# Content Authentication and Identification under Informed Attacks

Fokko Beekhof [1], Sviatoslav Voloshynovskiy [2], Farzad Farhadzadeh [3]

*CUI, University of Geneva*
*7, route de Drize*
*1227 Carouge*
*Switzerland*
[[1]Fokko.Beekhof, [2]svolos, [3]Farzad.Farhadzadeh]@unige.ch

***Abstract*—We consider the problem of content identification and authentication based on digital content fingerprinting. Contrary to existing work in which the performance of these systems under blind attacks is analysed, we investigate the information-theoretic performance under informed attacks. In the case of binary content fingerprinting, in a blind attack, a probe is produced at random independently from the fingerprints of the original contents. Contrarily, informed attacks assume that the attacker might have some information about the original content and is thus able to produce a counterfeit probe that is related to an authentic fingerprint corresponding to an original item, thus leading to an increased probability of false acceptance.**

**We demonstrate the impact of the ability of an attacker to create counterfeit items whose fingerprints are related to fingerprints of authentic items, and consider the influence of the length of the fingerprint on the performance of finite-length systems. Finally, the information-theoretic achieveble rate of content identification systems sustaining informed attacks is derived under asymptotic assumptions about the fingerprint length.**

## I. Introduction

In recent years, digital content fingerprinting is becoming a popular technique for the identification and authentication of digital content, physical objects and people. Digital fingerprints are extracted directly from any digital data including images, video, audio, text etc. [1], [2], and represent a short, robust and distinctive representation of the content. In the case of physical objects, the fingerprints are extracted from some unique unclonable features of the material in the object that can include, but are not limited to, optical images of micro-structures, magnetic taggants or fibers [3]. The content fingerprints to authenticate or identify people are calculated from biometrics such as irises, fingerprints, or the geometry of the face or hand [4].

Given that special measures are taken to protect the extracted fingerprints, security and protection systems can be build to distinguish between authentic and counterfeit objects by offering authentication or identification services. These systems must meet a number of requirements, some of which conflict and contradict each other, such as high accuracy of authentication and identification, low storage requirements and the protection of sensitive information. The attacks are assumed to target primarily the acceptance of counterfeit items by the system, as well to gather sensitive information to accomplish their primary target, hence the requirement to minimize the available information to unauthorized parties. In addition to classical requirements, the ability to withstand attacks, or *resilience*, is a crucial aspect of protection systems.

Despite the increasing number of publications on security and protection of content fingerprints [2], [3], [5] it is difficult or sometimes impossible to completely reduce the information leakages about the stored authentic fingerprints. Obviously, informed attackers may benefit from these leakages by developing more sophisticated attacks against the different elements of both identification and authentication systems.

The impact of the leakage on the performance of the system is generally measured by the achievable trade-off between the probability of miss $p_m$ for authentication, or the probability of incorrect identification $p_{ii}$ in identification, versus the probability of false acceptance $p_f$ in both scenarios. In an information-theoretic sense, the information leakage is measured directly between the data $\mathbf{X}(m)$ and its public version $\mathbf{B}_{\mathbf{X}}(m)$ in terms of mutual information $I(\mathbf{X}(m); \mathbf{B}_{\mathbf{X}}(m))$. It is important to point out that one of the goals of resilient system design consists in the minimization of the information leakage $I(\mathbf{X}(m); \mathbf{B}_{\mathbf{X}}(m))$, in order to reduce the information available to the attacker, as well as his general ability to attack the system. Nevertheless, such a theoretic approach does not directly indicate the impact of information disclosure on the system's performance. Therefore, most practical systems measure the success of attacks in terms of the Successful Attack Rate (SAR) $p'_f$, which is the probability of falsely accepting an item when the adversary is at an advantage due to some knowledge about the original $\mathbf{X}(m)$ [6]. In this case, the probability of false acceptance represents a blind attack, i.e., the attacker probes the system with a random-generated $\mathbf{X}'$, whereas the SAR corresponds to informed attacks. In the latter case, the adversary attempts to present a fake $\mathbf{X}'(m)$ that is statistically related to $\mathbf{X}(m)$. It should be pointed out that there are many publications on the empirical design of various spoofing attacks against specific elements of security

*WIFS'2012, December, 2-5, 2012, Tenerife, Spain.*
*The contact author is Prof. Voloshynovskiy (svolos@unige.ch).*
*978-1-4244-9080-6/10/$26.00* 

systems where a certain degree of success is reported. A few examples are SIFT spoofing [7], and human fingerprint [8] and iris spoofing attacks [9] in biometric applications.

