GEN KEY
Identity for all

GenKey BioHASH®

A radically different approach to biometrics

GenKey Netherlands B.V. has its statutory seat in Eindhoven, the Netherlands and is registered at the Dutch Chamber of Commerce under number 32132038.

The information presented in this document is subject to change without notice. GenKey assumes no responsibility for any errors that may appear in this document. This document may contain links to third-party websites that are not under the control of GenKey and GenKey is not responsible for the contents of any linked site or any link contained in a linked site, or any changes or updates to such sites.

Contents

A Introduction

A1 Protection of biometric information

Biometrics is a highly convenient and secure approach to verifying the identity of individuals. Biometrics such a as fingerprints cannot be forgotten and it is extremely hard to lend a fingerprint to another individual. Biometrics also give evidence that an individual is physically present at the time of the verification.

All biometric verification systems store biometric reference information in the form of biometric images or biometric templates. During verification a live biometric measurement is compared with the stored biometric reference information and a decision is made if the finger used to generate the stored reference information is the same as the offered live finger.

Due to its very nature, biometric information is sensitive and private and should be treated in a highly secure and anonymous manner. The stored biometric reference information can be protected using traditional encryption, however, this has an inherent vulnerability in that templates need to be decrypted to perform a biometric comparison making it temporarily available in cleartext during comparison and is then vulnerable to attack. The same is true for the live measurement: it is also important to prevent the live biometric measurement from being compromised in the verification process.

This brochure describes a fundamentally different approach to storing and comparing biometric information. We call this approach BioHASH® and its philosophy and implementation is based on the principle of "Privacy-by-Design".

A2 Handling biometric variability

A fundamental property of biometrics is variability: a fingerprint imprint (or any other biometric modality) is never exactly the same as the previous one, even if it stems from the same finger[1]. Conventional biometric solutions compensate for this variability at match-time, by accepting certain tolerances on the measurements. GenKey's BioHASH® is a radically different approach to biometrics in that it generates a reproducible, stable code from the varying biometrics.

- This stable BioHASH® code can be used in many different innovative ways, enabling biometrics to become an integral part of many security protocols. The many possible use cases are described in chapter B.
- The most basic application is that reproducible stable biometric code can be protected by hashing, following the ISO/IEC 24745 standard for Biometric Information

[1] With fingerprints, this variability can be caused by the way you put your finger on the scanner, environmental conditions like temperature and moisture, and by the amount of pressure applied on the sensor. Quite often, different sensors are being used for enrolment and verification.

Protection, providing benefits for secure storage, verification and authentication. This is described in more detail in chapter C.

BioHASH® was able to solve the challenge of deriving a stable code from ever-varying biometric input. This has been made possible by borrowing heavily from signal processing techniques such as extraction of stable components, noise reduction, error correction, etc. The stable code generated by BioHASH® or the hash value of the code can also be used as a seed in standard cryptographic operations, as a PIN code, and even in standardized client-server protocols. A list of use case options is described in the following section.

BioHASH® is not just an academic exercise. It is in practical use with millions of identity cards issued and thousands of transactions every day, as described under B5.

B BioHASH® Use Cases

In this section various application examples of BioHASH® are listed, ordered from the simplest to the more sophisticated use cases.

B1 Protecting biometric information

BioHASH® protects biometric information in the same way as PIN codes and passwords are protected in that a BioHASH® template is in essence the cryptographic hash of the biometric information represented as a reproducible stable code. Because BioHASH® is a keyless approach, there is no need for key storage and key management, such as using certificates in a PKI infrastructure.

At the process level, using BioHASH® in a biometric application is identical to using a regular non-protected biometric solution: there is enrolment, storage and verification. The biometric capture is transformed into a BioHASH® code and stored on a card, on-line or in a database. For a biometric verification, a live biometric probe is captured, hashed, and compared with the stored BioHASH® code in the hashed domain, again very similar to the way PIN codes and passwords are verified. Chapter C describes this use case in more detail.

B2 Small and secure template allows for low-cost storage media

The size of a BioHASH® template can be as small as a few hundred bytes. We call this format BioHASH®-C (For Classic). This makes it possible to store the information in standard 2D barcodes that can be printed on paper or small plastic cards.

A special category that can benefit from the intrinsic security of the stored templates, is when BioHASH® is printed on paper documents, like entry tickets, boarding passes or Visa inlays in passports. For these applications, there is no known biometric alternative that is as flexible, private, and cost effective. Once you have captured the biometrics, all you need is a normal printer!

