Report to Seminar-Talk

A Formal Study of the Privacy Concerns in Biometric-Based Remote Authentication Schemes

held at

Bonn-Aachen International Center for Information Technology

Seminar “Biometry & Security”
supervised by: Laila El Amani, Deniz Sarier
responsible: Prof. Joachim von zur Gathen

Markus Hirtsiefer
2010

Contents

Introduction

The use of biometrics in remote authentication systems has become more and more popular. Biometrics are comfortable to use, provide uniqueness for free, can not be transferred and, unlike passwords, they do not suffer from being either insecure or hard to remember. While a lot of work considering the security of biometric based in remote authentication systems has already been published, the privacy issues of such systems are not so well researched yet. This is surprising because privacy is an important issue and as our world gets more and more connected and thus control over private data becomes difficult, privacy cautions become even more important, especially with biometrics which can not be exchanged or revoked easily in case of data leakage.

The paper [1] from Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval, which my seminar-presentation and this report are based on, declares to be one of the first which provides some formal study about the privacy concerns in biometric based remote authentication systems. The nice thing about formal studies is that they may provide formal proofs that certain achievements can be guarantied by the system.

First we will see a very basic overview over the system described in [1], then the privacy concerns to be achieved by the system will be defined, assumption about the system will be explained, next there is a description of the system procedures and finally we will see whether one can proof some privacy properties under the assumptions made.

Basic Overview

System Structure

The system Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval propose, mainly consists of four components: A User who wants to use some service that needs authentication. A Sensor-Client which captures raw biometric data and extracts some template information. The system design is not restricted to some specific kind of biometric data, the biometrics used only have to comply with the assumptions described below. Furthermore there is one or more service-provider which offers some service to the user, and in case of a authentication request queries the 4th component, a database which stores the biometric information. Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval tried to use as few components as possible while still keeping certain information separated in different locations. This separation of information is one of their main design

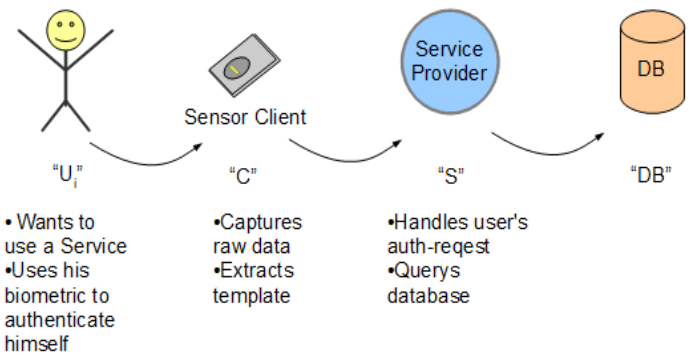

Figure 1: System Structure

ideas to provide the privacy desired and the way it is done strongly relates to the trust relationships which will be discussed in chapter assumptions.

Workflow

The proposed system supports two procedures: Enrollment and Verification. During enrollment new Users can register to the system with their biometric and a personal userid they may choose. When a registered user wants to use some service that needs authentication, a verification-process is instantiated: The user enters his userid and biometric at a sensor-client and the system checks whether we either have a validly registered user which is allowed to use the service or the authentication request has to be rejected. During all these operations we want some privacy issues to be concerned, which are described in the next chapter.

Privacy Concerns

To talk about privacy or even proof that a given system design is secure regarding some privacy concerns, the latter have to be defined precisely. Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval introduced the following properties which they want their system to guarantee:

Soundness and Impersonation Resilience

These two properties in essence mean that with high probability the system accepts the right users and rejects false ones. On the one hand of course this is an important privacy issue as we do not want user A to be able to log in into user B's email-account for example, but on the other hand it is primarily a security issue and in their paper [1] the designers tried to focus on pure privacy issues. They do give a formal definition for soundness, but they do not proof it. Impersonation Resilience is not proofed in detail either, but according to [1] it follows directly from soundness and the assumptions “Security Link” and “Liveness” which will be defined later in this report.