Finally, it is very important to mention that the information-theoretic analyses are performed under specific assumptions, in particular the use of infinitely long sequences or fingerprints. Contrarily, practical systems operate with finite-length fingerprints, which creates a crucial gap between theory and practice, as in channel coding in digital communication applications. As a result, little is understood about several open issues, namely:

1) how information-theoretic results can be mapped to practical protection systems;
2) the impact of the information leakage on the design of new spoofing attacks and their influence on a) the achievable identification rate $R$; b) the practical performance in terms of the $p_m$ and $p'_f$;
3) the impact of the fingerprint length $L$ on the leakage and the performance under informed attacks;
4) the optimal design of the detector or decoder when facing informed attacks.

There is a growing amount of publications on various spoofing attacks and despite their variety, all strive to produce a fake probe that should be accepted by the system, i.e., to maximize $p'_f$ using all available information. Instead of following the particular design behind all of these attacks, we will focus on binary content fingerprinting systems and assume without loss of generality that:

1) a *blind* attack consists in the generation of a probe randomly without taking any information about a particular $\mathbf{X}$ or $\mathbf{B_X}$ into account. This can be expressed as a probability of bit-error $p'_b = \frac{1}{2}$ in the binary domain;
2) an *informed* attack consists in the generation of a probe in an informed way such that the probability of bit-error between the original fingerprint and the fake fingerprint is $p'_b < \frac{1}{2}$.

The probability of bit-error for authentic items is denoted as $p_b$, and we assume that:

$$0 \leq p_b < p'_b \leq \frac{1}{2}. \tag{1}$$

The formulation of the problem in these terms allows one to analyze the effects of a partially successful attempt to challenge the system, without having to delve into the actual details of the particular approach that was taken to accomplish a lower bit-error rate. This generalization greatly simplifies the analysis of protection systems by making it sufficient to consider a bit-error rate that the system should be resilient to in order to defeat all attacks that can achieve upto that bit-error rate. The performance of authentication and identification systems will be analyzed along these assumptions.

**Notations.** Capital letters denote scalar random variables $X$, bold capital letters denote vector random variables $\mathbf{X}$, corresponding small letters $x$ and small bold letters $\mathbf{x}$ denote realizations of scalar and vector random variables, respectively. $\mathbf{B_X}$ and $\mathbf{b_x}$ are used to denote binary versions of $\mathbf{X}$ and $\mathbf{x}$. We use $X \sim B(L,p)$ to indicate that a random variable $X$ follows a Binomial distribution with parameters $L$ and $p$. $\mathbf{X}(m)$ denotes a vector $\mathbf{X}$ associated with a label $m$. Calligraphic letters $\mathcal{M}$ denote sets and $|\mathcal{M}|$ denotes the cardinality of the set. The Kullback-Leibler divergence between two Bernouilli distributions, is defined as:

$$D\left(\gamma \| p\right) = \gamma \log_2 \frac{\gamma}{p} + (1-\gamma)\log_2 \frac{1-\gamma}{1-p}. \tag{2}$$

The binary entropy is defined as $H_2(p) = 1 - D\left(p\|\frac{1}{2}\right)$.

## II. Mathematical Models of Binary Content Identification and Authentication

### A. Authentication

The authentication of items is a problem wherein it must be decided whether a certain item truly is the item it is claimed to be. The authentication problem consists of an *enrollment* and a *verification* stage. A diagram of authentication systems using digital fingerprints of some physical objects is shown in Fig. 1. In this analysis, we exemplify the considered setup

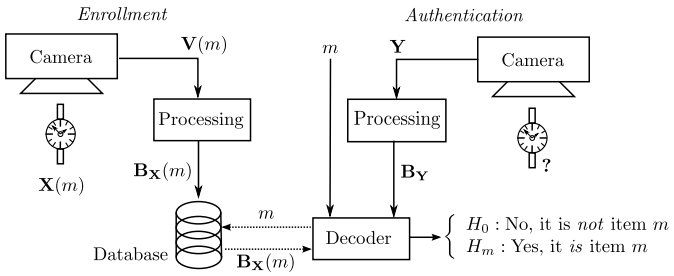



Fig. 1: Item authentication based on fingerprints.

based on physical items, but the same analysis can be applied to biometrics and digital content without loss of generality.

