Provable Performance Guarantees of Copy Detection Patterns

Joakim Tutt
Department of Computer Science
University of Geneva
Geneva, Switzerland
joakim.tutt@unige.ch

Slava Voloshynovskiy
Department of Computer Science
University of Geneva
Geneva, Switzerland
svolos@unige.ch

Abstract—**Copy Detection Patterns (CDPs) are crucial elements in modern security applications, playing a vital role in safeguarding industries such as food, pharmaceuticals, and cosmetics. Current performance evaluations of CDPs predominantly rely on empirical setups using simplistic metrics like Hamming distances or Pearson correlation. These methods are often inadequate due to their sensitivity to distortions, degradation, and their limitations to stationary statistics of printing and imaging. Additionally, machine learning-based approaches suffer from distribution biases and fail to generalize to unseen counterfeit samples. Given the critical importance of CDPs in preventing counterfeiting, including the counterfeit vaccines issue highlighted during the COVID-19 pandemic, there is an urgent need for provable performance guarantees across various criteria. This paper aims to establish a theoretical framework to derive optimal criteria for the analysis, optimization, and future development of CDP authentication technologies, ensuring their reliability and effectiveness in diverse security scenarios.**

Index Terms—**Copy detection patterns, Hamming distance, channel reliability, Neymann-Pearson test, channel aggregation, probability of error.**

I. Introduction

Copy Detection Patterns (CDPs) have emerged as indispensable tools in modern security frameworks, particularly in sectors where the integrity and authenticity of products are paramount [1]–[3]. Industries such as food, pharmaceuticals, and cosmetics heavily rely on CDPs to combat counterfeiting, ensuring consumer safety and maintaining product integrity. The urgency of this issue has been underscored by the proliferation of counterfeit vaccines during the COVID-19 pandemic, highlighting the critical need for reliable anti-counterfeiting measures. CDPs are essentially random binary patterns that possess high entropy, meaning the probability of black and white dots in these codes is equal to 0.5. These patterns are reproduced on physical media using high-resolution printing or laser engraving. Authentication of CDPs involves comparing the physical patterns to their corresponding digital templates. The fundamental idea is that any attempt to counterfeit these patterns will introduce additional distortions during scanning and reproducing processes, allowing the authentication system to distinguish between genuine and fake CDPs.

The challenge in CDP authentication is twofold. Firstly, attackers might use machine learning techniques to produce high-quality fakes. They can scan the printed CDPs from physical objects, preprocess them to estimate the original digital templates, and then reprint these estimates. These attacks, known as machine learning-based copy attacks, represent the most sophisticated and challenging threat, as described in [4]–[7]. Secondly, the imaging device used to acquire the patterns introduces additional degradation during the acquisition process, further complicating the authentication task. Another significant technical constraint is the need to authenticate CDPs based solely on the digital templates from which they are printed. Authentication based on physically scanned patterns is impractical due to the difficulty of managing the massive production volumes. Therefore, this study focuses on technologies that can authenticate printed CDPs using modern digital printing techniques and mobile phone acquisitions, relying only on the digital templates. Furthermore, the current methods for authenticating CDPs are often inadequate. Empirical setups typically utilize simplistic metrics like Hamming distances or Pearson correlation. While these metrics offer some insights, they fall short when accounting for the distortions and degradations that CDPs may encounter in real-world scenarios [8]. Machine learning methods, on the other hand, present their own set of challenges. Although they can provide sophisticated analysis, these methods are prone to distribution biases. When faced with new counterfeit samples that were not part of the training data, these models often fail to perform reliably [9]. This limitation is especially concerning given the dynamic and evolving nature of counterfeiting tactics.

In light of these challenges, there is a pressing need to develop a comprehensive analytical framework that provides provable guarantees on the performance of CDPs across a range of criteria. Such a framework would not only enhance the reliability and effectiveness of CDPs but also guide future developments in authentication technologies. This paper seeks to address this need by proposing a theoretical approach to derive optimal criteria for the analysis and optimization of CDP performance. Through this framework, we aim to ensure that CDP technologies can be systematically improved to meet the rigorous demands of modern security applications.

In previous publications [10], [11], we demonstrated that

S. Voloshynovskiy is a corresponding author.

This research was partially funded by the Swiss National Science Foundation SNF No. 200021_182063.

