GenKey BioHASH® Technology vs. Biohashing

Introduction

You are interested in Privacy-by-Design biometrics, and you want to know more. Your first act is typically to do an online search. You have a certain understanding of cryptographic functions, so during your quest you find the terms BioHASH® and Biohashing. A hash function is a non-reversible transformation, so it stands to reason that a Biometric Hash could be useful to prevent reverse engineering of protected biometric templates. This leads you to explore further and then the confusion might hit you, because the various articles claim completely different properties for the technology. This note will give you a deeper understanding of the differences and properties of two fundamentally different technologies for biometric protection.

GenKey BioHASH® vs. Biohashing

While GenKey BioHASH® technology and the technique commonly known as Biohashing may sound similar by name, the underlying technologies and applications are quite different. This paper provides a brief description of Biohashing and how it differs from GenKey BioHASH®. For more information regarding BioHASH®, please refer to "*Genkey BioHASH® - a Privacy-by-Design approach to biometrics*" available on the GenKey website at www.genkey.com.

What is Biohashing?

Biohashing is a technique that has been proposed and studied by various researchers and published in a variety of technical publications. The technique has been applied to various forms of biometrics such as face, finger and palm verification. The general idea of Biohashing is shown in Figure 1, which is to map a set of native biometric features $F=\{f_1, f_2, \ldots f_N\}$ to a bit code $B=\{b_1, b_2, \ldots b_N\}$ where the mapping function $M_u()$ is random and derived from a token R_u that is random and unique to each individual. To compare a new sample of a user's biometric $F'=\{f'_1, f'_2, \ldots f'_N\}$, the same token R_u is supplied which allows the same mapping function $M_u()$ to be performed and the resulting bit code $B'=\{b'_1, b'_2, \dots b'_N\}$ can be compared for binary similarity with the enrollment bit code B. In general, the mapping function consists of a user-specific orthonormal transformation T_u followed quantization Q.

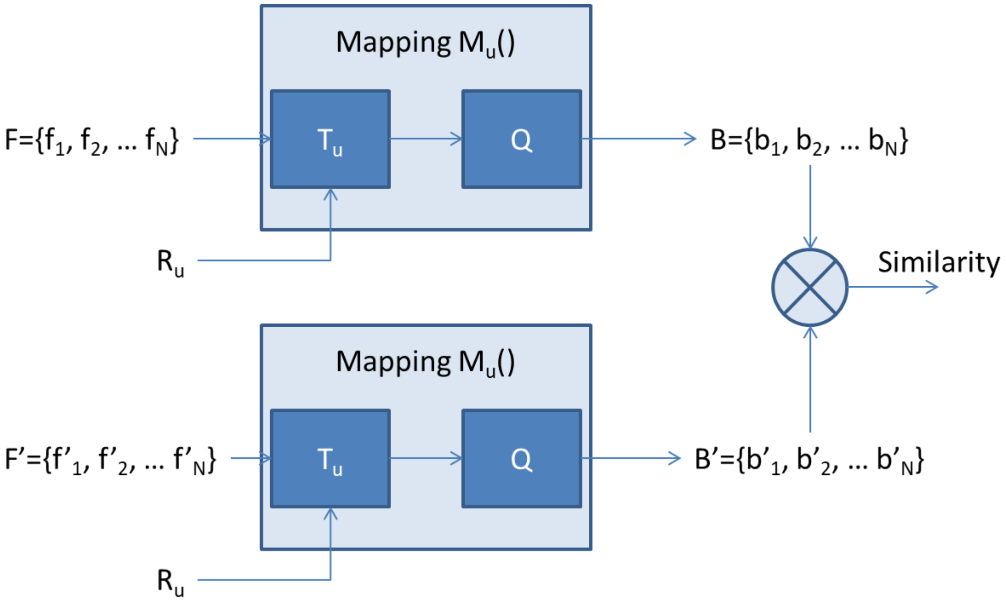
Mapping $M_u()$
$F=\{f_1, f_2, \dots f_N\}$
T_u
Q
$B=\{b_1, b_2, \dots b_N\}$
R_u
Similarity
Mapping $M_u()$
$F'=\{f'_1, f'_2, \dots f'_N\}$
T_u
Q
$B'=\{b'_1, b'_2, \dots b'_N\}$
R_u

Figure 1 Enrollment/Verification with Biohashing

One of the main claims of Biohashing is that it can achieve perfect accuracy even when the underlying biometrics have natural tendencies to produce false acceptance and false rejection errors. While Biohashing may be able to reach perfect accuracy, this accuracy gain over the native biometric accuracy can be shown to be due solely to the uniqueness of the user token R_u. In fact, as the feature length N increases, the importance of the biometric feature accuracy decreases thus allowing Biohashing to achieve perfect accuracy even with relatively low quality biometric algorithms.