At the enrollment stage, a fingerprint database is created by processing an observation $\mathbf{X}(m) \in \mathbb{R}^N$ of an item with label $m \in \mathcal{M}$ that is assumed to be noise-free. The labels associated to all enrolled items are denoted by the set $\mathcal{M}$. The fingerprinting procedure is assumed to encompass a stage reducing the dimensionality from $N$ to $L$, followed by a quantization that transforms the continuous data into binary, resulting in a fingerprint $\mathbf{B_X}(m) \in \{-1,1\}^L$ that is stored in the fingerprint database. In this setup, we do not assume any specific countermeasures used to protect the template $\mathbf{B_X}(m)$. We refer interested readers to [2], [5], [6] for analyses of various protection techniques.

At the verification stage, either the authentic $\mathbf{X}(m)$ or a counterfeit $\mathbf{X}'(m)$ is presented together with a claimed label $m$. The noisy acquired data $\mathbf{Y} \in \mathbb{R}^N$ is transformed into a binary fingerprint $\mathbf{B_Y} \in \{-1,1\}^L$. Both $\mathbf{B_Y}$ and $\mathbf{B_X}(m)$ are passed to the detector. Based on $\mathbf{B_Y}$ and $\mathbf{B_X}(m)$, the detector produces a decision to the corresponding binary hypothesis test. This test can be reformulated as a function of the Hamming distance $d^H(\mathbf{B_Y}, \mathbf{B_X}(m))$:

$$\begin{cases} H_0: & d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) > \gamma L, \\ H_m: & d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) \leq \gamma L, \end{cases} \tag{3}$$

where $\gamma L$ denotes a threshold controlling the performance of the authentication system, such that $p_b < \gamma < p'_b$.

The distribution of Hamming distances under hypothesis $H_m$ follows a Binomial pmf, i.e., $B(L, p_b)$. The distribution of Hamming distances under hypothesis $H_0$ corresponds to $B(L, \frac{1}{2})$ in the case of a blind attack and to $B(L, p'_b)$ in case of an informed attack. According to Fig. 2, one can envision

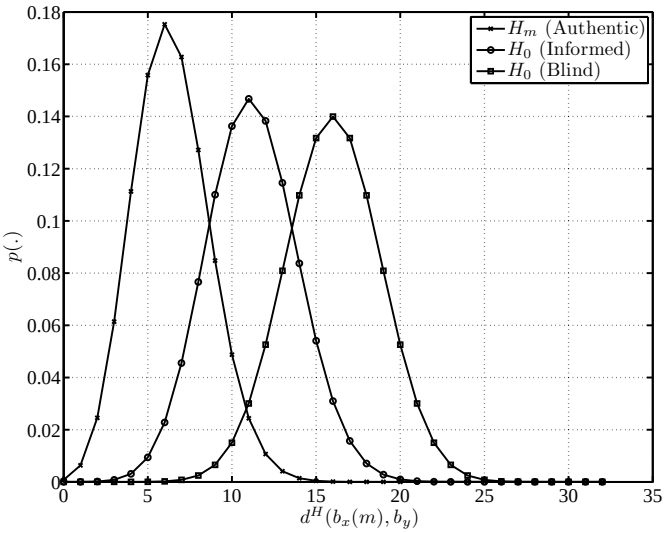



Fig. 2: The distribution of Hamming distances under hypotheses $H_m$, $H_0$ (blind), and $H_0$ (informed) in the authentication setup for $L = 32$, $p_b = 0.2$ and $p'_b = 0.35$.

the increase of the SAR $p'_f$ due to a greater overlap of the distributions when an informed adversary produces a fake $\mathbf{X}'$ such that in $p'_b < \frac{1}{2}$.

### B. Identification

As in the authentication problem, identification consists of an *enrollment* and a *verification* stage.

The enrollment stage is the same as for the authentication setup. At the verification stage, an item is presented, but, unlike in the authentication setup, no claim is made about its identity, i.e., the relevant label $m$. The index $m$ of the element in the database corresponding to the probe must be determined, or the item must be rejected. The output $\hat{m}$ is either $\emptyset$ for rejection (hypothesis $H_0$), or any valid $m \in \mathcal{M}$ (hypothesis $H_m$).

