Provable Security for the Fuzzy Fingerprint Vault

Johannes Merkle, Matthias Niesing, Michael Schwaiger
secunet Security Networks AG
D-45128 Essen, Germany
johannes.merkle@secunet.com,
matthias.niesing@secunet.com
michael.schwaiger@secunet.com

Heinrich Ihmor, Ulrike Korte
Bundesamt für Sicherheit in der Informationstechnik
D-53175 Bonn, Germany
heinrich.ihmor@bsi.bund.de
ulrike.korte@bsi.bund.de

***Abstract*—We investigate the security of privacy enhancing techniques for biometric applications.**

The *fuzzy vault* of Jules and Sudan is a technique that allows error tolerant authentication, while preserving the privacy of the reference data. Several publications have proposed its application to fingerprints in order to implement privacy-enhanced biometric authentication. While the heuristic security estimates given are promising, no rigid security analysis has been presented so far. We explore if and under what circumstances a provably secure fuzzy fingerprint vault can be implemented. Based on bounds on the loss of entropy for the general fuzzy vault and realistic models for minutiae distributions, we deduce lower bounds for attacks that attempt to recover the template. Furthermore, we show how to select optimal parameters and evaluate both, minimum minutiae match rates and minimum number of minutiae needed to obtain an appropriate security level. Our results indicate that a provable secure scheme is hard to achieve with current fingerprint technology.

***Keywords*-biometric template protection; fingerprint; fuzzy vault**

I. Introduction

The storage of biometric reference data (templates) introduces considerable risks for biometric authentication systems and raises serious concerns regarding privacy and data protection. In order to solve this issue, *biometric template protection* systems [1] use reference data which reveal only very limited information on the biometric trait. One of the most promising approaches is the fuzzy vault [2], which applies Reed-Solomon decoding to redundantly bind the biometric template to a randomly selected secret polynomial. This scheme only stores helper data that allow the recovery of the polynomial from a query template that is sufficiently "close" to the enrolled template.

Without any doubt, fingerprint minutiae are the biometric traits most widely deployed for authentication. On the other hand, the variety and extent of errors in minutiae measurements, particularly, frequent insertions, omissions and re-ordering of the measured minutiae, pose a considerable challenge to template protection schemes [3]. The fuzzy vault is able to tolerate such errors and hence is particularly interesting for minutiae-based authentication.

Several publications report successful implementation of the fuzzy vault scheme based on minutiae: The constructions of [4], [5] and [6] use minutiae locations, while [7] uses both location and orientation of the minutiae. While the error rates reported seem promising, the level of protection provided for the biometric data and the secret key remains open: either template security is not analyzed at all [6], or merely naive brute force attacks are considered [4], [5], [7]. The subsequent publication of more efficient attacks demonstrates that these security assessments have been too optimistic [8], [9].

For the fuzzy vault, theoretical results are known from which rigid security estimates could be deduced. In particular, Dodis, et al. [10] proved lower bounds for the loss of entropy which determines the maximum success probability of an attack trying to guess the template or the key from the helper data (see Section III-A for details). In addition, an attacker's success probability depends on the original entropy of the biometric feature vector - or, equivalently, its redundancy. Therefore, a realistic estimation of the entropy of the biometric feature vector is a key aspect for a sound security analysis.

In this publication, we explore if and under what circumstances a provable secure fuzzy fingerprint vault can be implemented. In particular, we generalize the bounds of [10] to the case where the minutiae and chaff points are chosen with a minimum distance to reduce false matchings. We also give an exact estimate for the entropy of a feature vector consisting of minutiae location data. Furthermore, we show, how the parameters can be optimized with respect to the resulting lower security bounds. Finally, we determine minimum minutiae match rates for a desired security level of 2^{50}.

This article is structured as follows: In Section II, we give a description of the scheme. In Section III, we conduct a theoretical analysis of its security and error robustness. Section IV presents methods for parameter optimization with respect to the deduced security bounds, and Section V provides results using empirical data. A conclusion is given in Section VI.

II. The fuzzy vault for fingerprints

We give a brief description of the fuzzy vault for fingerprints. The scheme uses minutiae locations and essentially matches the constructions of [4], [5] and [6] with minor modifications:

- According to [11], the minutiae of a single finger do not provide sufficient entropy to extract a secure cryptographic key. Therefore, we allow to use minutiae from more than one finger. The minutiae of the different fingers can be fused on a feature level.
- The polynomial is not evaluated on the biometric information but on the indices of the minutiae in the vault. More precisely, P is evaluated on the encodings $E(i)$ of the indices i, where E is an injective encoding function from $1, \ldots, r \subset \mathbf{Z}$ to $\mathbf{F}_q$. This modification minimizes the size of the function values stored as helper data and thus the loss of entropy (see Section III-A).

In addition, we assume that the query fingerprints is correctly aligned to the stored minutiae and chaff points. Furthermore, we do neither apply quantization to the minutiae location as done in [6] and [7], nor do we rely on minutiae matching by human experts like in [4] and [5], but use a tolerance parameter δ for the Euclidean distance between the matching points.

A. Enrollment

Let $k < t < r \leqslant q$. For each user, a random polynomial P of degree less than k over a finite field $\mathbf{F}_q$ is selected. The coefficients of this polynomial represent the secret key of the scheme. Then, a set T of t minutiae of the user is determined. This set of minutiae is amended by random chaff points, resulting in a set of r points, containing t genuine minutiae and $r - t$ chaff points. A minimum distance of d is enforced among minutiae and chaff points to reduce errors during verification. Furthermore, in order to ensure that minutiae and chaff points within the helper data are indistinguishable, they are lexicographically ordered.

For all genuine minutiae $\mathbf{m}_j$, where j is its index after applying the lexicographic order, $y_j = P(j)$ is computed. For each chaff point $\mathbf{m}_j$, where j is its index in the lexicographic order, a random value $y_j \neq P(j)$ is chosen. As helper data, the lexicographically ordered list of minutiae and chaff points, paired with the corresponding y_j values, is stored in the database.