While this may seem to provide an advantage to Biohashing, the primary contribution to the accuracy and security comes from the token itself. To achieve perfect accuracy the token itself must be strong enough to already serve as a unique identifier for the user so the biometric contribution to verification of the user is very low or even unnecessary. Additionally, achieving perfect accuracy requires an assumption that the token is never compromised as a compromised token will yield an accuracy level equivalent to the underlying biometric algorithm.

The key aspect of Biohashing is a user specific randomized orthonormal transformation matrix that is derived from the user's token. Any NxN orthonormal transformation matrix effectively produces a rotation of the N dimensional coordinate axis system about the origin. This is shown for N=2 in Figure 2 where different users produce different rotations T(R_u) of the coordinate axis system. Note that since this is rotation of the entire feature space, the Euclidian distance properties between biometric samples are preserved so samples that are close in the original feature space are also close in the transformed feature space.

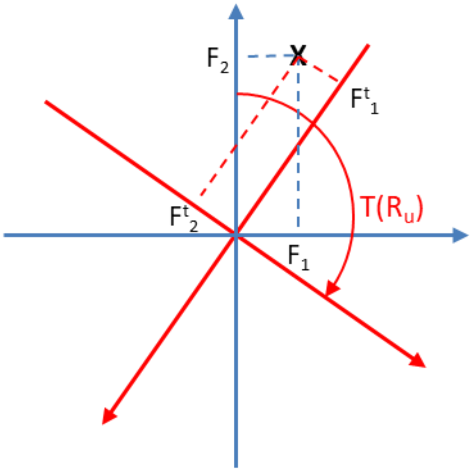

Figure 2 Orthonormal Transformation N=2

While the uniqueness of the rotation is somewhat limited in two dimensions, the amount of rotational freedom increases exponentially with the feature dimension N. Since a genuine verification will always have the correct token provided by definition, the only risk of a false rejection occurs when the similarity threshold is higher than the similarity threshold being used. However, for impostors the risk of false acceptance is related to both the similarity threshold and having access to the correct token. The number of unique tokens is limited to 2^N where N is the feature dimension, so as N increases the importance of the similarity threshold is reduced since the probability of presenting a matching token decreases. It should be clear that for an impostor to be falsely accepted for any reasonable value of N, obtaining a matching token is far more important than presenting a similar biometric feature set. This of course means that protection of the token is key factor in the security of a Biohashing system.

How is GenKey BioHASH® different?

With Biohashing, each user has a token that is required during the verification process. Much like a standard encryption key, the user token must be stored somewhere such as a database or a smart card and be retrieved whenever verification is required. By contrast, BioHASH® does not require any key storage or key management thus providing a self-contained and secure verification process that can be used in either online or offline verification applications. One could say that BioHASH® keys are generated on-the-fly during biometric verification leading to simplified system designs. One can build an asymmetric verification system using public/private key pairs without the need for certificates or any PKI infrastructure.

Unlike BioHASH®, Biohashing does not provide a stable numeric code so it is primarily suited for verification applications where biometric variations can be measured and a matching decision can be made on the basis of a matching threshold. BioHASH® creates a stable numeric code using various noise stabilization techniques such as noise reduction and error correction. The resulting stable code can be used for making verification decisions similar to Biohashing, but it can also be used to generate repeatable encryption keys and PIN codes.

BioHASH® does not claim perfect accuracy, however any such claim from a biometric system should be considered carefully as it is an unrealistic expectation of any biometric system, which is in essence a statistical comparison of properties that have a finite entropy. Nevertheless, the accuracy of BioHASH® has evolved over the years to the point where it is very practical in many real-life use cases. In a test on the public database FVC-2006 DB2 it has achieved FRR=1% and FAR=0.0004% for a single finger, purely based on the biometric data.

In the case of Biohashing, the claimed increased accuracy is a function of the strength and uniqueness of the token itself. If a use case calls for the use of a token, it is more powerful to use it in a two-factor authentication scheme to increase overall system robustness.

For more information on the benefits of BioHASH® technology, please refer to "*Genkey BioHASH® - a Privacy-by-Design approach to biometrics*" available on the GenKey website at *www.genkey.com*.