Definition 2. *A biometric-based authentication scheme is defined to be sound if it satisfies the following two requirements:*

1. *With an overwhelming probability, the service provider will accept an authentication request in the following case: sensor client sends (ID_i, b) in an authentication request, where $\mathsf{H}(b, b_i) \leq \lambda$ and b_i is the reference template registered for ID_i.*
2. *With an overwhelming probability, the service provider will reject an authentication request in the following case: sensor client sends (ID_i, b) in an authentication request, where $\mathsf{H}(b, b_i) > \lambda$ and b_i is the reference template registered for ID_i.*

Figure 2: Soundness Definition, copied from paper [1]

Identity Privacy

The primary privacy goal of the system proposed in [1] is to keep the link between a user's identity and his biometric-data secret, and this is exactly what is promised by Identity Privacy property. The biometric data itself was considered public information, which according to [1] is a usual assumption, the danger about using biometrics is not the information in the biometric data itself, but the fact that the data stays the same for each application which makes it is easy to track users over different applications, link usernames and perhaps to reveal some user's real identity.

Identity Privacy property guaranties that for any personalized username (Ids), an adversary knows nothing about the corresponding biometric template. Formally it is defined using kind of a Game:

Definition 3. *A biometric-based authentication scheme achieves identity privacy if* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ *has only a negligible advantage in the following game, where the advantage is defined to be* $|\Pr[e' = e] - \frac{1}{2}|$.

$$
\begin{array}{l}
\mathbf{Exp}_{\mathcal{A}}^{Identity\text{-}Privacy} \\
\left|\begin{array}{rl}
(i, ID_i, b_i^{(0)}, b_i^{(1)}, (ID_j, b_j)(j \neq i)) & \leftarrow \mathcal{A}_1(1^\ell) \\
b_i = b_i^{(e)} & \stackrel{R}{\leftarrow} \{b_i^{(0)}, b_i^{(1)}\} \\
\emptyset & \leftarrow \mathsf{Enrollment}(1^\ell) \\
e' & \leftarrow \mathcal{A}_2(Challenger; \mathsf{Verification})
\end{array}\right.
\end{array}
$$

Figure 3: Identity Privacy Definition, copied from paper [1]

The Adversary may choose some pairs of users/templates to be registered in the system. For one user i he provides two different templates $b_i^{(0)}$ and $b_i^{(1)}$. Randomly one of the two templates is chosen to be used for enrollment of user I. After the challenger has simulated all enrollments the Adversary A2 may issue a polynomial number of verifications for all users/templates but not for user i. Finally the Adversary should tell which template was used with user i. If he can not do better than guessing, the system supports Identity Privacy. More precisely the adversary even may do a bit better/worse[1], but his advantage may only be negligible which means smaller than 1 / [some polynomial of security parameter l].

Transaction Anonymity

Transaction Anonymity means that for every query issued by the service-provider, a malicious database knows nothing about which user is authenticating himself to the service-provider. This is important to prevent transaction statistics about users and service-providers. Similar to Identity Privacy, Transaction Anonymity-Property is defined by a kind of a Game as well:

1 Note the absolute value in the definition of advantage, it is important to prevent construction a successful adversary out of an adversary always guessing worse than random

Definition 4. *A biometric-based authentication scheme achieves transaction anonymity if an adversary* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)$ *has only a negligible advantage in the following game, where the advantage is defined to be* $|\Pr[e' = e] - \frac{1}{2}|$.