B. Authentication

We only consider an authentication in the verification scenario, where the identity of the user is known a priori.

In order to verify the identity of a user, the minutiae are measured from a query fingerprint. Then the matches between these minutiae and the minutiae and chaff points contained in the helper data are identified. Precisely, for each minutiae in the query fingerprint, the closest point in the helper data with Euclidean distance smaller than a threshold δ is identified. The indices of the matching minutiae and chaff points in the helper data, along with the corresponding y_i values, are used to recover the secret polynomial P (see Section II-C). If the number of genuine minutiae among the matches is sufficiently high (see Section II-C for a discussion), the polynomial can be recovered.

C. Recovery of the polynomial

The unlocking of the vault (during authentication) requires the recovery of the secret polynomial from a set of points (j_i, y_{j_i}), some of which (those resulting from matches with minutiae) lying on the polynomial P, while others (resulting from matches with chaff points) do not. For this task, an Reed-Solomon decoder is needed that on input $(j_1, y_{j_1}), \ldots, (j_x, y_{j_x}) \in \mathbf{F}_q^2$ with $x \geqslant k$, outputs $e_0, \ldots, e_{k-1} \in \{0, \ldots, q-1\}$, so that $y_{j_i} = P(j_i)$ holds for at least k of the (j_i, y_i) with $P(x) = \sum_{i=0}^{k-1} e_i x^i$, if such a polynomial exists. We assume that the Peterson-Berlekamp-Massey-decoder is used as suggested in [2]. (As pointed out in [4], this decoder degenerates to a brute-force polynomial interpolation for $x = k$.) This technique is successful if $(x+k)/2$ of the x points handed over to the decoder are correct. Although there are Reed-Solomon decoders that can decode with only $\sqrt{xk}$ correct points, they do not offer any advantage for the fuzzy vault, because $\sqrt{xk}$ is quite close to $(x+k)/2$ for typical parameters, and they are computationally much less efficient [2].

III. Security analysis

Throughout this article, let all logarithms be to the base 2. Furthermore, let P(X) denote the probability of an event X and let $\mathbf{E}_{a \leftarrow A}\left[f(a)\right]$ be the expectation of the function value of a random variable A. The *min-entropy* of a random variable A is given by

$$\mathbf{H}_\infty(A) := -\log(\max_a(\mathrm{P}(A = a))),$$

and the *average min-entropy* of A given B is defined as

$$\widetilde{\mathbf{H}}_\infty(A \mid B) := -\log\left(\mathbf{E}_{b \leftarrow B}\left[2^{-\mathbf{H}_\infty(A \mid B=b)}\right]\right)$$

For a biometric encryption scheme with feature vector T and helper data Y, we call $\mathbf{H}_\infty(T) - \widetilde{\mathbf{H}}_\infty(T \mid Y)$ the *loss of entropy.*

A. Minimum attack complexity

The following result shows that the security of the fuzzy vault for fingerprints can be lower bounded by the average min-entropy of the biometric feature vector given the helper data. The result is a trivial adaptation of Theorem 1 and Lemma 2 from [12], and holds (with according notations) for any biometric encryption scheme, in which the secret key and the helper data uniquely determine the biometric feature vector.

Theorem 1. *Any algorithm that takes as input the helper data Y and tries to output the secret polynomial $P(x) = \sum_i e_i x^i$ or the set of minutiae T has an average success probability of at most $2^{-\widetilde{\mathbf{H}}_\infty(T \mid Y)}$.*

B. Loss of entropy

By definition, the average min-entropy of the biometric feature vector given the helper data is the difference between the entropy of the feature vector and the loss of entropy. We now turn to the estimation of the latter quantity. Subsequently, let $\mathcal{M}$ be the set of all possible minutiae locations and $n = |\mathcal{M}|$ the number of possible values for a minutia or chaff point.

In [10], Lemma D.1, a lower bound for the loss of entropy in the original fuzzy vault scheme has been given. In the case $d = 1$, i.e., if the minimum distance is trivial and the minutiae and chaff points only need to be distinct, the result can be directly applied to our implementation. The proof is a simple adaptation of the proof of Lemma D.1 in [10].

Theorem 2. *If $d = 1$, the loss of entropy is at most $(t-k)\log q - \log\binom{r}{t} + \log\binom{n}{t} + 2$, i.e.,*

$$\widetilde{\mathbf{H}}_\infty(T \mid Y) \geqslant \mathbf{H}_\infty(T) - (t-k)\log q + \log\binom{r}{t} - \log\binom{n}{t} - 2.$$

Proof: By Lemma 3.1 in [10]

$$\widetilde{\mathbf{H}}_\infty(T \mid Y) \geqslant \mathbf{H}_\infty(T, Y) - \lambda,$$

where 2^λ is the number of possible values that Y can take. The information contained in T and Y is composed of four parts: The set of minutiae T, the set of chaff points, the y_i-values for the minutiae, and the y_i-values for the chaff points. The entropy of the $r-t$ chaff points is given by $\log\binom{n-t}{r-t}$, because they are randomly selected from all $n-t$ potential points that are distinct from the t minutiae. Given T, there is a one-to-one correspondence between the y_i-values for the minutiae and the random polynomial P; hence their entropy is $k \log q$. Finally, the y_i-values for the chaff points are randomly selected from $\mathbf{F}_q \backslash \{P(i)\}$, and therefore their entropy is $(r-t)\log q - 1$. This sums up to

$$\mathbf{H}_\infty(T, Y) = \mathbf{H}_\infty(T) + \log\binom{n-t}{r-t} + k\log q + (r-t)\log(q-1).$$

On the other hand, since the minutiae and chaff points in Y are in lexicographic order, we have $2^\lambda = \binom{n}{r} q^r$. Using $(r-t)\log\frac{q}{q-1} < q\log\frac{q}{q-1} \leqslant 2$ and $\binom{n}{r}\binom{r}{t} = \binom{n}{t}\binom{n-t}{r-t}$ this yields the result. □

This result can be interpreted as follows:

- The term $(t-k)\log q$ represents the information leaked by the redundantly encoded secret polynomial. Precisely, this term is composed of the $t\log(q)$ bits of information revealed by the y_i-values corresponding to the genuine minutiae and the $k\log(q)$ of information contained in the secret polynomial.
- The term $\log\binom{r}{t}$ estimates the amount of security contributed by "hiding" the t genuine minutiae among the r chaff points.
- The term $\log\binom{n}{t}$ refers to the information leaked by publishing T as part of the helper data.