During verification, the barcode is scanned, and the BioHASH® template is compared with a live measurement of the fingerprint.

For more demanding applications, slightly larger templates can be used, a few KBytes in size. These can be stored in the embedded memory of low-cost contactless ID cards, and they can be personalised (biographics, photo) using standard off-the-shelf card printers. Additional security features (like ID card number or personal information) can be included in the hash calculation and stored as part of the BioHASH® template. This protects also this additional information from tampering.

The larger templates, that we call BioHASH®-D (for Dual Layer), are also the basis for the fully private Client/Server biometric verification protocols described in B7, and for the creation of symmetric keys or asymmetric key pairs. The secret symmetric and private keys are never

stored, but regenerated on-the-fly from the live biometric only when a biometric verification is performed.

B3 BioHASH® to generate a PIN code from a finger

The stable code that is reproduced every time a succesfull verification against a BioHASH® template is done, can be used as a biometric PIN code. It can be used instead of or be combined with a regular PIN code to, for example, authenticate a money transfer from a mobile phone. It can also be used for people who tend to forget their PIN code or people who are not used to PIN codes in general. For this use case we recommend to use BioHASH®-D. Effective supported PIN code length for a single finger is 5 digits, typical for most banking-type applications. Longer codes can be supported when two fingers are used.

B4 BioHASH® as an encryption key

A slight extension of the concept of using BioHASH® as a PIN code, is to use it as an encryption key. This makes it possible to encrypt personal data with a key derived from this personal data (self-encryption). This enables much simpler system designs for protecting (biometric) information that need not rely on external keys, which are traditionally handled with certificates and PKI infrastructures. An obvious opportunity to exploit this is in multimodal biometrics, where conventional templates from one of the modalities can be protected by biometric encryption based on another modality. Both BioHASH®-C and BioHASH®-D support this use case.

B5 A biometric Proof-of-Presence

When there is a succesfull verification of a person against a BioHASH® code, exactly the same stable code is generated. This allows BioHASH® to generate a Proof-of-Presence by using this stable biometric code to generate a signature on the digest of a document. Since the correct signature can only be generated after a succesfull biometric verification, such a signature guarantees that the person was present. For this use case we recommend to use BioHASH®-D, although, as show in the example below, also BioHASH®-C can be used, albeit for slightly less-demanding cases.

A practical use case

In Ghana, the NHIA (National Healthcare Insurance Authority) has engaged with GenKey to combat fraud by issuing biometric Identity Cards for their clients containing BioHASH®. First objective was to ensure that members that carry Health Insurance cards identify themselves biometrically against that card, so that an insurance card cannot be abused by other persons. If an insured patient visits a healthcare facility, a succesfull verification of the patient against his card will also generate a Proof-of-Presence in the form of a so-called CVC (Claim Verification Code). This code includes references to the healthcare provider location, the time and day of treatment and it provides evidence that there has been a succesfull biometric

verification. The CVC is manually copied on the claim form that gets submitted to the insurer, who can independently verify the validity of the CVC. This addresses the issue of phantom claims by some healthcare facilities that filed claims for patients they had never treated. That is why Genkey's motto for its health care activities is:

"Less Fraud = More Care".

At this moment there are around 17.5 million active participants in this program. Similar programs are now also running for Kenya's National Hospital Insurance Funds, with the objective to register 4 million people.

B6 BioHASH® for all modalities

Although fingerprint has been heavily used as an example, and is currently the modality with the largest deployment, BioHASH® is very well suited to other biometric modalities such as vein, iris, voice and face. The basic principle remains the same throughout. First the raw biometric capture is transformed into a BioHASH® code which can be used in any of the above applications.

B7 BioHASH® for use in cryptographic protocols

This brochure gives some examples of how BioHASH® could be used. All those examples up till now are concerned, in some shape or form, with a local verification. There are however many and flexible ways to incorporate biometrics in cryptographic protocols made possible by the stable, reproducable code representating the biometric. An example is fully private biometric authentication towards a (possibly untrusted) server. With BioHASH®-D it is possible to derive a public-private key pair from a biometric. By sending the public key to a server and deriving the corresponding private key on the client from a live biometric measurement, standard public-private key authentication protocols can be used to perform a biometric authentication, giving the server evidence that the person corresponding to his public key is physically present. Those standardized public-private key protocols are well understood and scrutinized as they are routinely used to authenticate computers towards eachother. This approach to biometric authentication towards a server automatically inherits all the properties of those standardized protocols such as no information leakage when the servers turns out to be untrusted. For this use case BioHASH®-D is required.