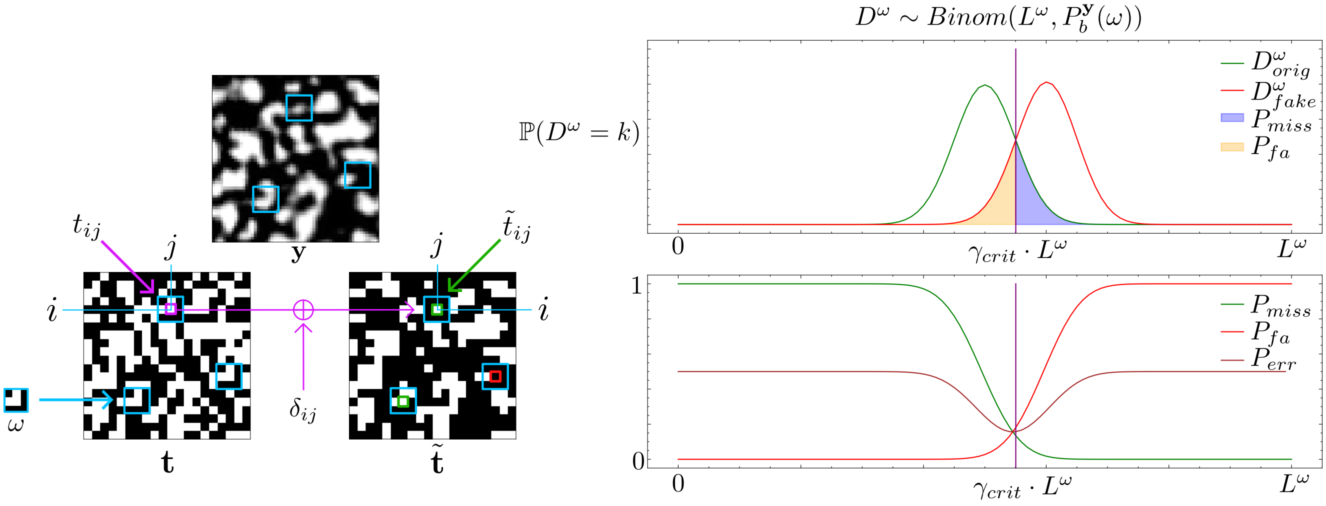

Fig. 1: An illustration of the theoretical framework under study. Given a fixed pattern ω and a CDP probe $\mathbf{y}$, one identifies all appearances of the pattern ω (left) in the digital template $\mathbf{t}$. For each location (i,j) of ω in $\mathbf{t}$, one can associate a random bit-flipping $\delta_{i,j}$. The Hamming distance D^ω between $\mathbf{t}$ and $\tilde{\mathbf{t}}$ for the pattern ω follows a Binomial distribution which is compared with the reference Binomial distributions of originals and fakes CDP. Based on the reference distributions, an optimal separation bound γ_{crit} is computed which minimizes the average probability of error.

different CDP patterns, composed of 3×3 elements, have varying probabilities of bit-flipping after printing, scanning or mobile phone image acquisition and binarization. We identified 512 different patterns and characterized their reliability based on the context of their 3×3 neighborhoods. However, those studies employed simple rules, using masking based on empirically optimized thresholds, and focused only on reliable channels characterized by low bit-flipping probabilities. Patterns with high bit-flipping probabilities were disregarded completely. Furthermore, simple metrics such as Hamming distance or Pearson correlation were used for the reliable patterns based CDP authentication.

In contrast, this new study explores more sophisticated rules that go beyond hard decision-making and the binary characterization of reliable and unreliable patterns. Instead, this paper introduces three different criteria for the measure of quality of the acquired CDP based on well-known statistical tests namely: The Hamming distance test; the cross-entropy between the probability of bit-flipping and a reference probability; and Neymann-Pearson likelihood ratio. We investigate different fusion rules and decision strategies for binarized CDP. Specifically, we consider two approaches: making a decision first and then aggregating the results versus aggregating everything first and then making a decision. This paper presents theoretical findings on how to optimize these strategies for enhanced CDP authentication.

II. Information flow in Copy Detection Patterns

In this section, we reformulate the classical authentication process of CDP using an information theoretic approach [12]. When trying to identify a probe sample $\mathbf{y}$, one typically uses binarization and compares the binarized probe $\tilde{\mathbf{t}}$ with the original template $\mathbf{t}$ using Hamming distance $d(\mathbf{t}, \tilde{\mathbf{t}})$. A decision threshold $\gamma \in [0,1]$ is then fixed to decide whether the probe $\mathbf{y}$ is genuine or counterfeited:

$$d(\mathbf{t}, \tilde{\mathbf{t}}) \gtrless \gamma B, \tag{1}$$

where B denotes the total number of symbols in $\mathbf{t}$. From the point of view of information theory and following the analysis done in [11], this process can be described as a Markov chain: $\mathbf{T} \rightarrow \mathbf{Y} \rightarrow \tilde{\mathbf{T}}$, where $\mathbf{T}$ is a random binary matrix in which each symbol has equal probability of being black or white, $\mathbf{Y}$ is a random matrix with real value entries between 0 and 1 and $\tilde{\mathbf{T}}$ is a binary matrix with fixed probability for each symbol. In this framework, each symbol transition can be described as a noisy binary channel:

$$\tilde{T}_{ij} = T_{ij} \oplus \delta_{ij}, \tag{2}$$

where $P_b(i,j) = \mathbb{P}(T_{ij} \neq \tilde{T}_{ij})$ is the probability of bit-flip for symbol T_{ij}, $\delta_{ij} = 1$ if there is a bit-flip and $\delta_{ij} = 0$, otherwise (see Fig. 1, left). In this setting, the focus is on being able to describe the posterior distribution $\mathbb{P}(\tilde{\mathbf{T}}|\mathbf{T})$.

The main contribution of [11] was to precisely describe the variability of the probability of bit-flipping $P_b(i,j)$ as a function of the pixel position (i,j) by introducing the Binary Pattern-based Channel (BPC) model. This model makes the assumption that the probability of bit-flipping of the symbol T_{ij} only depends on a small subset of symbols directly surrounding symbol T_{ij}, which we denote as ω_{ij}. Typically:

$$\omega_{ij} = \{T_{i \pm a, j \pm b} | 0 \leq a, b < h/2\}, \tag{3}$$

for an $h \times h$ square pattern centered around T_{ij}. Apart from this local dependency, the model assumes independence of the probability of bit-flipping with respect to the location of the pattern. That is, if $\omega_{ij} = \omega_{kl}$ are two identical patterns located in different coordinates, then:

$$P_b(i,j) = P_b(k,l). \tag{4}$$

Based on these assumptions, the characterization of the binary channels $T_{ij} \to \tilde{T}_{ij}$ reduces to the study of the binary channels $T_\omega \to \tilde{T}_\omega$, where T_ω indicates the central symbol of pattern ω. The advantage of this approach is the reduction of dimension, from the study of B different channels to $M = 2^{h^2}$ (in our experiments, $B = 51'984$, $h = 3$ and $M = 512$). According to this model, one can compute the Hamming distance for each pattern ω:

$$D^\omega = d^\omega(\mathbf{T}, \tilde{\mathbf{T}}) = \sum_{\substack{i,j:\\ \omega_{ij}=\omega}} \delta_{ij}, \tag{5}$$

and compute a pattern-wise probability of bit-flipping:

$$P_b^{\mathbf{y}}(\omega) = \frac{1}{L^\omega}\mathbb{E}[D^\omega], \tag{6}$$

where L^ω denotes the number of pattern ω appearing in $\mathbf{T}$ and $P_b^{\mathbf{y}}(\omega)$ is the probability of bit-flipping for pattern ω. As such, the model transforms a digital probe $\mathbf{y}$ into a feature vector of size M: $\{P_b^{\mathbf{y}}(\omega)\}_{\omega=0}^{M-1}$.[1]

III. A STATISTICAL VIEWPOINT

In this paper, we use the BPC model [11] as a baseline and describe the problem of authentication of CDP as a collection of independent channel-wise statistical tests. Under this model, each pattern ω gives rise to a random binomial variable D^ω independent of each other. Based on this mathematical description, we focus on different ways to measure channel performances and study how they can be optimally combined together to strengthen authentication. That is, given a probe $\mathbf{y}$, decide whether it is an original $\mathbf{x}$ or a fake $\mathbf{f}$. We denote $\mathbb{H}_0$ the hypothesis that probe $\mathbf{y}$ is an original and $\mathbb{H}_1$ the hypothesis that $\mathbf{y}$ is a fake. Under $\mathbb{H}_0$, the reference original codebook is denoted as $\mathcal{C}_0 = \{P_b(\omega)\}_{\omega=0}^{M-1}$ while, under $\mathbb{H}_1$, the reference fake codebook is denoted as $\mathcal{C}_1 = \{Q_b(\omega)\}_{\omega=0}^{M-1}$. Given an unidentified probe $\mathbf{y}$, and a fixed pattern ω, we perform the following steps (see Fig. 1 for an illustration):

- Identify each appearance of pattern ω in the reference template $\mathbf{t}$ and extract the central pixel of each of these patterns: $t_1, ..., t_{L^\omega}$;
- Binarize the probe $\mathbf{y} \mapsto \tilde{\mathbf{t}}$ and extract the corresponding central pixels: $\tilde{t}_1, ..., \tilde{t}_{L^\omega}$;
- Compute the bit-flipping values $\delta_k = t_k \oplus \tilde{t}_k$ for each $k = 1, ..., L^\omega$, where δ_k denotes the indicator function of bit-flipping with $\delta_k = 1$, if $t_k \neq \tilde{t}_k$, and 0, otherwise.