Since $\mathbf{H}_\infty(T) \leqslant \log\binom{n}{t}$, the lower bound (1) is positive (and hence meaningful) only if $q^{t-k} \leqslant \binom{r}{t} \leqslant \binom{q}{t} \leqslant q^t/(t!)$, which implies $q > (t/\mathrm{e})^{t/k}$. The exponent t/k defines the error correction capabilities of the scheme and, according to our experiments, must be larger than 1.5 to achieve a satisfactory false rejection rate (FRR). Therefore, we obtain a scheme with provable security according to Theorems 1 and 2 only if q is considerably greater than $(t/\mathrm{e})^{1.5}$.

For the case $d \geqslant 2$, we have to analyze the effect of the minimum distance to the number of possible choices for the chaff points and the possible values for the helper data Y. Subsequently, we will use the following definitions.

For a point $\mathbf{m} \in \mathcal{M}$ let $B_d(\mathbf{m})$ denote the set of points in $\mathcal{M}$ that have Euclidean distance to $\mathbf{m}$ smaller than d, and let $V_d = 1 + 4\sum_{i=1}^{\lceil d-1 \rceil}\left\lceil\sqrt{d^2 - i^2}\right\rceil$ be the number of integer points $(x, y) \in \mathbf{Z}^2$ with Euclidean norm smaller than d. Obviously, $|B_d(\mathbf{m})| \leqslant V_d$.

Since the minutiae and chaff points are selected with minimum distance d, the d-sphere centered at a selected point is excluded from the potential values for subsequent points. If the d-sphere neither juts out beyond $\mathcal{M}$ nor intersects with the d-spheres of the previously selected points, the number of possible choices for the next point is reduced by exactly V_d; otherwise, the reduction is smaller. On the other hand, if two selected points have distance $d' < 4d$, additional points lying between these points are excluded as potential choices for the subsequent points.

These effects make an exact estimation of the number of possible choices for the chaff points or the number of possible values for Y virtually impossible. However, for $rV_d \ll n$, the likelihood that a selected point is too close to the boundary of $\mathcal{M}$ or to a previously selected point is small. Furthermore, the effects of such cases can influence the number of choices in both directions; thus, the inaccuracies partially balance. Therefore, for $rV_d \ll n$, the approximation that, on average, each point reduces the number of choices for the subsequent points by V_d is quite accurate. Consequently, we can approximate the number of chaff points by $V_d^{r-t}\binom{n/V_d-t}{r-t}$ and the number of possible values for Y by $V_d^{r}\binom{n/V_d}{r}$. Analogously to Theorem 2, we obtain the following result:

Theorem 3. *For $rV_d \ll n$, the maximal loss of entropy is approximately $(t-k)\log q - \log\binom{r}{t} + \log\binom{n/V_d}{t} + t\log V_d + 2$, i.e., $\widetilde{\mathbf{H}}_\infty(T\,|\,Y) \geqslant E$ with*

$$\begin{aligned} E \approx \mathbf{H}_\infty(T) - (t-k)\log q + \log\binom{r}{t} \\ - \log\binom{n/V_d}{t} - t\log V_d - 2. \qquad (1) \end{aligned}$$

C. Entropy of the feature vector

The entropy of the feature vector T is defined by the maximum likelihood that it takes a certain instance M. Since for the parameters of interest the number of possible instances by far exceeds the number of persons for which minutiae information is available, we can estimate the entropy of T only by modeling its probability distribution. Several publications have tried to estimate the individuality of minutiae information in fingerprints, e.g., [13] and [14]. However, their analysis already takes into consideration the error tolerance of the minutiae matching algorithm and is therefore not applicable for the determination of the raw entropy $\mathbf{H}_\infty(T)$.

We model the probability distribution of T by a random process **Select_T**, where the t minutiae are successively chosen. The first minutia $\mathbf{m}_1$ is selected according to a distribution $\mathcal{D}$ defined over $\mathcal{M}$. All subsequent minutiae $\mathbf{m}_i$ are selected to the same distribution $\mathcal{D}$ restricted to the areas in $\mathcal{M}$ not covered by the d-spheres $B_d(\mathbf{m}_1), \ldots, B_d(\mathbf{m}_{i-1})$ around the previously chosen minutiae.

We do not assume any statistical dependency between the locations of the individual minutiae, except that they have the minimum distance d. Although it is known that minutiae tend to overdisperse on a small scale and to cluster on a large scale [15], we believe that these phenomena are sufficiently addressed by enforcing the minimum distance d and by the non-uniformity of the distribution $\mathcal{D}$. (The overdispersion observed in [15] could be due to the a minimum distance enforced by minutiae extraction algorithms, e.g., see [16].)