$$
\begin{array}{l}
\mathbf{Exp}_{\mathcal{A}}^{Transaction\text{-}Anonymity} \\
\left|\begin{array}{rl}
(ID_j, b_j)(1 \le j \le N) & \leftarrow \mathcal{A}_1(1^\ell) \\
\emptyset & \leftarrow \mathsf{Enrollment}(1^\ell) \\
\{i_0, i_1\} & \leftarrow \mathcal{A}_2(Challenger, \mathsf{Verification}) \\
i_e & \stackrel{R}{\leftarrow} \{i_0, i_1\} \\
\emptyset & \leftarrow \mathsf{Verification}(i_e) \\
e' & \leftarrow \mathcal{A}_3(Challenger; \mathsf{Verification})
\end{array}\right.
\end{array}
$$

Figure 4: Transaction Anonymity Definition, copied from paper [1]

The adversary, a malicious database in this case, may choose pairs of username and biometric template, which will be enrolled by the challenger Adversary A2 may then do a polynomial number of verifications for single users but for the last verification at this point he provides two users. The challenger chooses one of these by random and processes the verification protocol with it. After that Adversary A3 may do any number of further verifications and should finally guess which of the two users was randomly chosen and verified by the challenger. If the Adversary can not do better than guessing or more precisely if he only has a negligible advantage, than the system provides Transaction Anonymity.

Assumptions

So far we have seen a basic overview of the system and definitions of the privacy properties it should fulfill. To proof that the proposed System really does provide Identity Privacy and Transaction Anonymity, Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval made some assumptions:

Biometric Distribution

The Biometric distribution assumption basically requires that one can distinguish whether two biometric templates are form the same person or from two different persons. There should be a threshold for the distance between two templates which for the majority of templates allows to make that distinction correctly. In the paper [1] this formal definition is given:

> Biometric Distribution assumption: Let H be the distance function in a metric space (in this paper, we assume it to be Hamming space). Suppose b_i and b_j are the reference biometric templates for Alice and Bob, respectively. There is a threshold value λ, the probability that $\mathsf{H}(b_i, b'_j) > \lambda$ is close to 1 and the probability that $\mathsf{H}(b_i, b'_i) \leq \lambda$ is close to 1, where b'_i and b'_j are the templates captured for Alice and Bob at any time.

Figure 5: Definition of the Biometric Distribution Assumption, copied from paper [1]

Liveness

A further requirement is that all templates received at the sensor are always read live from a living human being who really owns these biometrics, and there must not be a way to fool the sensor with things like fingerprints on plastic or iris images on contact-lenses. These things have to be prevented by the sensor or by some security personal watching the enrollment/verification process.

Security Link

It is assumed that all connections between the different components is secure by means of confidentiality and integrity. Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval suggest using standard protocols such as SSL or TLS to implement this.

Trust Relationships

It seems the sensor is the most sensitive component because here the user-identity and the biometric-template, which should be kept unlinked, meet at a single point. The sensor client can access both data and thus fake enrollments or verifications easily, and probably there is no way to prevent this by system design, so Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval require the sensor to be always honest and trusted by the other components. Using the security link assumption or additionally sensor-certificates in case of indirect communication, this allows the service-provider and the database to trust in the liveness of templates they receive.

It is assumed that both the service-provider and the database do their job correctly, make the right decisions when it comes to authentication and store and provide the right datasets, but, they may be very curious and try figure out private information about users or transactions. Nevertheless they may not collude in their effort to collect private data, and this is a reason why the design decision was made to keep service-provider and database well separated.

Outside-adversaries may try to impersonate honest users or try to thread privacy concerns by other means, but due to security link assumption they are very limited.

Authentication Scheme

Having done all this preliminary work we will now have a closer look at the authentication scheme and its procedures.

Enrollment

During the enrollment procedure the user chooses a personalized user-id ID_i and registers his biometric at the sensor. The sensor client then extracts a reference template b_i from the raw data.

