Pseudonyme Biometrik: Ein signatur-basierter Ansatz

Martin Johns
Universität Hamburg, Fachbereich Informatik
mj@martinjohns.com

Abstract: Dieser Artikel stellt einen Ansatz zur Pseudonymisierung biometrischer Daten vor. Im Unterschied zu bisher bestehenden Ansätzen, setzt die Pseudonymisierung bereits bei der Berechnung der biometrischen Signatur an. Es werden Anforderungen und Vorrausetzungen für ein derartiges Verfahren erarbeitet und anhand des Algorithmus von John Daugman ein Beispiel für die Anwendung des Ansatzes gegeben.

1 Motivation

Bei der rechnergesteuerten Authentikation von Personen koexistieren die Methoden der Authentikation mittels Wissen (z.B. Passwörter), Besitz (z.B. Chipkarten) und Biometrik[1]. Die Methoden der Biometrik besitzen gegenüber anderen Authentikationsmethoden die spezifische Eigenschaft, dass ein biometrisches Merkmal stets eindeutig an eine Person gebunden, beständig und unveränderlich ist.

Die daraus resultierende eindeutige Beziehung zwischen biometrischer Signatur und dem Träger des biologischen Merkmals kann anwendungsbezogen gewünscht sein (keine Weitergabe oder Verlust des Authentikationsmittels möglich) beinhaltet aber Gefahren für das informationelle Selbstbestimmungsrecht des Einzelnen: Können Chipkarten entsorgt und Passwörter geändert werden, bleibt das biometrische Merkmal (im Idealfall) ein Leben lang erhalten. Wurde das selbe Merkmal im Laufe der Zeit zweimal verwendet, können ggf. die im Zusammenhang mit diesem Merkmal (bzw. der zugehörigen Signatur) abgelegten Datensätze zusammengelegt werden.

In den letzten Jahren kann ein Zuwachs an Anwendungen biometrischer Systeme verzeichnet werden. Je größer die Bedeutung dieser Systeme im täglichen Leben wird, desto größer wird auch die Gefahr des Missbrauchs. Aus diesem Grund sind Verfahren, die diese eindeutige Verbindung zwischen Signatur und Merkmalsträger auflösen, ohne die Authentikationseigenschaften der biometrischen Methoden zu beeinflussen, wünschenswert.

[1] Weitere Methoden, wie z.B. Authentikation über Ort oder Zeit, sind in dieser Aufzählung bewusst unberücksichtigt geblieben, da diesen Methoden per se kein Nachweis einer Identität zugrunde liegt, sondern dieser im Anwendungsfall lediglich durch äußere Umstände zugeordnet werden kann.

2 Grundlegendes

2.1 Pseudonymität

Der Begriff der *Pseudonymität* bezieht sich im Folgenden stets auf den Grad, in dem sich eine Beziehung zwischen einer tatsächlichen Person und einer digitalen Identität herstellen lässt. [PK01] definiert in diesem Zusammenhang die beiden Extreme *„anonymity“* (keinerlei Zuordnung zwischen digitaler Identität und Person möglich) und *„accountability“* (volle Zuordnung möglich). Der Bereich der Pseudonymität erstreckt sich zwischen diesen Extremen.

Der Begriff des Pseudonyms, wie er im folgenden Verlauf dieses Textes verwendet wird, beinhaltet eine $1{:}n$ Relation: Eine Person agiert unter verschiedenen Pseudonymen, aber ein Pseudonym kann nur von genau einer Person verwendet werden[2].

2.2 Die biometrische Authentikation

Bei einer Authentikation über biometrische Methoden wird aus einem biologischen Merkmal eine biometrische Signatur berechnet, die mit den eingelernten Daten in der Datenbank verglichen wird. Die Daten in der Datenbank entsprechen Zweitupeln, bestehend aus einer biometrischen Signatur und den zugehörigen Personeninformationen, insbesondere der Personenidentität. Eine Person weist somit stets eine digitale *Identität* nach, die jedoch nicht zwangsläufig mit der tatsächlichen Identität der Person übereinstimmen muss.