We can show the following result:

Theorem 4. *If T is chosen according to the random process* **Select_T** *and the maximum likelihood of a minutiae location is $1/\psi$, then*

$$\mathbf{H}_\infty(T) \geqslant \log\binom{\psi/V_d}{t} + t\log V_d$$

Proof: Let $\mathrm{P}(X)$ denote the probability o random even X. Furthermore, for $i = 1, \ldots, t$ let M_i let be the random variable of the i-th point output by **Select_T**. By M we denote the random variable chosen according to $\mathcal{D}$. Then by definition

$$\begin{aligned} 2^{-\mathbf{H}_\infty(T)} &= \max\left(\mathrm{P}\left(\{M_1, \ldots, M_t\} = \{\mathbf{m}_1, \ldots, \mathbf{m}_t\}\right)\right) \\ &\leqslant t!\max\left(\mathrm{P}\left(M_1 = \mathbf{m}_1, \ldots, M_t = \mathbf{m}_t\right)\right), \qquad (2) \end{aligned}$$

where the maximum is taken over all $\mathbf{m}_1, \ldots, \mathbf{m}_t$. The latter probability $\mathrm{P}\left(M_1 = \mathbf{m}_1, \ldots, M_t = \mathbf{m}_t\right)$ can be expanded to

$$\prod_{i=1}^{t} \mathrm{P}\left(M_i = \mathbf{m}_i \,|\, \forall j < i:\; M_j = \mathbf{m}_j\right).$$

The first term has an empty condition and is limited by $1/\psi$, while the other factors can be estimated as follows:

$$\begin{aligned} &\mathrm{P}\left(M_i = \mathbf{m}_i \,|\, M_1 = \mathbf{m}_1, \ldots, M_{i-1} = \mathbf{m}_{i-1}\right) \\ &\quad = \mathrm{P}\left(M = \mathbf{m}_i \,|\, M \notin B_d(\mathbf{m}_1) \cup \ldots \cup B_d(\mathbf{m}_{i-1})\right) \\ &\quad = \frac{\mathrm{P}\left(M = \mathbf{m}_i \wedge M \notin B_d(\mathbf{m}_1) \cup \ldots \cup B_d(\mathbf{m}_{i-1})\right)}{\mathrm{P}\left(M \notin B_d(\mathbf{m}_1) \cup \ldots \cup B_d(\mathbf{m}_{i-1})\right)} \\ &\quad \leqslant \frac{\mathrm{P}\left(M = \mathbf{m}_i\right)}{1 - \mathrm{P}\left(M \in B_d(\mathbf{m}_1) \cup \ldots \cup B_d(\mathbf{m}_{i-1})\right)} \qquad (3) \end{aligned}$$

By assumption, the numerator is at most $1/\psi$, while the probability in the denominator is limited by the term $\left|B_d(\mathbf{m}_1) \cup \ldots \cup B_d(\mathbf{m}_{i-1})\right|/\psi$, which is at most $(i-1)V_d/\psi$. Consequently, with (2) we obtain

$$2^{-\mathbf{H}_\infty(T)} = t!\prod_{i=1}^{t-1}\frac{1}{\psi - i\cdot V_d}.$$

The desired result now follows by elementary transformations. □

By combining Theorem 3 with Theorem 4 we obtain the following Theorem.

Theorem 5. *For $d \geqslant 1$, $\widetilde{\mathbf{H}}_\infty(T \,|\, Y) \geqslant E$ with*

$$E \approx \log\binom{\psi/V_d}{t} - (t-k)\log r + \log\binom{r}{t} - \log\binom{n/V_d}{t} - 2,$$

where $1/\psi$ is the maximum likelihood of a minutiae location.

IV. Optimization of Parameters

In this section we try to determine criteria for the optimal selection of parameters and to derive estimates on the achievable security according to Theorem 1 and Heuristic 5. We do this by estimating the maximum of E over t, k and r for a given decoding complexity.

A. Minimizing the fields size

In order to maximize the approximate lower bound for the remaining entropy according to Theorem 5, we set $q = r$. Furthermore, since $n > \psi \gg tV_d$, we have $\binom{n/V_d}{t}/\binom{\psi/V_d}{t} \approx (n/\psi)^t$. In general, we cannot assume $r \gg t$; therefore, we use the approximation

$$\binom{r}{t} \approx r^t \left(1 - \frac{t-1}{2r}\right)^t /(t!),$$

which is much tighter than $\binom{r}{t} \approx r^t/(t!)$. With Stirling's approximation for $t!$, this results in the estimate

$$E \lessapprox k \log r - t \log\left(\frac{nt}{\mathrm{e}\psi}\left(1 - \frac{t-1}{2r}\right)\right) - \frac{1}{2}\log(2\pi t) - 2. \qquad (4)$$

B. Selecting the minimum distance for minutiae

In (4), the remaining entropy is independent of the minimum distance d enforced for minutiae and chaff points. However, the parameter d limits the maximum number of chaff points to approximately $\rho n/V_d - t$, where ρ is the maximum packing density of d spheres in $\mathcal{M}$. On the other hand, d should not be smaller than δ, to prevent false matchings of minutiae in the query fingerprint with chaff points during authentication. Setting $d = 2\delta$ will already completely prevent such false matchings with minutiae that are also present in T, but smaller values might already reduce their number to a minimum.

C. Optimizing the degree of the polynomial

The parameter k must be set, so that with sufficient probability the secret polynomial can be recovered efficiently from a genuine query fingerprint. Subsequently we analyze the expected complexity of this task. Let m_{c} denote the number of *correct matches*, i.e., the matches between the query fingerprint and the genuine minutiae, and let m_{f} be the number of *false matches*, i.e., the matches between the query fingerprint and the chaff points. Obviously, decoding is only possible if $m_{\mathrm{c}} \geqslant k$.

It has been shown in [4] that, on average, the Reed-Solomon decoding of the polynomial using x points requires

$$\binom{m_{\mathrm{c}} + m_{\mathrm{f}}}{x}\left(\sum_{i=\max(\lceil\frac{x+k}{2}\rceil, x-m_{\mathrm{f}})}^{\min(x,m_{\mathrm{c}})} \binom{m_{\mathrm{f}}}{x-i}\binom{m_{\mathrm{c}}}{i}\right)^{-1}$$

trials, where the parameter x must fulfill $k \leqslant x \leqslant \min(2m_{\mathrm{c}} - k, m_{\mathrm{c}} + m_{\mathrm{f}})$. This expression is difficult to analyze theoretically. Numerical evaluation shows that for $m_{\mathrm{c}} - k \leqslant m_{\mathrm{f}} \leqslant m_{\mathrm{c}} + 2m_{\mathrm{c}}/(m_{\mathrm{c}} - k)$, the decoding complexity is minimized for $x = 2m_{\mathrm{c}} - k$. In this case, the sum collapses to the term for $i = m_{\mathrm{c}}$ and hence the minimum decoding complexity is

$$C_{\min}(m_{\mathrm{c}}, m_{\mathrm{f}}, k) = \binom{m_{\mathrm{c}} + m_{\mathrm{f}}}{2m_{\mathrm{c}} - k}\binom{m_{\mathrm{f}}}{m_{\mathrm{c}} - k}^{-1}. \qquad (5)$$