The BPC model defines that the bit-flipping values $\delta_1, ..., \delta_{L^\omega}$ are sampled according to a Bernoulli distribution with a fixed probability $P_b^{\mathbf{y}}(\omega)$ which can be estimated empirically as:

$$P_b^{\mathbf{y}}(\omega) \approx \frac{1}{L^\omega}\sum_{k=1}^{L^\omega} \delta_k. \tag{7}$$

The goal of authentication is now to decide whether this empirical probability is more likely to be $P_b(\omega)$ or $Q_b(\omega)$ and provide bounds on this decision [13].

A. Hamming distance test

In the following sections, all analyses are conducted on a fixed pattern ω and then extended to all ω. As such, and to ease the notations, we remove the explicit reference to the pattern in the indices, except when formulating final results:

$$D^\omega \to D, \quad L^\omega \to L, \quad P_b(\omega) \to P_b, \quad Q_b(\omega) \to Q_b.$$

Following the analysis in [12], the Hamming distance is described by a binomial distribution:

$$D = \sum_{k=1}^{L} \delta_k \sim Binom(L; P_b^{\mathbf{y}}). \tag{8}$$

If we suppose that we have access to reference codebooks for original and fakes, we can formulate the Hamming distance test and measure the probability of miss as:

$$\begin{aligned} P_m(\gamma) &= \mathbb{P}_{\mathbb{H}_0}(D > \gamma L) \\ &= \sum_{k=\lfloor \gamma L \rfloor + 1}^{L} \binom{L}{k} P_b^k (1-P_b)^{L-k}, \end{aligned} \tag{9}$$

and the probability of false acceptance as:

$$P_{fa}(\gamma) = \mathbb{P}_{\mathbb{H}_1}(D \leq \gamma L) = \sum_{k=0}^{\lfloor \gamma L \rfloor} \binom{L}{k} Q_b^k (1-Q_b)^{L-k}. \tag{10}$$

One way to choose an optimal decision threshold γ is to minimize the average probability of error as:

$$\min_{\gamma \in [0,1]} P_{err}(\gamma) = \min_{\gamma \in [0,1]} \frac{P_m(\gamma) + P_{fa}(\gamma)}{2}, \tag{11}$$

assuming both hypotheses $\mathbb{H}_0$ and $\mathbb{H}_1$ are equiprobable (see Fig. 1, right). Our goal, as defined in the following sections, is to better study these probability of errors and derive an optimal bound for each channel by making an explicit link between this approach and the Neymann-Pearson test.

B. Posterior likelihood approach

The authentication problem can be studied from the point of view of the likelihood function. Given a sequence of realisations $\delta_1, ..., \delta_L \in \{0, 1\}$ and a reference probability of bit-flipping P_b, one can measure the likelihood that these samples were drawn from a Bernoulli distribution with parameter P_b. We write:

$$\begin{aligned} \mathbb{P}(\tilde{t}_1, ..., \tilde{t}_L | t_1, ..., t_L) &= \prod_{k=1}^{L} \mathbb{P}(\tilde{t}_k | t_k) \\ &= \prod_{k=1}^{L} P_b^{\delta_k} \cdot (1-P_b)^{1-\delta_k} \\ &= P_b^D \cdot (1-P_b)^{L-D}, \end{aligned} \tag{12}$$

where D is the Hamming distance defined in (8) and follows a Binomial law. Note that this derivation is based on the

[1] Profitting from the fact that, on average, $L^\omega = B/M \approx 100$, one can estimate the probability of bit-flipping empirically for each of the patterns solely based on one pair $(\mathbf{t}, \tilde{\mathbf{t}})$ and build a feature vector for a given probe $\mathbf{y}$.

assumption of the BPC model that each bit-flipping δ_k is identically distributed and independent. Taking the logarithm on both sides, one arrives to the formula:

$$\log \mathbb{P}(\tilde{t}_1, ..., \tilde{t}_L | t_1, ..., t_L) = D \cdot \log P_b + (L - D) \cdot \log\left(1 - P_b\right). \tag{13}$$

These types of derivations are well known [13] and explicit the link between minimization of Hamming distance and maximization of log-likelihood. Taking the expectation with respect to the posterior distribution on both sides yields the following result:

Proposition III.1. *The average log-likelihood is given by:*

$$\frac{1}{L^\omega}\mathbb{E}[\log \mathbb{P}(\tilde{t}_1, ..., \tilde{t}_{L^\omega} | t_1, ..., t_{L^\omega})] = -\mathcal{H}(P_b^{\mathbf{y}}(\omega); P_b(\omega)), \tag{14}$$

where the right-hand side designates the cross-entropy between the unknown probe distribution $P_b^{\mathbf{y}}(\omega)$ and the reference distribution $P_b(\omega)$.