Im Folgenden werden folgende Symbole verwendet:

M	:	biometrische Rohdaten
S	:	biometrische Signatur
PS	:	pseudonymisierte biometrische Signatur
$f(M)$	:	biometrischer Algorithmus ($f(M) = S$)
$d(S, S')$	:	Abstandsmaß zweier biometrischer Signaturen
T	:	Schwellwert (zwei biometrische Signaturen werden als gleich erkannt, wenn $d(S, S') < T$)

3 Voraussetzungen und Ziele der pseudonymen Biometrik

Wie in Abschnitt 1 bereits motiviert, ist eine Authentikation über Biometrik eindeutig für eine Person. Anonyme oder pseudonyme Nutzung ist somit nicht (oder lediglich eingeschränkt) möglich. Sobald zwei unterschiedliche Ressourcen mit dem selben biometri-

[2] [PK01] klassifiziert eine Reihe von verschiedenen Arten eines Pseudonyms (*person, role, relationship, role-relationship, transaction*). Ziel des vorgestellten Verfahrens ist eine Durchsetzung aller Arten von Pseudonymität, bei denen eine eindeutige Beziehung zwischen Pseudonym und Person besteht.

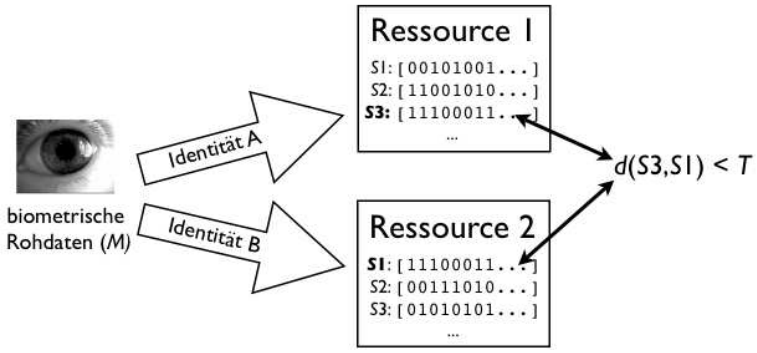

Abbildung 1: Identifizieren von Signaturen

schen Verfahren[3] gesichert werden, ist es durch die eindeutige Beziehung zwischen dem Merkmalsträger und der biometrischen Signatur möglich, Nutzer der beiden Ressourcen miteinander zu identifizieren. Dieses ist unabhängig von der mit den Signaturen verbundenen Identität: Die Daten eines Nutzers, zu dem bei zwei unterschiedlichen Ressourcen verschiedene Identitäten abgelegt sind, können aufgrund der Eindeutigkeit seiner biometrischen Signatur dennoch miteinander identifiziert werden (siehe Abb. 1).

Unter dem Gesichtspunkt der Pseudonymität können folgende Anforderungen formuliert werden:

(F1) Eine Möglichkeit der Nutzung eines Dienstes eines Anbieters unter zwei verschiedenen Identitäten (z.B. Nutzung von Webspeicherplatz).

(F2) Eine Möglichkeit der Nutzung zweier Dienste des selben Anbieters unter verschiedenen Identitäten (z.B. Teilname an Diskussionforen und persönliche Email).

(F3) Eine Verhinderung der Zusammenlegung von personenbezogenen Datensätzen (z.B. nach Zusammenschluss zweier Dienstanbieter).

(F4) Eine Verhinderung der illegalen Weitergabe der biometrischen Signatur an dritte (z.B. durch einen Einbruch in die Datenbank eines Dienstanbieters).

Die Voraussetzung für eine effektive Pseudonymisierung biometrischer Daten ist, unabhängig von der verwendeten Methode, dass keine Speicherung der biometrischen Rohdaten stattfindet. Bei einer Speicherung dieser Daten sind alle Pseudonymisierungsmaßnahmen hinfällig, da aus den Rohdaten stets eine unverfälschte Signatur berechnet werden kann. Die Instanz, welche die Kontrolle über die Erfassungseinheit besitzt, bestimmt somit den Erfolg der Pseudonymisierung. An dieser Stelle müssen für eine Durchsetzung entsprechende Kontrollmaßnahmen[4] existieren.

[3] Definiert durch die Kombination aus dem verwendeten biometrischen Algorithmus und dem aufgenommenen biologischen Merkmal.

[4] oder entsprechendes Vertrauen.

4 Bestehende Ansätze

Es wurden bisher einige Ansätze entwickelt, die eine pseudonyme Nutzung biometrischer Methoden ermöglichen sollen:

- Verwendung verschiedener Merkmale:

  Für jede verwendete Identität einer Person nutzt diese zur Authentikation ein anderes biologisches Merkmal [Köh99, S. 7]. Solange keines der Merkmale doppelt verwendet wird, ist eine Identifikation von Datensätzen nicht möglich.

- Hashen oder Verschlüsseln der Signaturen:

  Vor der Speicherung in der Referenzdatenbank wird die Signatur mit einen Einweg-Verfahren verschlüsselt. [Don99] erweitert diesen Ansatz mit einem für jede Identität spezifischen Wert, der in die Verschlüsselung eingeht.

