POSTER: Biometric Data Transformation for Cryptographic Domains and its Application

Shoukat Ali
shoukatali@fau.edu
Florida Atlantic University
Boca Raton, Florida

Koray Karabina
kkarabina@fau.edu
Florida Atlantic University
Boca Raton, Florida

Emrah Karagoz
ekaragoz2017@fau.edu
Florida Atlantic University
Boca Raton, Florida

ABSTRACT

A large class of biometric template protection algorithms assume that feature vectors are integer valued. However, biometric data is generally represented through real-valued feature vectors. Therefore, secure template constructions are not immediately applicable when feature vectors are composed of real numbers. We propose a generic transformation and extend the domain of biometric template protection algorithms from integer-valued feature vectors to real valued feature vectors. We show that our transformation is accuracy-preserving and verify our theoretical findings by reporting the implementation results using a public keystroke dynamics dataset.

CCS CONCEPTS

• **Security and privacy** → **Biometrics**; *Authentication.*

KEYWORDS

biometric security, template protection, keystroke dynamics

ACM Reference Format:
Shoukat Ali, Koray Karabina, and Emrah Karagoz. 2019. POSTER: Biometric Data Transformation for Cryptographic Domains and its Application. In *WiSec '19: ACM Conference on Security and Privacy in Wireless and Mobile Networks, May 15–17, 2019, Miami, FL, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3317549.3326302

1 INTRODUCTION

Biometrics-based cyber security technologies offer significant advantages in authentication, identification, and access control mechanisms. The popularity of biometrics technologies and their worldwide deployment makes biometric applications and databases natural targets in cyber attacks on a large scale. Since 1994 [2, 11], there have been tremendous research and development efforts for creating secure biometric schemes. In the most general terms, we can classify biometric template protection methods under three main categories: biometric cryptosystems (BC), cancelable biometrics (CB), secure multiparty computation based biometrics (also known as keyed biometrics) (SC, or KB), and hybrid biometrics (HB).

WiSec '19, May 15–17, 2019, Miami, FL, USA

ACM ISBN 978-1-4503-6726-4/19/05.
https://doi.org/10.1145/3317549.3326302

In BC and SC (whence in HB), cryptographic functions and transformations are the main tools to create secure templates. By construction, the underlying cryptographic primitives are defined over some particular discrete domains, and therefore, feature vectors are supposed to be some binary, or integer-valued vectors. More generally, a large class of template protection algorithms tend to assume that feature vectors are integer valued, and the similarity scores are calculated based on Hamming distance, set difference distance, or edit distance; see [8, 9]. However, biometric data is generally represented through real-valued feature vectors as in the case of face recognition [5, 6, 10] and keystroke dynamics [1, 3, 4, 7]. Therefore, many of the known secure template constructions, including the examples given above, would not be immediately applicable when feature vectors are composed of real numbers.

2 AN ACCURACY-PRESERVING TRANSFORMATION

In this section, we present our method to extend the domain of biometric template protection algorithms from integer-valued feature vectors to real-valued feature vectors. We also derive some theoretical estimates on the accuracy-preserving properties of our construction.

Definition 1 (The Scale-then-Round transformation (StR_s)). *For a real-valued vector $x = (x_1, x_2, \ldots, x_n)$, the map $\mathsf{StR}_s : \mathbb{R}^n \to \mathbb{Z}^n$ is defined as*

$$\mathsf{StR}_s(x) = (\lfloor sx_1 \rceil, \lfloor sx_2 \rceil, \ldots, \lfloor sx_n \rceil)$$

where s is a positive real number and $\lfloor \cdot \rceil$ is the nearest integer function.

Let GenP and ImpP denote the list of genuine and impostor pairs, respectively. Corresponding to these lists, let GenP′ and ImpP′ denote the lists of transformed version of GenP and ImpP, respectively, defined as

$$\mathsf{GenP}' = \{(\mathsf{StR}_s(x), \mathsf{StR}_s(y)) : (x, y) \in \mathsf{GenP}\}$$
$$\mathsf{ImpP}' = \{(\mathsf{StR}_s(x), \mathsf{StR}_s(y)) : (x, y) \in \mathsf{ImpP}\}$$

Next, we provide some theoretical estimates on the new system's False Accept Rate (FAR) and False Reject Rate (FRR) as a function of the original system's error rates. Note that, for a distance function d on $\mathbb{R}^n$ and $t \in \mathbb{R}^+$, we can write down the error rates of the original and the new system as follows:

$$\mathsf{FAR}(t) = \frac{\#\{(x, y) \in \mathsf{ImpP} : d(x, y) \leq t\}}{\#\mathsf{ImpP}},$$
$$\mathsf{FRR}(t) = \frac{\#\{(x, y) \in \mathsf{GenP} : d(x, y) > t\}}{\#\mathsf{GenP}},$$
$$\mathsf{FAR}'(T) = \frac{\#\{(X, Y) \in \mathsf{ImpP}' : d(X, Y) \leq T\}}{\#\mathsf{ImpP}'},$$

Table 1: The FRR and FAR values for the subjects *s055* and *s049* at three threshold points by computing the MD using original feature vectors. The transformed feature vectors at EER threshold using the (selected) scalars $s = 93$ and 3100.