Proof. The proof is straightforward from (13) since $D \sim Binom(L, P_b^{\mathbf{y}})$, and thus $\frac{1}{L}\mathbb{E}[D] = P_b^{\mathbf{y}}$. ■

Formula (14) is very useful to test the adequacy of probe $\mathbf{y}$ to a certain reference probability P_b. In practice one can use (7) to estimate these quantities empirically. This result also gives new insights on the Posterior Log Likelihood (PLL) metric introduced in [10].

Corollary 1. *The PLL metric can be computed as an average on all channels:*

$$PLL(\mathbf{y}; \mathcal{C}) = -\sum_{\omega=0}^{M-1} L^\omega \cdot \mathcal{H}(P_b^{\mathbf{y}}(\omega); P_b(\omega)). \tag{15}$$

C. Neymann-Pearson optimal bound

Using Neymann-Pearson terminology, we will designate the type I error in this problem, as accepting the probe as an original when it is a fake (false acceptance). The type II error would be to reject a probe when it is an original (true miss). The problem is, by nature, asymmetrical as it is much more harmful in real life scenarios to accept a fake product than to throw away a genuine one (think for instance of fake drugs). Neymann-Pearson lemma [14], then provides an optimal decision criterion for a significance test at level α. That is, if we fix the Type I error to be equal to α, the optimal decision statistic is given by the likelihood ratio test:

$$\mathcal{R} = \frac{\mathbb{P}_{\mathbb{H}_0}(\tilde{t}_1, ..., \tilde{t}_L | t_1, ..., t_L)}{\mathbb{P}_{\mathbb{H}_1}(\tilde{t}_1, ..., \tilde{t}_L | t_1, ..., t_L)} > \rho. \tag{16}$$

Proposition III.2. *The likelihood ratio test* (16) *is equivalent to:*

$$\mathcal{H}(P_b^{\mathbf{y}}(\omega); Q_b(\omega)) - \mathcal{H}(P_b^{\mathbf{y}}(\omega); P_b(\omega)) > \frac{\log \rho}{L^\omega}, \tag{17}$$

and this test is asymptotically optimal (with $L^\omega \to +\infty$) as long as $Q_b(\omega) \neq P_b(\omega)$.

Proof. By computing $\frac{1}{L}\log \mathcal{R}$ and following the same derivation as (12), one arrives to the following formula:

$$\frac{1}{L}\log \mathcal{R} = \frac{D}{L}\log\frac{P_b}{Q_b} + \frac{L-D}{L}\log\frac{1-P_b}{1-Q_b}. \tag{18}$$

Equation (17) then follows by taking the expectation on both sides as in (14). Under hypothesis $\mathbb{H}_0$, the random variable $D \sim Binom(L; P_b)$ follows a binomial distribution and thus:

$$\begin{aligned}\frac{1}{L}\mathbb{E}[\log \mathcal{R} | \mathbb{H}_0] &= P_b \cdot \log\frac{P_b}{Q_b} + (1 - P_b) \cdot \log\frac{1-P_b}{1-Q_b} \\ &= \mathcal{D}_{KL}(P_b || Q_b),\end{aligned} \tag{19}$$

where $\mathcal{D}_{KL}(P_b || Q_b)$ denotes the Kullback-Leibler divergence between the Bernoulli distributions given by P_b and Q_b. With similar computations one can show that:

$$\frac{1}{L}\mathbb{E}[\log \mathcal{R} | \mathbb{H}_1] = -\mathcal{D}_{KL}(Q_b || P_b). \tag{20}$$

Finally, with similar approach, one shows that the variance $\mathrm{Var}(\frac{\log \mathcal{R}}{L}) \to 0$, when $L \to +\infty$ and so these quantities concentrate around their means, yielding asymptotic optimality, independant of ρ. ■

Note that (17) can be seen as a direct extension of the approach in [11]. Indeed, the PLL metric was solely based on a single cross-entropy term, whereas the Neymann-Pearson test is based on the difference of two cross-entropies. PLL can thus be considered as an analogy to a one-class decision based solely on P_b, while NP-test is a two-class decision which requires the knowledge of both P_b and Q_b.