B8 Self-Sovereign Identities

Self-Sovereign Identities (SSIs) is a new approach to identities on-line. The main underlying idea is that the individual himself is in full control of all his identity attributes contained in an electronic wallet on a device owned and trusted by the individual. Those attributes can be issued and signed by a trusted authority such as a government, university or bank, or they can be generated by the individual himself. During authentication towards a service, the individual

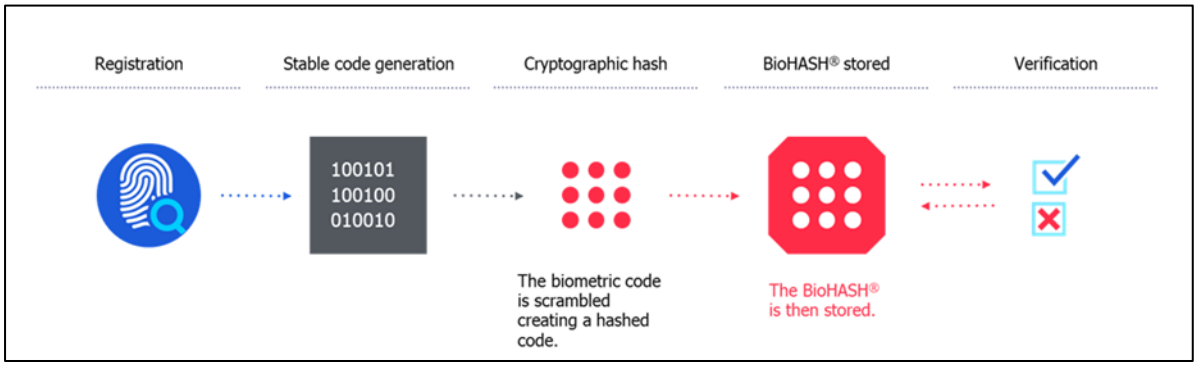
decides which information is revealed to the service and the wallet then interacts with the service revealing the minimum possible amount of information (often only a single bit). An example of an SSI initiative is Sovrin (www.sovrin.org).

BioHASH® allows standardized SSI schemes to be extended with biometric attributes adding an irrefutible and fully private proof of personal presence at the time of the interaction with the service. For these use cases BioHASH®-D is required.

C The simplest case in more detail: BioHASH® as a secure biometric template

C1 Stable codes and biometric hashing

One of the fundamental properties of a cryptographic hash function is the fact that even the smallest change in the input will generate a completely different hash output. This property is routinely used for tamper detection of documents or computer programs often in combination with digital signatures to detect even minute changes. Also, Blockchain technology builds on this notion by forming a distributed chain of hashes that becomes a ledger that cannot be modified or tampered with. A more basic application for using hash functions is to protect information. For example, the storage of PIN codes and passwords are protected by hash functions. Due to the properties of the hash functions, when there is even a single typo in a password or PIN code, verification will fail. When protecting PIN codes this way, this is of course a desirable property, but when applying a hash to biometrics, the variability over different biometric samples of the same finger creates a problem. The stability that is created by BioHASH® resolves this issue, and enables properties that are not readily available in conventional biometrics, like revokability and diversification across multiple applications.



Cryptographic One-Way functions

In our daily environment we see many situations that are "one-way" by nature, meaning that a transformation in one direction is very common, but a reverse transformation is very rare or even impossible. If you drop a glass and it shatters in pieces, it is very hard to recreate the glass from the fragments, although you could consider melting all of the pieces to re-create a glass from the melt. In other cases the reverse direction can be completely impossible: when a cigarette gets burned, it is physically impossible to recreate the original from the smoke and the ashes. What is however possible, is to collect some smoke and/or ashes and run an analysis, e.g. with a mass spectrometer.

In this way you can characterize the actual variety of tobacco that was used, even though the cigarette no longer exists.

In cryptography there is an important class of algorithms that have the same One-Way functionality as described above. It is easy to convert an input to an output, but virtually impossible to reverse-engineer the input based on the output. The best known and most widely used form of this is a cryptographic hash function. This is routinely used in the protection of PIN codes and passwords. BioHASH® uses exactly that approach and applies it to biometric data, and in the cigarette analogy only the characterizations are stored, never the original biometrics. That is what makes BioHASH® a truly second generation biometric approach, different from all traditional biometrics.