- Auslagerung von Teilen der biometrischen Authentikation:

  Dieser Ansatz verfolgt eine Auftrennung des biometrischen Algorithmus, so dass Teile der Authentikation nicht von der Partei, gegenüber der die Authentikation stattfindet, durchgeführt werden. Die Auslagerung kann den ganzen Prozess der biometrischen Authentikation betreffen (wie in [Ble98] vorgestellt wird) oder lediglich Teile wie z.B. den Vergleich der Signaturen.

  Die betreffenden Bestandteile des Algorithmus können z.B. entweder auf Smartcards, die unter der Kontrolle des Merkmalsträgers stehen, verlagert werden oder unter der Kontrolle vertrauenswürdiger dritter Instanzen, wie z.B. Trustcenter, stehen.

Jedes dieser Verfahren hat Vor- und Nachteile, die hier nicht erörtert werden können. Allen Ansätzen ist gemein, dass sie in die Berechnung der Signatur nicht eingreifen. Somit bleibt die eigentliche Ursache des behandelten Problems, die Eigenschaft der Eindeutigkeit der Signatur für eine Person, weiterhin gegeben.

5 Der signatur-basierte Ansatz

Die grundlegende Idee des in diesem Text beschriebenen Ansatzes ist es, in die Berechnung der biometrischen Signatur S einen Maskierungs-Wert k zu integrieren (siehe Abb. 2):

$$PS = f(M, k) \tag{1}$$

Der Maskierungs-Wert k ist für eine gegebene Identität eindeutig. Diese Art der Berechnung soll folgende Eigenschaften besitzen:

(P1) $d\left(f(M,k), f(M',k)\right) < T$, wenn M und M' von der selben Person stammen

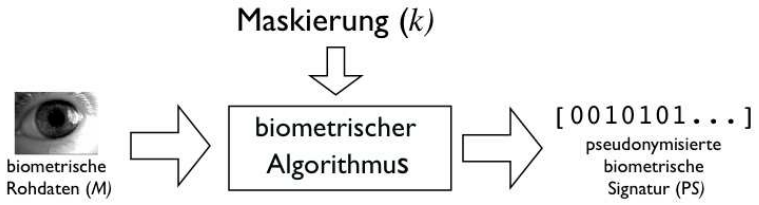

Abbildung 2: Pseudonymisierter biometrischer Algorithmus

(P2) $d\left(f(M,k), f(M',k)\right) > T$, wenn M und M' von verschiedenen Personen stammen

(P3) $d\left(f(M,k), f(M,k')\right) > T$, wenn k ungleich k'

Für eine zuverlässige Pseudonymisierung ist es weiterhin wichtig, dass keine deterministische Zuordnungsmöglichkeit zweier pseudonymisierter Signaturen derselben Person existiert. Daraus ergeben sich weitere Anforderungen:

(R1) Es darf keine Umkehrfunktion f^{-1} geben, die die pseudonymisierte Signatur PS in die unbearbeitete (nicht pseudonymisierte) Signatur[5] S überführt.

(R2) Es darf keine Vergleichsfunktion d', die eine Identifikation zweier pseudonymisierter Signaturen derselben Person erlaubt, existieren.

Die Punkte (P1) - (P3) formulieren die Anforderungen an die Pseudonymisierung, die Punkte (R1) - (R2) die Anforderungen an die Robustheit des Verfahrens. Die Pseudonymisierung kann an verschiedenen Punkten ansetzen:

1. Vor der Signaturberechnung:

   Über den biometrischen Rohdaten wird eine Transformation gelegt:

   $$PS = f(M,k) = f(t(M,k)) \tag{2}$$

2. Während der Signaturberechnung:

   Die Art der Signaturberechnung wird von k beeinflusst (siehe Abbschnitt 6).

3. Eine Kombination beider Ansätze

Da eine Methode der signatur-basierten Pseudonymisierung bereits zum Zeitpunkt der Signaturgenerierung greift, also umittelbar in die Signaturgenerierung integriert ist, ist diese Methode immer spezifisch für einen ausgewählten biometrischen Algorithmus.

[5] oder in die biometrischen Rohdaten M.

5.1 Geeignete biometrische Algorithmen

Nicht jeder biometrische Algorithmus ist für eine signatur-basierte Pseudonymisierung geeignet.