A diagram of systems using digital content fingerprints to address the identification problem is shown in Fig. 3. Based on $\mathbf{B_Y}$ and $\mathbf{B_X}(m)$, the decoder produces a decision corresponding to a multiple hypotheses test. In practical terms, the decision can be made by the following rule, for $p_b < \gamma < p'_b$:

$$\begin{cases} H_0 : \forall m \in \mathcal{M} : d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) > \gamma L, \\ H_1 : d^H(\mathbf{B_Y}, \mathbf{B_X}(1)) \leq \gamma L, \\ \quad \cdots \\ H_m : d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) \leq \gamma L. \end{cases} \tag{4}$$

The event that there is more than one $m$ for which $d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) \leq \gamma L$ is considered as an error. List decoding, i.e., accepting multiple candidates, is considered

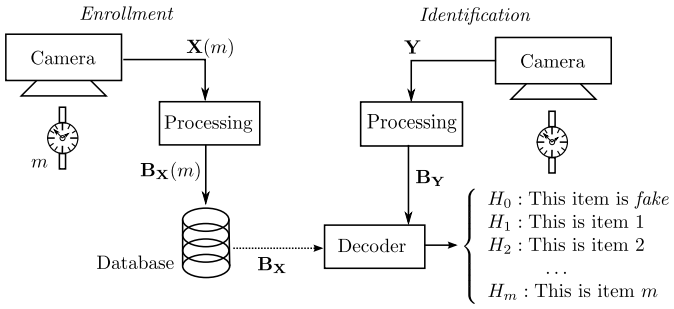



Fig. 3: Item identification based on fingerprints.

in related work [10], [11]. The distribution of Hamming distances between the $\mathbf{B_Y}$ and $\mathbf{B_X}(\mathbf{m})$ under hypothesis $H_0$ and $H_m$ are the same as in the authentication setup; while the distribution of distances between $\mathbf{B_Y}$ and $\mathbf{B_X}(n \neq m)$ always follows $B(L, \frac{1}{2})$.

## III. Attacks on Content Fingerprint Systems

The *attacker* is a malicious entity that wishes to create counterfeit items that are accepted as originals. In literature, it is frequently assumed that an attacker only creates items with features that are unrelated to enrolled genuine items [2], [6], [12], [11], [10]. If we denote the probability of bit-error in the equivalent Binary Symmetric Channel (BSC) of the informed attacker with $p'_b$, then the assumption that counterfeit items are unrelated implies that $p'_b = \frac{1}{2}$. It would then seem reasonable to expect that a producer of counterfeit items makes an effort to let said items resemble genuine items, and consequently $p'_b < \frac{1}{2}$. It remains largely unknown how object protection systems perform when the attacker has at least a moderate degree of success in reducing $p'_b$. Our contribution consists of a revised performance analysis for authentication and identification systems under such an informed attack.

## IV. Performance Evaluation

### A. Authentication

The performance of authentication systems is evaluated in terms of the probability of miss $p_m(\gamma)$ and the SAR $p'_f(\gamma)$. We define $p'_f(\gamma)$ as the probability of false acceptance for an informed attacker. An informed attacker operates with a probability of bit-error $p'_b < \frac{1}{2}$, by using available information which can include $\mathbf{B_X}(m)$. The probability of miss represents the chance that the system decides that the object under investigation is counterfeit although it is in fact the authentic object $m$ that it is claimed to be; $p_m(\gamma)$ is defined as:

$$\begin{aligned} p_m(\gamma) =& \Pr\left[d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) > \gamma L \mid H_m\right] \\ =& \sum_{k=\lceil \gamma L \rceil + 1}^{L} \binom{L}{k} p_b^k (1-p_b)^{L-k} \\ \leq& 2^{-LD(\gamma \| p_b)}, \end{aligned} \tag{5}$$

where $\mathbf{B_Y}$ is generated from $\mathbf{B_X}(m)$. The bound in (5) is according to large deviation theory [2]. Conversely, $p'_f(\gamma)$ represents the probability that any counterfeit object $\mathbf{X}'(m)$

is falsely accepted as the authentic object $m$, i.e., if $\mathbf{B_Y}$ is generated from $\mathbf{B'_X}(m)$, then:

$$\begin{aligned} p'_f(\gamma) &= \Pr\left[d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) \leq \gamma L \mid H_0\right] \\ &= \sum_{k=0}^{\lfloor \gamma L \rfloor} \binom{L}{k} {p'_b}^k (1-p'_b)^{L-k} \\ &\leq 2^{-LD(\gamma \| p'_b)}. \end{aligned} \tag{6}$$

The main goal of the attacker is to maximize the SAR $p'_f(\gamma)$. By selecting the threshold $\gamma L$, one can achieve a reasonable trade-off between $p_m(\gamma)$ and $p'_f(\gamma)$, which can be done based on the classical Neyman-Pearson framework. However, due to strict technical requirements in relevant commercial sectors such as the pharmaceutical or luxury industry, both probabilities should be made asymptotically small. For this reason, one can target the minimization of both probabilities of error simultaneously according to Bayes' rule, by minimizing the maximum of both:

$$p_{max}^{AUTH}(\gamma) = \max\left(p_m(\gamma), p'_f(\gamma)\right). \tag{7}$$

Therefore, the goal of the designer of the authentication system is to find a suitable threshold $\gamma_{opt}^{AUTH} L$ that minimizes $p_{max}^{AUTH}(\gamma)$:

$$\gamma_{opt}^{AUTH} = \arg\min_{\gamma} \left[\max(p_m(\gamma), p'_f(\gamma))\right]. \tag{8}$$