In the case $m_{\mathrm{f}} = m_{\mathrm{c}} - k$, we have $x = 2m_{\mathrm{c}} - k = m_{\mathrm{c}} + m_{\mathrm{f}}$ and $C_{\min}(m_{\mathrm{c}}, m_{\mathrm{f}}, k)$ evaluates to 1. For $m_{\mathrm{f}} = m_{\mathrm{c}} - k + i$ with $i = 1, 2, \ldots, m_{\mathrm{c}}/(m_{\mathrm{c}} - k) - 1$ equation (5) yields

$$C_{\min}(m_{\mathrm{c}}, m_{\mathrm{f}}, k) = \frac{(2m_{\mathrm{c}} - k + 1)\cdots(2m_{\mathrm{c}} - k + i)}{(m_{\mathrm{c}} - k + 1)\cdots(m_{\mathrm{c}} - k + i)}.$$

This equation shows that, for $m_{\mathrm{c}} - k \leqslant m_{\mathrm{f}} < m_{\mathrm{c}} - k + m_{\mathrm{c}}/(m_{\mathrm{c}} - k)$, the minimum decoding complexity increases exponentially with $i = m_{\mathrm{f}} - m_{\mathrm{c}} + k < m_{\mathrm{c}}/(m_{\mathrm{c}} - k)$. Numerical evaluation reveals that the exponential growth continues (with slowing pace) for $m_{\mathrm{f}} - m_{\mathrm{c}} + k \geqslant m_{\mathrm{c}}/(m_{\mathrm{c}} - k)$. Consequently, we find that the decoding complexity is an exponential function in $m_{\mathrm{f}} - m_{\mathrm{c}} + k$.

On the other hand, the number m_{c} of correct matches will typically disperse considerably between different authentications due to variations in the fingerprint image quality. Thus, if k is larger than the expectation of $m_{\mathrm{c}} - m_{\mathrm{f}}$, the fraction of cases in which decoding is not feasible anymore, can become quite high. As a consequence, we set k to the expectation of $m_{\mathrm{c}} - m_{\mathrm{f}}$ in order to optimize the remaining entropy while limiting the decoding complexity.

Remark: Depending on the specific distribution of the number of correct matches and the requirements on decoding complexity imposed by the application scenario, it may be appropriate to select higher values for k. We will investigate the impact of increasing k in our numerical evaluation in Section V-C.

We estimate the mean values for m_{c} and m_{f} as follows:

- It is reasonable to assume that the average number of correct matches is a linear function of t, i.e., $m_{\mathrm{c}} = \mu t$, where μ is the average match rate independent of t.
- If $rV_\delta \ll n$, the number of points in $\mathcal{M}$ covered by the tolerance areas $B_\delta(\mathbf{m}_i)$ around the chaff

points $\mathbf{m}_i$ can be estimated as $(r-t)V_\delta$. (Since minutiae of the query fingerprint that lie within the tolerance area of a chaff points can still be correctly matched with a minutiae in T, this estimate is even conservative.) Therefore, we can estimate the average number m_f of false matches by $\tau(r-t)V_\delta/n$, where τ is the average number of surplus minutiae from the query fingerprints, i.e., the average number of minutiae in the query fingerprints that do not match with the stored minutiae.

As we set k to the expectation of $m_\mathrm{c} - m_\mathrm{f}$, these estimations yield

$$k = t\mu - (r-t)\frac{\tau V_\delta}{n}. \tag{6}$$

Using approximation (4) this yields $E \lessapprox f(t,r)$ with

$$f(t,r) = \left(t\mu - (r-t)\frac{\tau V_\delta}{n}\right)\log r - t\log\left(\frac{nt}{\mathrm{e}\psi}\left(1-\frac{t-1}{2r}\right)\right) - \frac{1}{2}\log(2\pi t) - 2.$$

D. Maximizing the Bound for the Entropy

For fixed δ, n, μ and s, we try to estimate the maximum remaining entropy E by finding the maximum of the function $f(t,r)$ over r. The maximum is assumed, where the first derivation $\frac{\partial f(t,r)}{\partial r}$ is zero. It is easy to see that this is equivalent to $t^2+a(r)t+b(r)=0$ with $a(r) = 2\mu nr+\tau V_\delta r(3+\ln(r))$ and $b(r) = -2\tau V_\delta r^2(\ln(r)+1)$. For $r>0$, one of the two solutions is negative and can thus be neglected. Consequently, for every r, $f(t,r)$ takes its maximum at

$$t_0(r) = -a(r)/2 + \sqrt{a(r)^2/4 - b(r)}.$$

Consequently, the function $f(t_0(r),r)$ upper bounds E for a given r, and the maximum of $f(t_0(r),r)$ over r yields a general upper bound for E. Thus, we can estimate the best provable security bound according to Theorems 1 and 2 that can be achieved for given δ, n, μ and s, by numerically determining the maximum of $f(t_0(r),r)$ over the relevant range of r. As argued in Section IV-A, it is reasonable to set $d \geqslant \delta$; hence, the relevant range is given by $1 \leqslant r \leqslant 0.45n/V_\delta$, where the factor 0.45 represents the density of the random packing by chaff points [4].

Since for fixed $t, r \geqslant 1$, the value $f(t,r)$ is monotonically increasing with the match rate μ, we can determine the minimum value μ_min for which the maximum of $f(t_0(r),r)$ is greater than a certain security level S. Since $E \lessapprox f(t_0(r),r)$, this value μ_min is an approximate lower bound for the average match rate required to obtain a scheme with security 2^S according to Theorem 1 and Theorem 5, so that in the average case the polynomial can be recovered with one trial.