Since biometric data varies from time to time and it is unlikely to get twice the same result reading the very same biometric again, a way has to be found to deal with these little changes or noise in data. One way is to make the template comparison robust against small deviations, the other way, which Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval chose in their paper [1], is to use a secure sketch which stores some extra information for registered templates that allows to recover a slightly modified version of a template back to the original reverence template. Of course this may only be done for small deviations from the reference template or otherwise user A's biometric might be restored to user B's reference template and thus user A was able to impersonate user B. It is reasonable to choose the threshold for the deviation allowed according to the threshold given in the biometric distribution assumption. A common way to implement secure sketches is to use fuzzy extractors/vaults, we had some presentations about that during our seminar session [3]. For the original description of the secure sketch see [4].

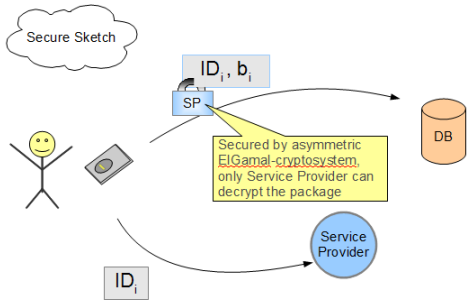

Figure 6: Enrollment procedure

Once the userid is entered and secure sketch and reference-template have been computed, the sensor sends the userid to the service-provider and an encrypted tuple of userid ID_i and reference-template b_i to the database. The (Idi, bi)-tuple has to be sent and stored encrypted because nobody (except the always honest and trusted sensor) should be able to link some user's id with the corresponding template. Here Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval used the asymmetric ElGamal-cryptosystem to encrypt the tuple in a way that only service-provider could decrypt it, we will later see why. Of course this package may never reach the service-provider which is not allowed to gather information about the correspondence between templates and userids either.

Verification

Once a user is registered he may want to use some service-provider's service that needs authentication. He enters his userid id and biometric b* at the sensor. The sensor than uses the secure sketch to compute the adjusted template b'.

The sensor then sends the userid to the service-provider plus some extra information required to query the database. This extra information contains ID_i, ID_{sp} and b' but is first encrypted for the service-provider then further encrypted with the database's public key and finally signed by the sensor. So in total a package $(ID_i, M, \text{sign}(ID_{sp} \| M, SK_{sensor}))$ with

$M = \text{enc}(X, PK_{DB}), X = \text{enc}(g_s^{ID_i \| ID_{sp} \| b_i'}, PK_{SP})$ is send to the service-provider

When the package arrives at the service-provider it notes the user's number i and forwards the rest of the package to the database which will first check the signature and -if valid- will then decrypt the outer encryption, which will result in the situation shown in the 4th figure on the right. Note that till now there was no point where either database or service-provider could link userid and template without helping each other, so identity privacy seems to be assured.

Whats left to do now is let the database do a comparison between the query and the stored dataset(s) and report to the service-provider whether the authentication is valid or failed. Even though the tuples (ID,b) are binary matching it is not possibly to do simply binary compare X and the stored record because both are encrypted with ElGamal cryptosystem which involves using a random number to generate different ciphertexts every time the same plaintext is encrypted. The Database can not decrypt and compare afterwards either, because by intention it does not have the key to decrypt.

Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval used a nice trick to do the comparison and, even better, do it in a way that only the service-provider can understand the result:

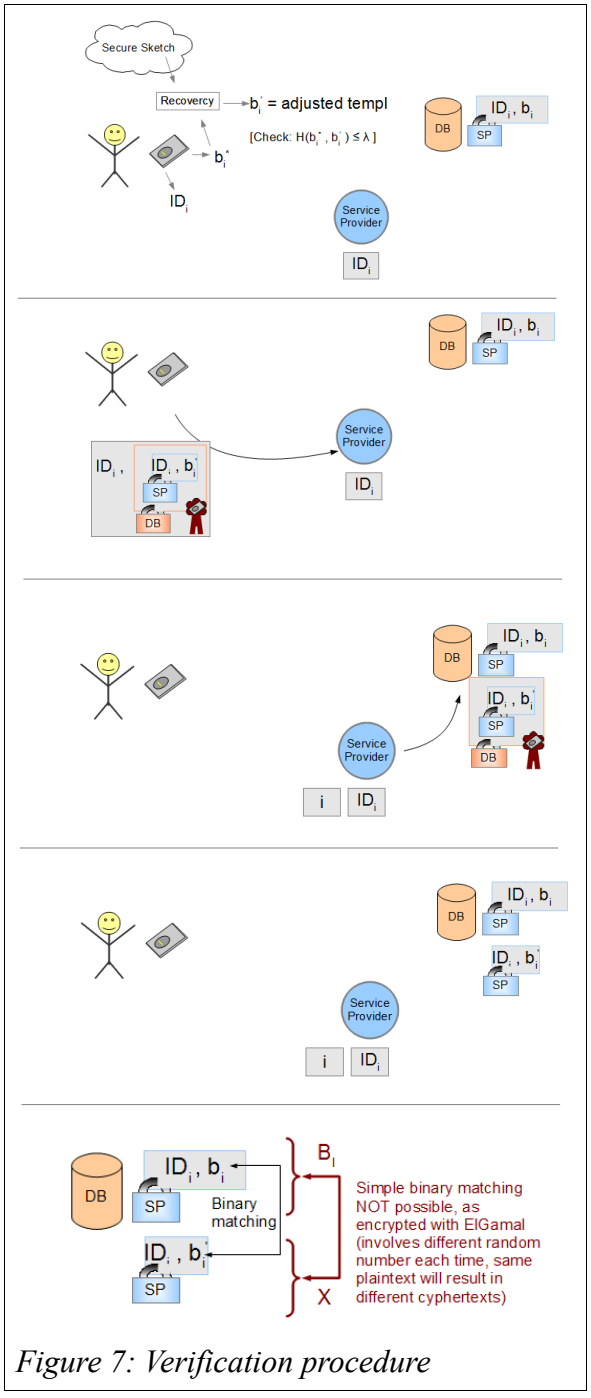

Figure 7: Verification procedure

They defined an operator Ø as $((c1, c2) Ø (c3, c4))^x = \left(\left(\frac{c1}{c3}\right)^x, \left(\frac{c2}{c4}\right)^x\right)$ for some integer x and El-Gamal-ciphertexts c1..c4, and used it to compute the query-result $R = (X Ø B_l)^{St}$ for the still encrypted query X, the dataset B_l and some

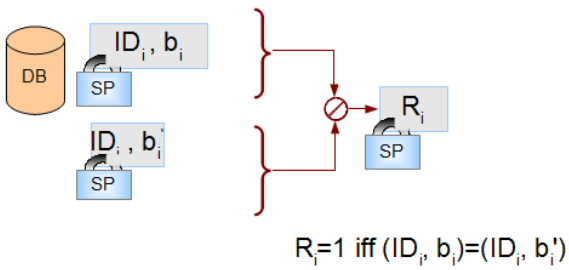

Figure 8: Ø Operator

randomly chosen integer $St \in_R \mathbb{Z}_{q_{SP}}$. $Dec(R,SK_{SP})$ will be 1 iff $Dec(X,SK_{SP})==DEC(B_I,SK_{SP})$, but with the privacy-advantage that the latter computations do not have to be done and the relationship between ID and template remains hidden.

A disadvantage of the solution seen so far is that the database contains a lot of datasets and every time the service-provider queries the database it either has to compare the query with all datasets and transfer all the results back which would be a computational and communicational disaster, or, the service-provider had to tell the database which record to match with but this would eliminate the transaction anonymity property. As a solution to this issue Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval propose using a Private Information Retrieval algorithm such as [5] which allows to reduce the computational and communication overhead without touching transaction anonymity.