C2 BioHASH® protects biometric information

The stable code generated from the biometric by GenKey's BioHASH® can be used as input to a cryptographic hash function. This protects the code and any sensitive and personal information that might still be present in the stable code[2]. Furthermore, before hashing the code, additional random information is added making it possible to derive different hash values codes from the same biometric. This prevents coupling of different databases (function creep) and allows virtually unlimited revocation/renewal of identifiers.

[2] The stable code is already heavily abstracted and compressed from the original images

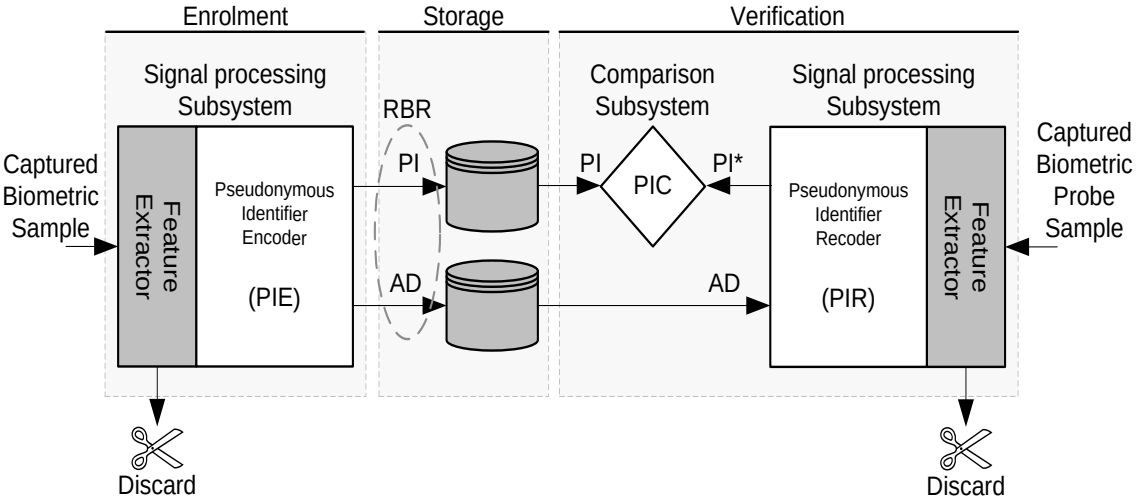
C3 BioHASH® complies with ISO/IEC 24745

The ISO/IEC 24745 standard for Biometric Information Protection provides guidance for the protection of biometric information under the following privacy and security requirements:

Irreversibility and confidentiality: protect biometric information against unauthorized access or disclosure during storage and transmission.

Unlinkability: prevents biometric references being linkable across applications or databases.

Renewability and revocability: allows re-issuing biometric references while at the same time revoking the earlier biometric reference.



PI: Pseudonymous Identifier
AD: Auxiliary Data
PIC: Pseudonymous Identifier Comparator

The linkage between this standard picture and the actual BioHASH® implementation is as follows:

The feature extractor is a key element in BioHASH®, and together with PIE it creates a template. PIE adds elements like the random seed, an application ID and salting, plus redundancy to enable the use of error-correcting codes later on in the verification process. The template consists of two parts, the PI and the AD. Neither of them reveals any information about the biometric features that were used in creating them. In the simplest example, PI and AD are stored together[3]. During verification, the AD gets combined with a life probe, error correction takes care of any finger placement differences, and this produces a candidate PI*. This creates a binary match with the PI from the template, if the template and the probe came from the same finger. In this process of generating PI*, the randomness that was introduced in PIE cancels out.

[3] There are applications that can benefit from storing PI and AD separately, for instance in a client/server model. See description under B5.

D System considerations

It is obvious that BioHASH® has many benefits, so a natural question to ask is whether there are any concerns when applying this technology. The short answer is yes, and this chapter describes how GenKey tackles those concerns.

D1 Brute Force biometric attacks

Biometrics is a statistical science. The chosen operating point for any biometric matching is always a trade-off between the probability that the system accepts an imposter (FAR, False Accept Rate) and that the system rejects a genuine user (FRR, False Reject Rate). Typical range for FAR is between 1 in 10,000 and 1 in 100,000. So if an attacker is allowed to run a database of 100,000 entries against any form of protected biometric reference information, he is very likely to find a working biometric that will allow him to force entry. This needs to be prevented at system level. Examples of counter measures are:

- Limit the number of attempts. Just like many PIN-based systems, biometric systems can limit the number of attempts to 3.
- Increase the barrier against brute force entry. If a single finger can reach a FAR of 1 in 10,000, then 2 fingers combined could theoretically reach a FAR of 1 in 100,000,000. In reality this will be less because fingers from the same individual have properties that are more correlated, but in practice a barrier of 1 in 10,000,000 can be achieved.
- Multi-factor system design. Rather than uniquely relying on biometrics, combining it with a token/mobile phone (something you have) or PIN code or password (something you know) can significantly strengthen system security.