	s055				*s049*			
	Original			**Transformed**	**Original**			**Transformed**
Threshold	$t_L = 1.177$	$t = 1.510$	$t_R = 1.843$	$\lfloor 93 \times t \rceil = 140$	$t_L = 6.387$	$t = 6.72$	$t_R = 7.053$	$\lfloor 93 \times t \rceil = 625$
FRR	0.095	0.010	0.005	0.010	0.620	0.480	0.255	0.485
FAR	0.008	0.012	0.024	0.012	0.392	0.480	0.584	0.484
Threshold	$t_L = 1.500$	$t = 1.510$	$t_R = 1.520$	$\lfloor 3100 \times t \rceil = 4681$	$t_L = 6.71$	$t = 6.72$	$t_R = 6.73$	$\lfloor 3100 \times t \rceil = 20832$
FRR	0.010	0.010	0.010	0.010	0.485	0.480	0.470	0.480
FAR	0.012	0.012	0.012	0.012	0.480	0.480	0.484	0.484

$$\mathsf{FRR}'(T) = \frac{\#\{(X,Y) \in \mathsf{ImpP}' : d(X,Y) > T\}}{\#\mathsf{ImpP}'}.$$

Our first result is the following theorem that assures the existence of a scalar s that can be used to obtain a new biometric system that now takes integer-valued vectors as input, deploys the same d in its matching algorithm, and runs at false accept rate $\mathsf{FAR}'(st)$ and false reject rate $\mathsf{FRR}'(st)$ that are arbitrarily close to $\mathsf{FAR}(t)$ and $\mathsf{FRR}(t)$ of the original system. We omit the proof due to space constraints.

Theorem 2.1. *Let d_p be the Minkowski distance defined on $\mathbb{R}^n$, and let $X = \mathsf{StR}_s(x)$, $Y = \mathsf{StR}_s(y)$ as before. For a given $\epsilon > 0$, if a scalar s is chosen such that $s \geq n^{1/p}/\epsilon$, then*

$$\mathsf{FAR}(t-\epsilon) \leq \mathsf{FAR}'(st) \leq \mathsf{FAR}(t+\epsilon),$$
$$\mathsf{FRR}(t+\epsilon) \leq \mathsf{FRR}'(st) \leq \mathsf{FRR}(t-\epsilon).$$

3 APPLICATION OF THE NEW TRANSFORMATION

In this section, we evaluate our theoretical findings over publicly available keystroke dynamics dataset [7]. Our reasoning for choosing these datasets is that it is widely referenced in the literature, and the feature vectors are represented as real-valued vectors.

To verify and support the theoretical bounds in Theorem 2.1, we use Manhattan Distance (MD) to determine the distance between two biometric feature vectors. After computing the error rates in the dataset, we select the two subjects that show minimum and maximum equal error rate (EER). The subject *s055* and *s049* have minimum and maximum EER, respectively. We provide the EER and the threshold values in Table 1. We also provide $t_L = t - \epsilon$ and $t_R = t + \epsilon$ where ϵ is determined according to Theorem 2.1 based on the length of the feature vector and selected scalar. In the upper half of the table 1, we have $t_L = t - (31/93) = t - 0.333$ and $t_R = t + (31/93) = t + 0.333$ where $s = 93$ is our scalar. In the lower half of the table 1, we have $t_L = t - (31/3100) = t - 0.01$ and $t_R = t + (31/3100) = t + 0.01$ where $s = 3100$ is our scalar. Furthermore, using the same evaluation technique, we compute the error rates for *s055* and *s049* using the transformation StR_s on the feature vectors. The error rates at transformed EER thresholds with the scalars 93 and 3100 are provided in Table 1. Clearly, all the computed values agree with the conclusion of Theorem 2.1.

4 CONCLUSION AND FUTURE WORK

We proposed a generic method to extend the domain of biometric template protection algorithms from integer-valued feature vectors to real-valued feature vectors. We prove that our method is accuracy-preserving in the sense that the accuracy of the new system can be made arbitrarily close to the accuracy of the original system. This allows real-valued feature vectors to be used as input to some cryptographic algorithms, whence to enhance the security of the matching algorithms while preserving the accuracy rates of the original (non-cryptographic) systems. Therefore, our next objective is to implement template protection algorithms in combination with our proposed method over a large class of biometric data.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation (award number 1718109). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] Salil P Banerjee and Damon L Woodard. 2012. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research* 7, 1 (2012), 116–139.
[2] A. Bodo. 1994. Method for producing a digital signature with aid of a biometric feature. *German patent DE* 42, 43 (1994), 908.
[3] Patrick Bours and Hafez Barghouthi. 2009. Continuous authentication using biometric keystroke dynamics. In *The Norwegian Information Security Conference (NISK)*, Vol. 2009.
[4] Michael Fairhurst and Márjory Da Costa-Abreu. 2011. Using keystroke dynamics for gender identification in social network environment. (2011).
[5] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. 2007. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.* Technical Report 07-49. University of Massachusetts, Amherst.
[6] Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. 2009. Cancelable iris biometrics and using error correcting codes to reduce variability in biometric data. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE, 120–127.
[7] Kevin S Killourhy and Roy A Maxion. 2009. Comparing anomaly-detection algorithms for keystroke dynamics. In *Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP International Conference on.* IEEE, 125–134.
[8] K. Nandakumar and A. Jain. 2015. Biometric template protection: Bridging the performance gap between theory and practice. *IEEE Signal Processing Magazine* 32 (2015), 88–100.
[9] I. Natgunanathan, A. Mehmood, Y. Xiang, G. Beliakov, and J. Yearwood. 2016. Protection of Privacy in Biometric Data. *IEEE Access* 4 (2016), 880–892.
[10] Christian Rathgeb, Frank Breitinger, Christoph Busch, and Harald Baier. 2014. On application of bloom filters to iris biometrics. *IET Biometrics* 3, 4 (2014), 207–218.
[11] G. Schmidt, C. Soutar, and G. Tomko. 1996. Fingerprint controlled public key cryptographic system (1996). *US Patent* 5541994 (1996).