The error exponents corresponding to $p_m(\gamma)$ and $p'_f(\gamma)$ can be defined as:

$$E_m(\gamma) = \lim_{L\to\infty} -\frac{1}{L}\log_2 p_m(\gamma), \tag{9}$$

$$E'_f(\gamma) = \lim_{L\to\infty} -\frac{1}{L}\log_2 p'_f(\gamma), \tag{10}$$

so that the optimization problem given in (8) can be reformulated as:

$$\gamma_{opt}^{AUTH} = \arg\max_{\gamma} \left[\min(E'_f(\gamma), E_m(\gamma))\right]. \tag{11}$$

The maximization problem (11) is difficult to solve, instead we can solve for an equal error rate. That occurs at the equality $E'_f(\gamma) = E_m(\gamma)$ for which we can derive from (6) and (5) that:

$$\gamma_{eq}^{AUTH} = \frac{\log\frac{1-p'_b}{1-p_b}}{\log\frac{p_b(1-p'_b)}{p'_b(1-p_b)}}. \tag{12}$$

In case of a blind attack, i.e., $p'_b = \frac{1}{2}$, the threshold coincides with the one derived in [12].

To exemplify the impact of informed attacks Fig. 4 shows a Receiver Operatonal Characteristic (ROC) for blind ($p'_b = 0.5$) and informed attacks ($p'_b = 0.3, 0.4$) for $L = 128$ and $p_b = 0.1$. As expected, informed attacks considerably degrade the performance of the authentication system.

### B. Identification

The analysis for identification is slightly more complex than for authentication. Under hypothesis $H_m$, we can define and bound the probability of *incorrect identification* $p_{ii}$ [2]:

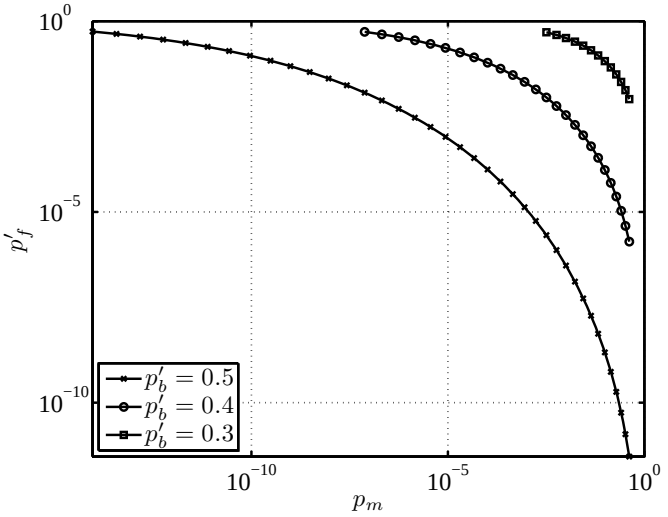



Fig. 4: Authentication ROC curves: blind attack ($p'_b = 0.5$) and informed attacks ($p'_b = 0.3, 0.4$) for $L = 128$.

$$\begin{aligned} p_{ii}(\gamma) =& \Pr\Big[d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) > \gamma L \\ & \bigcup_{n\neq m} d^H(\mathbf{B_Y}, \mathbf{B_X}(n)) \leq \gamma L \mid H_m\Big] \\ \leq& \sum_{\lfloor \gamma L \rfloor + 1}^{L} \binom{L}{k} p_b{}^k (1-p_b)^{L-k} + \sum_{n\neq m} \sum_{k=0}^{\lfloor \gamma L \rfloor} \binom{L}{k} \frac{1}{2^L} \\ \leq^{(a)}& 2^{-LD(\gamma\|p_b)} + 2^{-L\left[D\left(\gamma\|\frac{1}{2}\right)-R\right]}, \end{aligned} \tag{13}$$

where $|\mathcal{M}| \leq 2^{LR}$ and $R$ is the identification rate; (a) follows from the union and Chernoff bounds [11].