D2 Limited biometric entropy

Biometrics are claimed to be unique for every individual, but in reality there is a limit to the amount of differentiation that can be distinguished by biometric systems, or even by human observers. Many celebrities have body doubles that could impersonate them, and any individual biometric modality has “doubles” in the global population as well. The amount of differentiation is often referred to as entropy. This presents a natural limit to the size of the stable number that BioHASH® can derive from each modality. Unfortunately, typical sizes are in the order of 20 to 30 bits per single measurement, certainly not enough to be used directly as a cryptographic key where lengths of at least 128 bits are required[4]. So again, counter measures are required at the system level. Some examples:

[4] These numbers are for fingerprints. Face and voice are slightly worse, iris and vein are better, but none of them is good enough by itself to reach a strength that is deemed adequate by cryptographic standards

- Multi-modal and multi-finger biometrics. Multi-modal refers to the case where different modalities are combined (e.g. fingerprint and fingervein), and multi-finger refers to the case where 2 or 4 fingers are scanned.
- Stronger modalities. It is known that vein and iris have more entropy than voice, face or fingerprint.
- Multi-sample fingerprints. By taking two or more samples from the same finger at enrollment (which is a one-time event) and allowing a retry upon verification (only if the first attempt fails), we can substantially improve the entropy, and therefore the matching and security characteristics of both BioHASH®-C and BioHASH®-D.
- Salting. This is the cryptographic term when a secret with limited entropy (the biometrics) gets enhanced by a much stronger key length from a system-level secret. Together they create a combination that has a sufficient key length (mainly due to the system-level secret) and a high variability (a different secret is applied for each individual due to the biometrics). The variability makes the salted secret much more resilient to attacks compared to the traditional alternative (symmetric encryption) of applying the same secret to protect information from everyone system-wide. GenKey uses a WhiteBox implementation for the Salting process.
- Differentiation and randomness. When a template that is derived from a biometric property is always constant, it becomes more and more vulnerable as multiple samples are produced. When the template contains a random element for every new instance, attacking the template becomes much more complicated. This is the case both for BioHASH®-C and BioHASH®-D

D3 Reverse Engineering attacks.

Matching of BioHASH® templates is typically performed on a biometric verification device or server. No matter how secure the template may be, if an attacker can get hold of the live measurement when a user tries to verify, the biometrics for that user will have been compromised. So at system and device level, counter measures are necessary, such as:

- Shielding of some critical pieces of software. On mobile devices this is often done in ARM's TrustZone, on a PC a TPM (Trusted Platform Module) can be used, and for servers there is access control, TXT (Trusted eXecution Technology) and similar technologies.
- Protection against malware, trojans etc. should be part of every device and system environment.
- Dedicated measures against reverse engineering, run-time emulation and debugging for the core SDK that performs biometric processing. Genkey uses an external library from a reputable security firm to perform this function.

Genkey has been one of the inventors of protecting biometric information using hash functions and has a number of patents and many manyears of experience in designing well-balanced biometric solutions at the system level, and the resulting systems have been proven in mass-deployments.

E Summary and Conclusion

The ability to derive a stable code from ever-varying biometrics opens up many applications. These range from field-proven and standardized (ISO/IEC 24745) to speculative, such as using biometric Identities in a SSI/Blockchain environment, where currently nobody can predict where the ultimate exact sweet spot may reside. However, it is for sure that deployment for these technologies can best be done by people that have years of experience in the type of system level considerations that are relevant for this field. So please consult with our experts when you have an application idea that might benefit from BioHASH®.

About GenKey

We're experts in biometrics. We help millions of people in Africa, and other emerging markets, to identify who they are. Together with partners, we deliver large scale biometric identity programs, with broad experience in national elections and healthcare.

GenKey has roots in Biometric Privacy-by-Design technology. Our unique approach to storing and comparing biometric information has evolved from an academic proposition to a mature state which is applicable wherever the privacy and security of end-users' biometric data are important. The technology is protected by more than 10 patents.

Identity for All

info@genkey.com

www.genkey.com