Grundsätzlich lässt sich sagen, dass biometrische Algorithmen, die einen „semantischen[6] Ansatz“ der Signaturberechnung verfolgen, sich schlechter eignen, eine passende Methode zu finden. Diese Algorithmen sind i. A. darauf angewiesen, dass die Struktur der zu verarbeitenden biometrischen Rohdaten von vorhersehbarer Form ist. Eine Transformation der Rohdaten vor der Signaturberechnung kann somit Probleme verursachen (eine Verzerrung eines Fingerabdruckbildes kann z. B. das Erkennen von Minutien verhindern). Weiterhin repräsentiert ein Teil der biometrischen Signatur bei diesen Algorithmen stets eine spezifische Ausprägung des Merkmals. Eine Pseudonymisierung der Signatur entspräche somit ggf. einer Auswahl, Addition oder Subtraktion von Merkmalen.

Vielen biometrischen Algorithmen, die syntaktische Methoden verwenden, ist gemein, dass ihre Signaturberechnung von hohem Abstraktiongrad ist. Dieser äußert sich meist in einer Behandlung der biometrischen Rohdaten mit numerischen Mitteln, die verlustbehaftet ist.

Besonders geeignet sind Algorithmen, die Schritte des Zusammenfassens (wie z.B. das Berechnen des Mittelwertes wie in [GZT00]) oder der Vergröberung (wie z.B. die Abbildung eines komplexen Waveletkoeffizienten auf den Quadranten seiner Phase in [Dau93]) von numerischen Werten beinhalten, da diese Schritte einen Aufschluss auf die genauen Eigenschaften der biometrischen Rohdaten aus der Signatur verhindern.

5.2 Zur Geheimhaltungsbedürftigkeit des Maskierungs-Werts k

Je nach verwendeter Methode der Pseudonymisierung, gestaltet sich der Umgang mit dem Maskierungs-Wert k unterschiedlich:

1. Geheimhaltung von k notwendig

   Ist die Herstellung einer Relation zweier pseudonymisierten Signaturen $PS_1 = f(M, k_1)$ und $PS_2 = f(M, k_2)$ durch die Kenntnis der k_i möglich, verbietet sich eine Speicherung oder Veröffentlichung von k. In diesem Fall sollte k vom Nutzer verwaltet werden.

2. Geheimhaltung von k nicht notwendig

   Kann auch mit Kenntnis der k_i kein Zusammenhang zwischen PS_1 und PS_2 her-

[6] In [Joh03] werden biometrische Algorithmen in die Klassen des *semantischen* und des *syntaktischen* Vorgehens eingeteilt: Semantische Algorithmen analysieren die biometrischen Rohdaten auf einer „verstehenden“ Ebene; den Mustern der Daten werden dabei tatsächliche Eigenschaften des Merkmals zugeordnet (wie z.B. die Minutien der Fingerabdrücke). Syntaktische Algorithmen betrachten die biometrischen Rohdaten auf abstrakte Weise, ohne Rückschlüsse auf tatsächliche Ausprägungen vorzunehmen. Die Daten werden in diesem Fall meist nach der Vorverarbeitung mit statistischen Verfahren verglichen (siehe z.B. [GZT00]).

gestellt werden, kann k zusammen mit den anderen Identitätsdaten in der biometrischen Datenbank der Ressource gespeichert werden.

6 Ein Beispiel: Pseudonymisierung während der Signaturberechnung

Das folgenden Beispiel beziehen sich auf die Algorithmus von John Daugman [Dau93] zur Iriserkennung. Es dient in erster Linie dem Zweck, die Wirkungsweise des beschriebenen Ansatzes zu illustrieren und nachzuweisen, dass eine praktische Umsetzung der signatur-basierten Pseudonymisierung biometrischer Daten möglich ist. Es erhebt nicht den Anspruch, die Anforderungen aus Abschnitt 5 vollständig zu erfüllen.

Der Algorithmus von John Daugman zur Iriserkennung [Dau93] basiert auf den Vergleich der Phaseninformationen komplexer Waveletkoeffizienten bezogen auf festgelegte Analysepunkte. Die Phase der Waveletkoeffizienten ist stark von der relativen Position des Schwerpunktes des Wavelets zu den Bilddaten abhängig. Bereits kleine Verschiebungen der Position können sich in starken Änderungen der Phaseninformationen äußern [WFNM99].

Ein möglicher Ansatzpunkt für eine Pseudonymisierung der Signaturberechnung ist somit die Platzierung der Analysepunkte: Eine Verschiebung dieser Punkte führt zu einer Änderung der Phaseninformation der zugehörigen Waveletkoeffizienten und somit zu einer reproduzierbaren Änderung der Signatur.