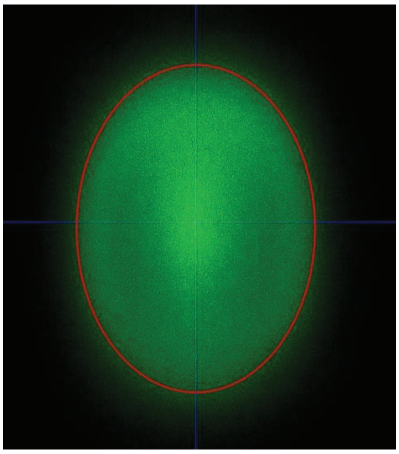

Figure 1. Distribution of minutiae positions in 82800 fingerprints and the ellipse $\mathcal{E}$ from where minutiae are considered. The brightness of pixels corresponds to the frequency of minutiae occurrence at this position.

V. Results

We evaluate whether and to what extend a (heuristically) provably secure fuzzy fingerprint vault is feasible. In particular, for different values for δ and for typical values for n, ψ, τ we determine the minimum match rates required to achieve a security of 2^{50} according to Theorem 1 and Theorem 5. We compare these minimum match rates with match rates observed in practice.

A. Evaluation of Minutiae Distribution

In order to estimate n and ψ, we have empirically determined the statistical distribution of minutiae locations. We evaluated the location of 5.8 million minutiae extracted from 82800 imprints that were taken from 9200 fingers with 3 different sensors having 500 DPI.[1] For this evaluation, the fingerprints had been pre-aligned, so that the center of mass of all minutiae coincides with the image center and the longest distance between two minutiae locations was vertically aligned.

It turned out that 83% of all minutiae occurred in an area defined by an ellipse that covers approximately 87000 pixels, which roughly corresponds to 2.25 cm^2. Outside this ellipse, the density of minutiae decreases drastically. Therefore, it is reasonable to restrict the fuzzy vault to minutiae and chaff points inside this area. This gives an estimate $n \approx 87000 \cdot f$, where f is the number of fingers from which the minutiae are gathered. The distribution of the minutiae positions and the ellipse are shown in 1.

The maximum frequency of a minutiae location was 112, which corresponds to a maximum probability of

[1] The fingerprints were taken from a non-public database set up in the course of a previous project of the Bundesamt für Sicherheit in der Informationstechnik.

	$\delta = 5$	$\delta = 7$	$\delta = 10$
$s = 20$	82.2%	89.6%	97.0%
$s = 35$	87.9%	95.5%	-
$s = 50$	91.7%	99,1%	-

Table I
Minimum match rates required to achieve a provable security of 50 bits.

	$\delta = 5$	$\delta = 7$	$\delta = 10$
$s = 20$	0.67%	0.54%	0.41%
$s = 35$	0.54%	0.38%	-
$s = 50$	0.45%	0.28%	-

Table II
Linear factor, by which the minimum math rates given in Table I decrease with increasing k.

a minutiae location inside the ellipse of approximately $112/5800000/0.83 \approx 2^{-15.4}$. This results in an approximation $n/\psi \approx 2$. This approximation is independent from the number of fingers used for the fuzzy vault, as both ψ and n scale linearly with the number of fingers.

B. Estimating the number of surplus minutiae

According to [3], a good-quality live-scan fingerprint has 20–70 minutiae. Since $f(t, r)$ decreases with an increasing average number τ of minutiae from the query fingerprints not matching with genuine minutiae, it might be a good idea to use only the most reliable minutiae of the query fingerprints, e.g., by evaluating minutiae quality indices output by the feature extraction algorithm. However, the extent of the filtering should be carefully balanced with the match rates achieved with the reduced number of minutiae. We will subsequently assume that τ equals $s \cdot f$ with $s \leqslant 50$, where s is the average number of surplus minutiae from a single query fingerprint, i.e., the average number of minutiae from a single query fingerprint that are not matching with genuine minutiae.

C. Numerical Parameter Optimization

In the previous Sections, we found the approximations $n/\psi \approx 2$ and $\tau/n \approx s/87000$ from empirical data. Using these estimations and various values for δ and s, we applied the method described in Section IV-D to determine the minimum match rate required to achieve a security level of 2^{50} according to Theorem 1 and Theorem 5. We numerically computed the maximum value of the function $f(t_0(r), r)$ over the range $1 \leqslant r \leqslant 0.45n/V_\delta$ using the computer algebra system PARI/GP. The minimum match rates, at which this maximum exceeds 2^{50}, are listed in Table I for different values of δ and s. A "–" denotes that a remaining entropy of 50 is not achieved at all.

The security bounds are very sensitive to changes of the match rate. For instance, for the parameters given in Table I, a decrease of the match rate by only 2% results in a reduction of the achievable security of 12 to 38 bits; a larger reduction is observed for higher match rates.

As explained in Section IV-C, under specific circumstances it may be reasonable to select k greater than our choice $k_0 := t\mu - (r - t)\tau V_\delta/n$, particularly if the dispersion of the number of correct matches is small, or if a larger decoding complexity is acceptable. In these cases, the match rate required for a certain security level decreases. In particular, setting $k = k_0 + \epsilon$ increases the entropy estimation 4 by $\epsilon \log(r)$. For a given match rate μ, this results in the same amount of entropy as setting $k = k_0$ with match rate $\mu + \epsilon/t$. Thus, for a given security level, decreasing k by ϵ compensates an decrease of the match rate by ϵ/t. As a consequence, the minimum match rates required for a security level of 2^{50} with $k = k_0 + \epsilon$ can be estimated by subtracting $\epsilon/t_{\max}$ from the values given in Table I, where $t_{\max}$ is the value of $t_0(r)$, for which $f(t_0(r), r)$ is maximal. We give the respective values of $1/t_{\max}$ in Table II.

