirical experiments is compared with the theoretically ideal plot that would be expected from completely uncorrelated bits. The ideal plot is generated as follows.

For completely uncorrelated signature bits, the outcome of $cor_{j,k}^i$ is like a coin toss, so we have $P(cor_{j,k}^i = 1) = P(cor_{j,k}^i = -1) = 0.5$. Hence, the probabilities for the outcome of each $cor_{j,k}$ value are:

<table>
<thead>
<tr>
<th>$cor_{j,k}$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1.0 = \frac{-m}{m}$</td>
<td>$\binom{0}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$\frac{-m+1}{m}$</td>
<td>$\binom{1}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$\frac{-m+2}{m}$</td>
<td>$\binom{2}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$\frac{-1}{m}$</td>
<td>$\binom{\frac{m-1}{m}}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$0$</td>
<td>$\binom{\frac{m}{m}}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$\frac{1}{m}$</td>
<td>$\binom{\frac{m+1}{m}}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$\frac{m-2}{m}$</td>
<td>$\binom{\frac{m-2}{m}}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$\frac{m-1}{m}$</td>
<td>$\binom{\frac{m-1}{m}}{m} \cdot 0.5^m$</td>
</tr>
<tr>
<td>$1.0 = \frac{m}{m}$</td>
<td>$\binom{m}{m} \cdot 0.5^m$</td>
</tr>
</tbody>
</table>

The fact that

$$\sum_{k=0}^{m} \binom{m}{k} \cdot 0.5^m = 2^m \cdot 0.5^m = 1.0$$

confirms this consideration.
4.2 Evaluation Capabilities of New Metric

When the occurrences of all empirically observed $\text{cor}_{j,k}$ outcomes are to be compared to the ideal distribution, the ideal probabilities can be multiplied by $\frac{n(n-1)}{2}$ (amount of $\text{cor}_{j,k}$ values) to get the expected amounts of the empirical outcomes. Alternatively, the observed outcomes can be divided by $\frac{n(n-1)}{2}$, which is preferable as it allows the comparison of PUFs with different $m$ and $n$.

Figure 2 shows the plot from an actual study done by the authors with $m = 72$ FPGA and a ring-oscillator (RO)-PUF producing $n = 3160$ signature bits. The red $\times$ plot stems from a flawed implementation in which some ROs suffered from systemic biases. As a result, several signature bits are the same on all devices. This poor bit-aliasing is indicated by the high occurrences of $\text{cor}_{j,k} = -1$ and $\text{cor}_{j,k} = 1$, because a pair of bits with bit-aliasing – that are either 0 or 1 on all devices – is always correlated. The blue + plot stems from an improved implementation overcoming the RO biases. The improvement is immediately visible in the plot.

Fig. 2. The evaluation of the new metric is done with a plot. Each dot represents how often (x-value) a $\text{cor}_{j,k}$ (y-value) was observed. For completely uncorrelated bits, the occurrences are expected to be in a bell curve around zero.

The theoretical ideal of completely uncorrelated signature bits (plot of yellow stars) is not achieved, because of another interesting phenomenon made visible by the new metric: There are more occurrences around $\text{cor}_{j,k} = \pm \frac{1}{3}$ than in the ideal case. The reason for this is that the signature bits of RO-PUFs are generated by comparing two ROs per bit. Some bits use the same RO as others, such that the frequency ranking of that RO has an influence on both bits, giving them a slightly increased overall correlation.

Finally, we revisit the example from the beginning of this section, where the signatures of $\frac{m}{2}$ devices have been the bitwise inverse of the other $\frac{m}{2}$ devices’ signatures. While $BA_j$ and equivalently $HD_{\text{metric}}$ do not register this undesirable
anomaly, the new metric produces a plot showing that there are only occurrences of $\text{cor}_{j,k} = -1$ and $\text{cor}_{j,k} = 1$, because each “0” bit is 100% positively correlated with every other “0” bit, and 100% negatively correlated with every “1” bit.

5 Conclusion

We showed, that the CSM is not only able to identify the same anomalies as the traditional metrics, but that it furthermore detects correlations between signature bits the traditional metrics cannot detect. However, $BA_j$ is still a valuable addition to CSM, because $BA_j$ allows to single out signature bits with poor bit-aliasing directly. The plot visualisation of CSM does not allow to immediately recognise which bits are correlated, it only shows if there are any. However, the generation algorithm can easily be setup to write down the indices $j$ and $k$ of all bits with suspicious $\text{cor}_{j,k}$ outcomes.

For future work, the authors intent to take into account further means of quantifying the uniqueness of PUF signatures, e.g. Shannon Entropy [11] or the effectiveness of data compression algorithms [15,3].

References