Conversely, $p'_f(\gamma)$ represents the probability that any counterfeit object $\mathbf{X}'(m)$ is falsely accepted:

$$\begin{aligned} p'_f(\gamma) =& \Pr\Big[d^H(\mathbf{B_Y}, \mathbf{B_X}(m)) \leq \gamma L \\ & \bigcup_{n\neq m} d^H(\mathbf{B_Y}, \mathbf{B_X}(n)) \leq \gamma L \mid H_0\Big] \\ \leq& \sum_{k=0}^{\lfloor \gamma L \rfloor} \binom{L}{k} {p'_b}^k (1-p'_b)^{L-k} + \sum_{n\neq m} \sum_{k=0}^{\lfloor \gamma L \rfloor} \binom{L}{k} \frac{1}{2^L} \\ \leq& 2^{-LD\left(\gamma\|p'_b\right)} + 2^{-L\left[D\left(\gamma\|\frac{1}{2}\right)-R\right]}. \end{aligned} \tag{14}$$

Note that one of the components of the bounds (13) and (14) is common to both $p_{ii}$ and $p'_f$.

A threshold can be found based on the maximization of the error exponents of the components of $p_{ii}$ and $p'_f$. Let $E_m = D\left(\gamma\|p_b\right)$, $E'_f = D\left(\gamma\|p'_b\right)$ and $E_c = D\left(\gamma\|\frac{1}{2}\right) - R$, then:

$$\gamma_{opt}^{ID} = \arg\max_{\gamma} \left[\min\left(E_m, E'_f, E_c\right)\right]. \tag{15}$$

We will again develop a threshold for an equal-error rate rather than finding an exact solution to (15). An important

difference with the authentication setup is the impact of the number of items in the codebook, or the value of $R$. Let $R_{eq}$ be the rate for which $E_m = E'_f = E_c$, then the equality $E_m = E'_f$ leads to (12), and we can subsequently derive from $E_c = E_m$ that:

$$\begin{aligned} R_{eq} =& D\left(\gamma_{eq}^{AUTH} \| \frac{1}{2}\right) - D\left(p_b \| \gamma_{eq}^{AUTH}\right) \\ =& 1 + p_b \log_2\left(\gamma_{eq}^{AUTH}\right) + (1-p_b)\log_2\left(1-\gamma_{eq}^{AUTH}\right). \end{aligned} \tag{16}$$

Depending on the rate $R$, there are thus two possible scenarios with different conclusions:

1) $E_m = E_c < E'_f$ when $R_{eq} < R \leq 1 - H_2(p_b)$;
2) $E_m = E'_f \leq E_c$ when $0 < R \leq R_{eq}$.

In the first case, the rate is high, and the probability of error is dominated by $E_m$ and $E_c$, while $E'_f$ is insignificant. In other words, the probability that an item $\mathbf{X}'(m)$ is accepted as item $m$ is insignificant compared to the chances that authentic items are falsely rejected or that $\mathbf{X}'(m)$ is accepted as any authentic item $n \neq m$. In this case, attacks do not significantly impact the performance of identification. In this case, the threshold $\gamma_{eq}^{ID} = p_b$ [2]. The error exponents and the corresponding $\gamma_{eq}^{ID}$ are shown in Fig. 5.

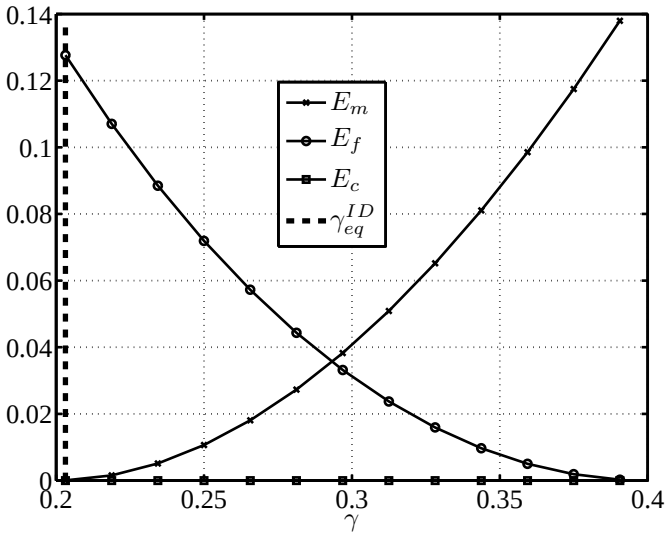



Fig. 5: Error exponents: $p_b = 0.2$, $p'_b = 0.35$, $R = 1 - H_2(p_b)$.

In the second case, the probability of error is dominated by $E_m$ and $E'_f$, as shown in Fig. 6. Contrary to scenarios without attacks, lowering the rate will not significantly improve the performance because $E_c$ does not significantly influence the average probability of error. Consequently, the performance of identification and authentication are comparable. In this case, the threshold for identification is as in (11), the same as for authentication.