To get a feeling for the number of minutiae and thus for the number of fingers needed for a provable secure scheme, we evaluate the minimum value t for which we still obtain a remaining entropy of 2^{50} for a given μ. For this evaluation we apply the following method:

First, we observe that $t_0(r)$ is continuous and unbounded for $r > 0$ and is zero for $1/e$. Thus, for every $t' > 0$ there is a r' with $t' = t_0(r')$; by definition of $t_0(r)$, this pair (t', r') maximizes the function $f(t, r')$ over t. Consequently, it suffices to search through all pairs $(t_0(r), r)$ to find the minimal t with $f(t, r) \geqslant 2^{50}$.

On the other hand, the approximation of the remaining entropy E by the continuous function $f(t, r)$ will result in an artificially smooth curve for the minimal t. In particular, in the definition of f we have replaced k by a real number, whereas in practice, k can only take integer values. The small deviations in k implied by the truncation result in significant variations of E and hence, of the minimal t required for a certain value of E. To obtain a more realistic estimation of the minimal t, we set $k_0(r) = \lfloor t_0(r)\mu - (r - t_0(r))\tau V_\delta/n \rfloor$ and determine the minimal $t_0(r)$ for that (4) yields at least a value of $E \geqslant 2^{50}$ with $t = t_0(r)$ and $k = k_0(r)$.

Fig. 2 shows the minimal number t of minutiae required for a security of 2^{50} as a function of the average match rate μ for various parameters δ and s.

These curves also allow estimating the impact of selecting a larger k to the minimum value t of minutiae. As explained above, selecting $k = k_0(r) + \epsilon$ compensates a decrease of the match rate by ϵ/t. Therefore, for small ϵ, the minimum value of t yielding a security of 2^{50} with

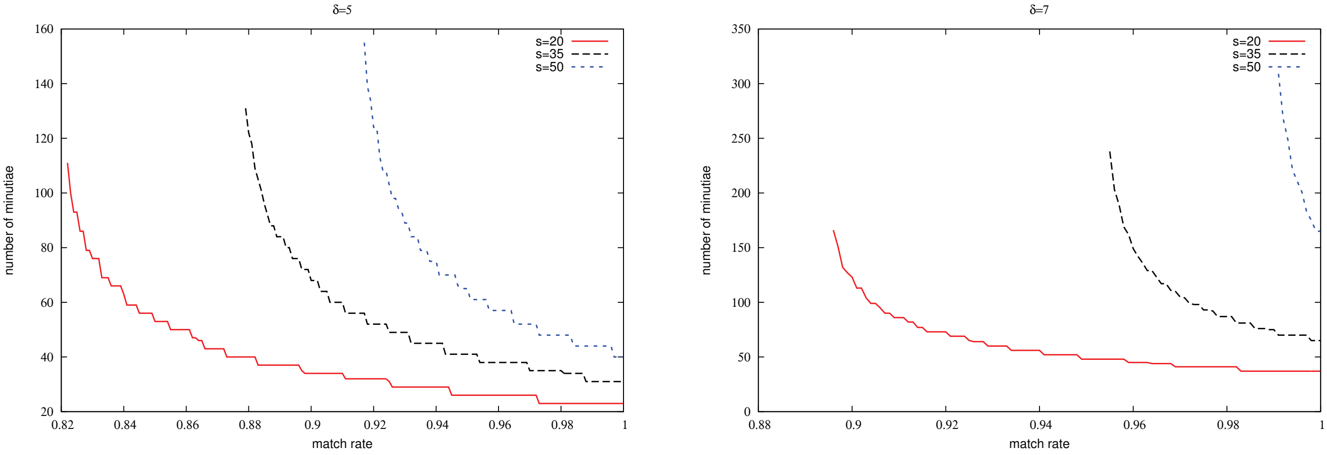

Figure 2. Dependency of the minimal number t of minutiae on the average match rate μ for a security of $E \geqslant 2^{50}$, for (a) $\delta = 5$ and (b) $\delta = 7$, respectively, and different numbers s of surplus minutiae per finger.

$k = k_0(r) + \epsilon$ and a match rate μ can be estimated as the value of t indicated in Fig. 2 for $\mu + \epsilon/t_0$, where t_0 is the value of t indicated in Fig. 2 for μ.

VI. Conclusions

Our analysis shows that a provably secure fuzzy fingerprint vault is not easy to achieve in practice.

- The minimum match rates required are quite high. According to [3], matchings conducted by a human expert results in rates of approximately 90%. Automatic matching algorithms only operating on the minutiae data will yield considerably lower rates. For instance, the distribution of the distance of matching minutiae reported in literature (see [17]) implies that a tolerance below 10 pixels distance will significantly reduce the match rate.
- The match rates can be greatly enhanced by applying a quality filter for the minutiae during enrollment. However, this reduces the number of minutiae available for the biometric feature vector and hence the maximum value of the parameter t.
- The minimum number of minutiae needed for a provable secure scheme is considerably higher than the numbers used in previous implementations. If quality filtering is applied during enrollment to improve match rates, we should not expect more than 30 reliable minutiae per finger. Thus, more than two fingers must be used.
- Provable security can only be achieved at reasonable match rates if the number of false matches, i.e. matches with chaff points, is effectively reduced. Therefore, it is imperative to apply minutiae quality filtering during authentication to reduce the number of surplus (non-matching) minutiae in the query fingerprint. However, the filtering may reduce the match rate and needs to be adjusted carefully.
- The storage of additional helper data (e.g., singular points, ridge density, minutiae orientations) could surely improve minutiae match rates. However, this data potentially leaks information on the feature vector and thereby decreases the remaining entropy. (This is particularly true for minutiae orientations that bear strong dependencies with their positions.) A solution could be to apply another layer of biometric encryption to protect this data, as suggested in [18], but the amount of information leaked by this layer should be carefully analyzed.
- The security can be increased by choosing higher values for k. However, this may render the decoding infeasible in many cases, resulting in higher false rejection rates. The maximum value of k yielding acceptable error rates should be carefully fine tuned on the basis of practical tests.

As a summary, it is questionable whether a provably secure fuzzy fingerprint vault can be achieved given the limited minutiae detection reliability of current technology. The minimum match rates and number of reliable minutiae needed for this goal imply that several fingers must be used, multiple concerted optimizations need to be utilized and a very high quality of the fingerprint images must be ensured.