Proposition III.3. *The Neymann-Pearson test and the Hamming distance test $\mathcal{R} > \rho$ and $D < \gamma L$ are equivalent if $P_b \leq Q_b$ and:*

$$\frac{\log \rho}{L} = \mathcal{D}_{KL}(\gamma || Q_b) - \mathcal{D}_{KL}(\gamma || P_b). \tag{21}$$

When $\rho = 1$, that is when both likelihoods are equal, we can derive the critical separation boundary independent of L:

$$\gamma_{crit} = \left(1 + \frac{\log P_b - \log Q_b}{\log(1 - Q_b) - \log(1 - P_b)}\right)^{-1}. \tag{22}$$

Proof. To prove (21), we start with relation (18) and use the facts that $D < \gamma L$, $\log(\frac{P_b}{Q_b}) \leq 0$ and $\log(\frac{1-P_b}{1-Q_b}) \geq 0$ to derive the lowerbound (21). Equation (22) is obtained from (21) by setting $\rho = 0$ and isolating γ. ■

IV. Channels aggregation strategies

All statistical approaches described in Section III, were computed independently for each channel and different approaches to optimal thresholds selection were given. Applying a decision threshold to each channel transforms the real-valued feature vector $\{P_b^{\mathbf{y}}(\omega)\}_{\omega=0}^{M-1}$ into a binary decision vector: $\{\Delta_\omega\}_{\omega=0}^{M-1}$, where $\Delta_\omega \in \{0, 1\}$. In both situations, we are left with the problem of optimally aggregating these pattern-wise scores into a final decision score S_{final}. We propose here a generic way of describing linear aggregation strategies. A linear aggregation strategy is a weighted sum:

$$S^{AD}_{final} = \sum_{\omega} P^{\mathbf{y}}_b(\omega) \cdot \alpha_\omega \quad \text{or,} \quad S^{DA}_{final} = \sum_{\omega} \Delta_\omega \cdot \alpha_\omega, \quad (23)$$

where the coefficients $\alpha_\omega \in \mathbb{R}$ are chosen according to an aggregation strategy. A final decision based on this aggregation is performed by applying a threshold to S_{final}. These two approaches will later be referred to as Aggregate first - Decide next (AD) for the left one and Decide first - Aggregate next (DA) for the one on the right. In [10], [11], the (AD) aggregation was performed according to two simple criterions:

(S1) *Averaging*: Taking the average of all statistics with equal contribution: $\alpha_\omega = 1$.

(S2) *Reliable patterns* [10]: Selecting only those patterns for which $P_b(\omega) \leq \mu$ for a fixed threshold μ and taking the average of all statistics on these patterns:

$$\alpha_\omega = \begin{cases} 1, & \text{if } P_b(\omega) \leq \mu, \\ 0, & \text{else.} \end{cases}$$

These approaches, although showing promising results, were chosen empirically and lacked a deeper investigation which could lead to a better understanding of the decision process. In this paper, we now introduce a systematic approach, based on the average probability of error:

(S3) *Minimal Probability of error*: Fixing a threshold $\nu \in [0, 1]$ and only aggregating the patterns which have a probability of classification error smaller than ν:

$$\alpha_\omega = \begin{cases} 1, & \text{if } P_{err}(\omega) \leq \nu, \\ 0, & \text{else.} \end{cases}$$

(S4) *2C-SVM optimal*: Training a 2-class linear SVM [15] on the statistics and extract the coefficients learned by the model as "optimal" machine learning coefficients α_ω. We then sort them according to their absolute value and aggregate the k firsts: $|\alpha_{\omega_1}| \geq |\alpha_{\omega_2}| \geq |\alpha_{\omega_3}| \geq \dots$.

V. Experimental results

A. 1x1 Indigo base smartphone dataset

The dataset[2] of CDP that we use in the following experiment is fully presented in [11]. It consists of 1440 unique triples of digital template, original printed and fake CDP $(\mathbf{t}, \mathbf{x}, \mathbf{f})$ enrolled with an *iPhone 12 Pro* during 6 capture runs. The CDP acquired in this way are then processed using histogram matching to a reference CDP and binarized using Otsu's method. In the following experiments, the dataset is divided into two randomly chosen train set and test set, which consist of 500 triples. The train set is used to compute reference codebooks of originals $\{P_b(\omega)\}_{\omega=0}^{M-1}$ and fakes $\{Q_b(\omega)\}_{\omega=0}^{M-1}$.