## V. Countermeasures

An interesting effect of attacks is that although an attacker only strives to increase $p'_f$, a change in $p'_b$ to lower values causes a change in the optimal threshold, which also affects

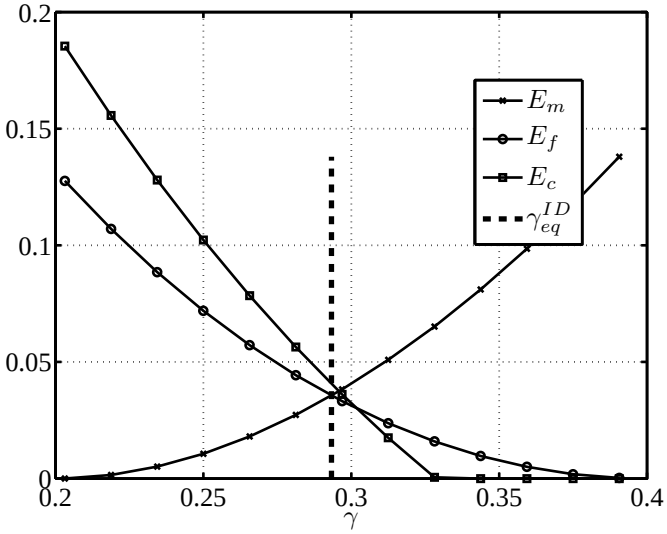


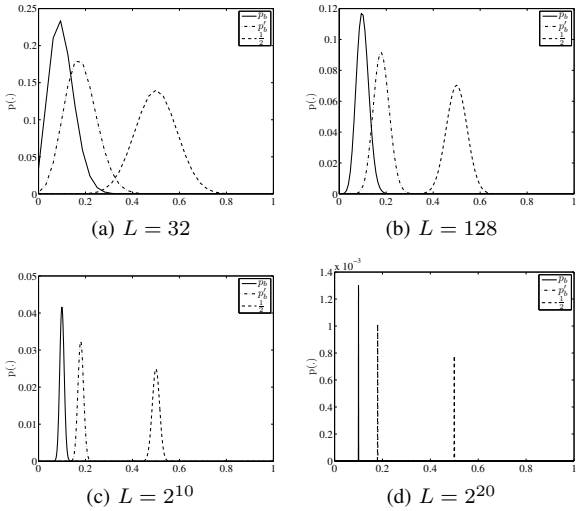

Fig. 6: Error exponents: $p_b = 0.2$, $p'_b = 0.35$ and $R = R_{eq}$.

$p_m$ and potentially $p_{ii}$. In this Section, we will propose a countermeasure to attacks that can regain the lost performance.

### A. Authentication

Clearly, the overlap between the distributions for authentic and counterfeit items based on $\mathbf{B}_{\mathbf{X}}(m)$ is greater than that for authentic and randomly created items. The analysis shows that informed or partially informed attacks pose a far more serious threat to the system than random attacks, and therefore, the former have a far greater influence on the limits of authentication systems based on binary fingerprints. We envision a possible countermeasure: increase the fingerprint length $L$. The effects of the countermeasure are shown in Fig. 7, which shows the distributions of Hamming distances between $\mathbf{B}_{\mathbf{X}}(m)$ and $\mathbf{B}_{\mathbf{y}}$ for authentic items (left-most pmf), related counterfeited items (middle pmf), and the classic case of unrelated counterfeit items (right-most pmf). The overlap between the distributions decreases as $L$ increases, and the increase of $L$ also justifies the low $p_f$ in information-theoretic work [2], [5], [10], [12], where the fingerprints are treated under typicality conditions assuming large $L$. Therefore, the informed attacks are somewhat underestimated in information-theoretic settings, while they represent a more significant threat in practical setups. The effect of increasing the fingerprint length $L$ on practical authentication systems can be seen in Fig. 8, where ROC curves are plotted for $p_b = 0.2$ and $p'_b = 0.35$ and varying $L$. With each successive increase in $L$, the performance increases significantly. An increase of the fingerprint length is likely to increase the amount of information that an attacker can learn about $\mathbf{X}(m)$ based on $\mathbf{B}_{\mathbf{X}}(m)$, which is considered undesirable in, for example, biometric applications. In other words, the presence of attacks can have consequenses that are beyond performance issues.

Therefore, the analysis of the trade-off between the information leakage and its impact on $p'_b$, versus the length of the



(a) $L = 32$ (b) $L = 128$

(c) $L = 2^{10}$ (d) $L = 2^{20}$

Fig. 7: Binomial distributions for increasing $L$.