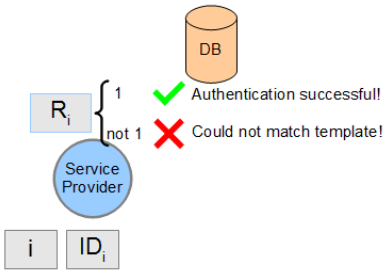

Figure 9: Verification result

Analysis of the proposed scheme

Soundness and Impersonation Resilience

The soundness of the scheme follows from biometric distribution assumption and the soundness. It is mentioned in the paper [1] that even if service-provider and database collude, they can not recover biometrics easily, because these are encoded in form of $(gs)^{ID_s \| ID_i \| b_i}$ and $(gs)^{ID_s \| ID_i \| b_i'}$ and thus are protected by the Discrete Logarithm assumption as long as the templates provide sufficient entropy.

Identity Privacy

Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval state in two lemmas that their proposed scheme achieves Identity Privacy, the proof itself was not included in the version of their paper [1] this report is based upon, but they say it was proofable and there would be a full version published where all the proofs were included.

> **Lemma 1.** *The proposed scheme achieves identity privacy against malicious $\mathcal{S}$, based on the semantic security of the ElGamal scheme and the existential unforgeability of the signature scheme.*
>
> **Lemma 2.** *The proposed scheme achieves identity privacy against malicious $\mathcal{DB}$, based on the semantic security of the ElGamal scheme.*

Figure 10: Lemmas from [1] regarding Identity Privacy

Transaction Anonymity

For Transaction Anonymity there is a lemma too, but the associated proof has to be looked up in the full version of paper [1] as well.

> **Lemma 3.** *The proposed scheme achieves transaction anonymity against malicious $\mathcal{DB}$, based on the semantic security of the ElGamal scheme and the security (user privacy) of the PIR protocol.*

Figure 11: Lemma from [1] regarding Transaction Anonymity

Conclusion

We have seen that privacy is an important issue in biometric based remote authentication systems, two concrete privacy concerns, Identity Privacy and Transaction Anonymity were formally defined and we learned about an authentication scheme Q. Tang, J. Bringer, H. Chabanne and D. Pointcheval proposed, which, given certain assumptions, can proofably archive the properties Identity Privacy and Transaction Anonymity. It is nice to have such proofs, but especially if it comes to practical application of such a scheme we should have a very close look at the assumptions made. A method for revocation was not discussed, but in their paper [1] the authors refer to [6] and [7] where an interesting concept of "cancelable biometrics" is introduced.

References

[1] Q. Tang, J. Bringer, H. Chabanne, D. Pointcheval: A Formal Study of the Privacy Concerns in Biometric-Based Remote Authentication Schemes. In: L.Chen, Y.Mu, W.Susilo (eds.) ISPEC 2008, LNCS 4991, pp.56-70, 2008, Springer Berlin Heidelberg

[2] J. Bringer, H. Chabanne, M. Izabachène, D. Pointcheval, Q. Tang, S. Zimmer: An application of the Goldwasser-Micali cryptosystem to biometric authentication. In: J. Pieprzyk, H. Ghodosi, E. Dawson (eds.) ACISP 2007. LNCS, vol. 4586, pp.96–106, 2007, Springer Heidelberg

[3] Seminar Biometry & Security, Bonn-Aachen International Center for Information Technology, winter-term 2009/2010, see for related presentations and further reading: http://cosec.bit.uni-bonn.de/students/teaching/09ws/09ws-sem/

[4] Dodis, Y., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 523–540. Springer, Heidelberg (2004)

[5] Gentry, C., Ramzan, Z.: Single-database private information retrieval with constant communication rate. In: Caires, L., Italiano, G.F., Monteiro, L., Palamidessi, C., Yung, M. (eds.) ICALP 2005. LNCS, vol. 3580, pp. 803–815. Springer, Heidelberg (2005)

[6] Bolle, R.M., Connell, J.H., Ratha, N.K.: Biometric perils and patches. Pattern Recognition 35(12), 2727–2738 (2002)

[7] Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. IBM Systems Journal 40(3), 614–634 (2001)