B. Reliability and authentication performance

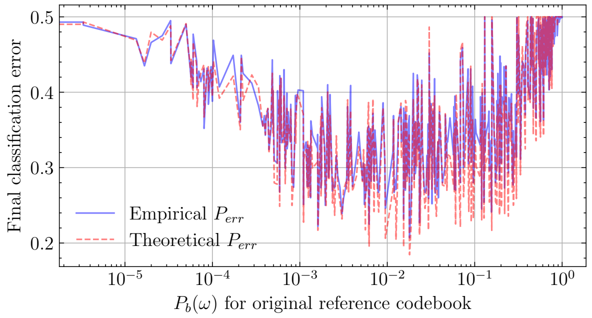

Fig. 2: Average probability of error between originals and fakes samples as a function of the probability of bit-flipping of each channel. The optimal probability of error is computed using the theoretical model (in red) and empirically on the test set (in blue). Both curves show a close match.

The results of this experiment are shown in Fig. 2. For each pattern ω independently, we measure the minimal average classification error P_{err} both theoretical (red dashed curve) by referring to the optimal threshold selection γ_{crit} derived in (22) for $L^\omega = 100$, and experimental (blue curve) by performing an optimal threshold selection on the test set. The results show that the theoretical predictions of the average probability of error based on the BPC model closely match the empirical measure based on the testing set. Indeed, the red and blue curve are very close together. Finally, we also note that pattern reliability, that is low probability of bit-flipping, does not necessarily imply low classification error[3]. Indeed, the plot clearly indicates that the best performing patterns are those located between $P_b(\omega) = 10^{-3}$ and $P_b(\omega) = 2 \cdot 10^{-2}$. This observation is a clear motivation for us to investigate strategy (S3) based on a bound ν on P_{err} for the aggregation strategy and compare it with strategy (S2).

C. Aggregation strategies

In this second experiment, we test the various aggregations strategies introduced in Section IV. Fig. 3 shows the final average error of classification as a function of the number k of patterns used for authentication. The selection of the k best patterns depends on the strategy used for aggregation and is fully described in Section IV.

Surprisingly, even though the strategy (S3) based on the minimal P_{err} is optimal for each pattern taken individually, aggregating according to this strategy does not improve the final decision score on this dataset. Indeed, the aggregation strategy (S2) based on the minimization of $P_b(\omega)$ introduced in [11] still outperforms this theoretically optimal criterion, while still being solely based on a reference codebook of originals. Finally, the fully supervised aggregation scheme (S4) based on linear 2-class SVM performs optimally and, contrary to the other aggregation schemes, does not worsen when the number of patterns used gets too big. This approach, although theoretically interesting, requires the knowledge of

[2] The dataset is publicly available and can be found here: https://sipcloud.unige.ch/index.php/s/tYKffnKNRgSwBAN

[3] The idea being that an all-white or all-black pattern, while being very reliable ($P_b(\omega)$ is almost 0) is completely useless for classification since fakes will not be wrong when estimating these either and so $Q_b(\omega) = P_b(\omega)$ for those patterns making them indistinguishable from a theoretical point of view.

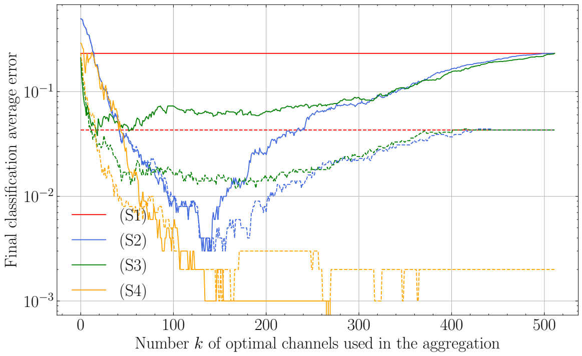

Fig. 3: A plot of the final classification error probability based on the final score S_{final} for each aggregation strategy as a function of the number of channels aggregated. Continuous curves correspond to (AD) and dashed curves correspond to (DA). The best k channels are different for each strategy and depend on the optimality criterion.

both original CDP and fake CDP to be applied and we conjecture that it should be prone to errors against other types of unseen fakes [9].

D. Final classification performance

In this third experiment, we compare the final aggregated average classification error for the different aggregating strategies introduced in Section IV. A direct comparison is done between single-shot aggregated performance and multi-shot aggregated performance, where all 6 capture runs are used and fused before computing the Hamming distance test, effectively increasing the number L^{ω} of symbols in each CDP by a factor of 6. (The behaviour of these symbols is of course strongly correlated but still enhances the prediction capacity.)