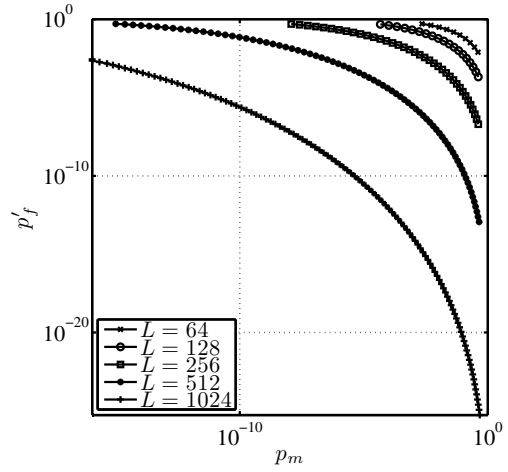



Fig. 8: Authentication ROC curves, for $p_b = 0.2$, $p'_b = 0.35$ and varying $L$.

fingerprint $L$ for the resulting ROC, represents an interesting research problem.

### B. Identification

The performance of identification degrades significantly due to informed attacks versus blind attacks if the rate $R$ is sufficiently low, but the loss can be successfully countered by increasing $L$. If the rate is sufficiently high, the influence of informed attacks is negligible compared to other sources of error, and no countermeasures should be necessary. Nonetheless, an increase in $L$ will decrease $p_m$ and $p'_f$, and thus improve performance.

## VI. Conclusions

We have analyzed blind and informed attacks against authentication and identification systems based on digital content fingerprints. We have shown that an attacker who strives to counterfeit a particular item can decrease the performance of practical authentication systems, both in terms of false acceptance and false rejections. Additionally, we have suggested thresholds that approximate an optimal resilient design. Nonetheless, as long as the probability of bit-error for fake items is higher than for authentic items, the system can be made resilient to attacks by increasing the fingerprint length. When the fingerprint length grows asymptotically long, the results coincide with information-theoretic results.

Identification setups fall into two categories, depending on the ratio between the number of items in the database and the fingerprint length. If the ratio is sufficiently low, the performance is equivalent to that of authentication. If the ratio is sufficiently high, the probability of error is dominated by errors due to the large number of authentic items in the database. The probability of error due to attacks is then insignificant, and subsequently does not play a role. Under the assumptions about the probabilities of bit-error, the resilience to attacks and overall performance can be improved by increasing the fingerprint length.

It remains an open question how an increase of the fingerprint length affects the bit-error rate of the attacker.

## Acknowledgment

This paper was partially supported by SNF project 200020-134595.

## References

[1] A. L. Varna and M. Wu, "Modelling and analysis of correlated binary fingerprints for content identification," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1146–1159, September 2011.

[2] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *Information Theory Workshop*, Dublin , Ireland, August 30 – September 3 2010.

[3] P. Tuyls, B. Skoric, and T. Kevenaar, *Security with Noisy Data: On Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Springer, 2007.

[4] A. Jain, A. A. Ross, and K. Nandakumar, "Introduction to biometrics," 2011.

[5] T. Ingatenko, "Secret-key rates and privacy leakage in biometric systems," Ph.D. dissertation, University of Eindhoven, June 2009.

[6] Y. Wang, S. Rane, S. Draper, and P. Ishwar, "An information-theoretic analysis of revocability and reusability in secure biometrics," in *Workshop on Information Theory and its Applications*, San Diego, CA, February 2011.

[7] T.-T. Do, E. Kijak, L. Amsaleg, and T. Furon, "Enlarging hacker's toolbox: deluding image recognition by attacking keypoint orientations," in *IEEE ICASSP*, Kyoto, Japan, March 2012.

[8] T. van der Putte and J. Keuning, "Biometrical fingerprint recognition: Don't get your fingers burned," in *Fourth Working Conference on Smart Card Research and Advanced Applications (CARDIS)*, 2000.

[9] S. Venugopalan and M. Savvides, "How to generate spoofed irises from an iris code template." *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 385–395, 2011.

[10] P. Moulin, "Statistical modeling and analysis of content identification," in *Information Theory Applications*, San Diego, USA, 2010.

[11] F. Farhadzadeh, S. Voloshynovskiy, and O. Koval, "Performance analysis of identification system based on order statistics list decoder," in *IEEE International Symposium on Information Theory*, Austin, TX, June, 13–18 2010.

[12] F. M. Willems, "Information theory and biometrics," Keynote Lecture at the Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, Germany, October 15–17 2010.