Nevertheless, it is important to note that our analysis is based on theoretical lower bounds for security that are far from being tight. Practical attacks are not able to exploit all information revealed by the helper data (in particular, the y_i values) and it is questionable if they will ever be. The underlying computational problem of the fuzzy vault scheme (Reed-Solomon decoding) is believed to be hard for $k+1 < t < \sqrt{rk}$, and it is known that random instances of this problem are as hard as the worst case, see [19]. (For very large fields sizes, it is known to be NP-hard, see [20].) Therefore, practical security may be achieved for a fuzzy fingerprint vault scheme even though we are not able to prove it in terms of information theory.

Finally, provably secure biometric template protection schemes may still be achievable using completely different constructions. For instance, there exist approaches to apply the fuzzy commitment scheme to fingerprints, e.g., [21], [22] or [23]. As shown in [12], the entropy loss in the fuzzy commitment is much lower than in the fuzzy vault. However, we are not aware of any comprehensive security analysis for these approaches based on estimations for the feature vector's entropy and the error correction required without manual alignment of the fingerprints.

Acknowledgments

This work was conducted as part of the project "BioKeyS Pilot-DB" of the Bundesamt für Sicherheit in der Informationstechnik.

We would like to thank Patrick Jäger for his assistance in the numerical calculations and visualization.

References

[1] J. Breebaart, C. Busch, J. Grave, and E. Kindt, "A reference architecture for biometric template protection based on pseudo identities," in *Proc. BIOSIG 2008*, ser. LNI, vol. 137. Gesellschaft für Informatik, 2008, pp. 25–38.

[2] A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proc. IEEE Int. Symp. Inf. Theory*, 2002, p. 408.

[3] U. Uludag, S. Pankanti, S. Prabhakar, and A. Jain, "Biometric cryptosystems: Issues and challenges," *Proc. IEEE*, vol. 92, no. 6, pp. 948–960, 2004.

[4] C. Clancy, N. Kiyavash, and D. Lin, "Secure smartcardbased fingerprint authentication," in *WBMA '03: ACM SIGMM workshop on Biometrics methods and applications*, 2003, pp. 45–52.

[5] U. Uludag, S. Pankanti, and A. Jain, "Fuzzy vault for fingerprints." in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA)*, ser. LNCS, vol. 3546. Springer, 2005, pp. 310–319.

[6] U. Uludag and A. Jain, "Securing fingerprint template: Fuzzy vault with helper data," in *IEEE Workshop on Privacy Research In Vision*, 2006, pp. 163–169.

[7] K. Nandakumar, A. Jain, and S. Pankanti, "Fingerprint-based fuzzy vault: Implementation and performance." *IEEE Trans.Inf. Forensics Security*, vol. 2, no. 4, pp. 744–757, 2007.

[8] E.-C. Chang, R. Shen, and F. W. Teo, "Finding the original point set hidden among chaff," in *Proc. ACM Symp. on Information, Computer & Communication Security (ASIACCS)*, 2006, pp. 182–188.

[9] P. Mihailescu, A. Munk, and B. Tams, "The fuzzy vault for fingerprints is vulnerable to brute force attack," in *Proc. BIOSIG 2009*, ser. LNI, vol. 155. Gesellschaft für Informatik, 2009, pp. 43–54.

[10] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data." *SIAM J. Computing*, vol. 38, no. 1, pp. 97–139, 2008.

[11] R. Plaga, "Biometric keys: Suitable uses and achievable information content," *Int. J. Inf. Secur.*, vol. 8, no. 6, pp. 447–454, 2009.

[12] U. Korte, M. Krawczak, J. Merkle, R. Plaga, M. Niesing, C. Tiemann, H. Vinck, and U. Martini, "A cryptographic biometric authentication system based on genetic fingerprints," in *Proc. Sicherheit 2008*, ser. LNI, vol. 128. Gesellschaft für Informatik, 2008, pp. 263–276.

[13] S. Pankanti, S. Prabhakar, and A. Jain, "On the individuality of fingerprints," *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, p. 805, 2001.

[14] Y. Zhu, S. C. Dass, and A. Jain, "Statistical models for assessing the individuality of fingerprints." *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3-1, pp. 391–401, 2007.

[15] J. Chen and Y. S. Moon, "A statistical study on the fingerprint and minutiae distribution," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2006, pp. 169–172.

[16] C. Watson, M. Garris, E. Tabassi, C. Wilson, M. McCabe, S. Janet, and K. Ko, *User's Guide to NIST Biometric Image Software (NBIS)*, National Institute of Standards and Technology, 2007.

[17] A. Jain, S. Prabhakar, and S. Pankanti, "On the individuality of fingerprints," in *Proc. IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. I:805–812.

[18] A. Nagar, K. Nandakumar, and A. Jain, "Securing fingerprint template: Fuzzy vault with minutiae descriptors," in *Proc. Int. Conf. on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[19] A. Kiayias and M. Yung, "Cryptographic hardness based on the decoding of reed-solomon codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2752–2769, 2008.

[20] V. Guruswami and A. Vardy, "Maximum-likelihood decoding of reed-solomon codes is NP-hard," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2249–2256, 2005.

[21] P. Tuyls, A. Akkermans, T. Kevenaar, G. J. Schrijen, A. Bazen, and R. Veldhuis, “Practical biometric authentication with template protection.” in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA)*, ser. LNCS, vol. 3546. Springer, 2005, pp. 436–446.

[22] S. Draper, A. Khisti, E. Martinian, A. Vetro, and J. Yedidia, “Using distributed source coding to secure fingerprint biometrics,” in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007, pp. II–129–II–132.

[23] H. Xu, R. Veldhuis, T. Kevenaar, A. Akkermans, and A. Bazen, “Spectral minutiae: A fixed-length representation of a minutiae set,” in *Proc. IEEE Comp. Soc. Workshop in Biometrics*, 2008, pp. 1–6.