When looking at Table I, a striking observation is the outstanding enhancement of performance when using multi-shot prediction. Indeed, all aggregation strategies benefit from this approach. We analyze this improvement as an illustration of Prop. III.2. Without much surprise, the fully supervised linear 2C-SVM aggregation strategy (S4) performs the best. It is still remarkable to notice that such an approach is able to perfectly separate originals and fakes when all other approaches still miss by a few. Finally, the (DA) approach also shows an interesting enhancement in performance compared to the (AD) approach which was used in the previous papers.

VI. Conclusion

In this article, we presented a theoretical analysis of the Binary Pattern-based Channel Model [11]. We derived optimal bounds for each channel for the classification task and introduced a structured description for the aggregation strategies of the different channels. By comparing with a real dataset of smartphone acquired CDP, we demonstrated that the BPC model correctly simulates the authentication process and allows to reason about the design of optimal aggregation strategies. In future works, we plan to pursue the analysis of the interplay between the theoretical BPC model and the real life CDP datasets. Extending this research to new types of fakes and continue to explore the design of optimal aggregation strategies.

TABLE I: Results of minimal average P_{err} (in percent) for the aggregation of channels based on four different strategies with Aggregate-Decide (AD) or Decide-Aggregate (DA). Both single-shot and multi-shot settings are compared.

		Aggregation Strategy			
		(S1)	(S2)	(S3)	(S4)
AD	single-shot	20.29	0.36	3.07	0.10
	multi-shot	14.30	0.04	0.13	0.00
DA	single-shot	4.12	0.40	0.96	0.11
	multi-shot	1.20	0.05	0.06	0.00

References

[1] J. Picard, "Digital authentication with copy-detection patterns," in *Optical Security and Counterfeit Deterrence Techniques V*, vol. 5310. International Society for Optics and Photonics, 2004, pp. 176–183.

[2] J. Picard, P. Landry, and M. Bolay, "Counterfeit detection with qr codes," in *Proceedings of the 21st ACM Symposium on Document Engineering*, 2021, pp. 1–4.

[3] I. Tkachenko, A. Trémeau, and T. Fournel, "Protection de documents par impression d'éléments anticopies," in *Sécurité multimédia 2 : Biométrie, protection et chiffrement multimédia*, Jul. 2021.

[4] R. Chaban, O. Taran, J. Tutt, T. Holotyak, S. Bonev, and S. Voloshynovskiy, "Machine learning attack on copy detection patterns: are 1x1 patterns cloneable?" in *IEEE International Workshop on Information Forensics and Security (WIFS)*, December 2021.

[5] R. Yadav, I. Tkachenko, A. Trémeau, and T. Fournel, "Estimation of copy-sensitive codes using a neural approach," in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 2019.

[6] E. Khermaza, I. Tkachenko, and J. Picard, "Can copy detection patterns be copied? evaluating the performance of attacks and highlighting the role of the detector," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2021, pp. 1–6.

[7] O. Taran, S. Bonev, and S. Voloshynovskiy, "Clonability of anti-counterfeiting printable graphical codes: a machine learning approach," in *IEEE ICASSP*, 2019, pp. 2482–2486.

[8] R. Chaban, O. Taran, J. Tutt, Y. Belousov, B. Pulfer, T. Holotyak, and S. Voloshynovskiy, "Printing variability of copy detection patterns," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, Shanghai, China, December 2022.

[9] B. Pulfer, R. Chaban, Y. Belousov, J. Tutt, O. Taran, T. Holotyak, and S. Voloshynovskiy, "Authentication of copy detection patterns under machine learning attacks: A supervised approach," October 2022.

[10] J. Tutt, O. Taran, R. Chaban, B. Pulfer, Y. Belousov, T. Holotyak, and S. Voloshynovskiy, "Mathematical model of printing-imaging channel for blind detection of fake copy detection patterns," in *IEEE WIFS*, Shanghai, China, December 2022.

[11] ——, "Authentication of copy detection patterns: A pattern reliability based approach," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 3124–3134, 2024.

[12] F. Beekhof, S. Voloshynovskiy, and F. Farhadzadeh, "Content authentication and identification under informed attacks," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5 2012.

[13] T. Cover and J. Thomas, *Information theory and statistics*. Wiley, 2012, ch. 11, pp. 347–408.

[14] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Phil. Trans. Roy. Soc. Lond. A*, vol. 231, no. 694-706, pp. 289–337, 1933.

[15] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, may 2011.