Untersuchungen, die im Rahmen der Erstellung dieses Textes durchgeführt wurden, haben ergeben, dass durchschnittlich bei einer horizontalen Verschiebung des Wavelets um 10 Prozent der effektiven Breite des Wavelets (b) sich der entsprechende Anteil der Signatur verändert[7]. Verschiebungen in vertikaler Richtung haben ebenfalls Einfluss auf die resultierende Signatur, aber nicht von selber Gewichtung. Die Signatur bezieht sich insgesamt auf 1024 Analysepunkte. Ein Maskierungs-Wert k entspräche somit einem Vektor

$$v_i = (v_{i_x}, v_{i_y}), \qquad i \in [1, 1024] \qquad v_{i_{(x,y)}} \in [-b, b] \tag{3}$$

wobei die v_{i_j} der jeweiligen (zufälligen) Verschiebung des Analysepunkts in horiziontaler bzw. vertikaler Richtung[8] entsprechen.

Die resultierende Version der Signaturberechnung wäre somit (abgeleitet aus [Dau93]):

$$h_{i_{\{Re,Im\}}} = \mathrm{sgn}_{\{Re,Im\}} \int_{\rho} \int_{\phi} I(\rho,\phi)\; e^{-i\omega(\theta_i + v_{i_x} - \phi)} e^{-\frac{(\tau_i + v_{i_y} - \rho)^2}{\alpha^2}} e^{-\frac{(\theta_i + v_{i_x} - \phi)^2}{\beta^2}} \rho \, d\rho \, d\phi. \tag{4}$$

Diese Modifikation des Algorithmus bietet bei einer kleinen Anzahl von zu vergleichenden

[7] Bei diesen Werten besteht eine Abhängigkeit vom Größenverhältnis Wavelet / Bilddaten.

[8] Daugman verwendet in seinem Algorithmus Polarkoordinaten bezüglich des Augenmittelpunktes. Eine vertikale Verschiebung entspricht somit einer radialen Verschiebung und eine horizontale Verschiebung einer in Dreh-Richtung.

Datensätzen (also einer eingeschränkten Zahl von Pseudonymen pro Person) eine effektive Pseudonymisierung ohne eine Notwendigkeit der Geheimhaltung von k. Bei ein

er großen Anzahl von Pseudonymen pro Person können unter Verwendung von k über statistische Methoden Rückschlüsse auf Relationen zwischen den pseudonymisierten Signaturen ermittelt werden.

7 Fazit

Unter dem Aspekt der zu erwartenden wachsenen Verbeitung biometrischer Methoden ist es nötig rechtzeitig Aspekte des Datenschutzes zu berücksichtigen. Der in diesem Text beschriebene Ansatz bietet ein Verfahren das alleinstehend oder in Kombination mit

anderen Ansätzen die in Abschnitt 3 beschriebenen Nachteile der biometrischen Authentikation ausgleicht. Es steht noch aus, bestehende biometrische Algorithmen auf ihre Eignung für eine Erweiterung durch signatur-basierte Pseudonymisierung zu prüfen und diese Erweiterungen zu entwickeln. Weiterhin ist es möglich bereits beim Entwurf neuer biometrischer Algorithmen Methoden der signatur-basierten Pseudonymisierung umzusetzen.

Literatur

[Ble98] Gerrit Bleumer. Biometric yet Privacy Protecting Person Authentikation. In *Information Hiding*, LNCS, pages 99 – 110, Berlin Heidelberg, 1998. Springer.

[Dau93] John Daugman. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. In *Transactions on Pattern Analysis and Machine Intelligence*, volume 15, pages 1148 – 1161, November 1993.

[Don99] Lutz Donnerhacke. Anonyme Biometrie. *DuD*, 3/99:151 – 154, 1999.

[GZT00] C. Garcia, G. Zikos, and G. Tziritas. Wavelet Packet Analysis for Face Recognition. In *Image and Vision Computing*, volume 18(4), pages 289 – 297, 2000.

[Joh03] Martin Johns. Anwendung von Wavelets in der biometrischen Authentikation. Master's thesis, Universität Hamburg, Fachbereich Informatik, Hamburg, 2003.

[Köh99] Marit Köhntopp. Technische Randbedingungen für einen datenschutzgerechten Einsatz biometrischer Verfahren. In *Patrik Horster (Hg.): Sicherheitsinfrastrukturen; Proceedings zur Arbeitskonferenz Sicherheitsinfrastrukturen 1999*. Vieweg, 1999.

[PK01] Andreas Pfitzmann and Marit Köhntopp. Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology. In H. Federrath, editor, *Designing Privacy Enhancing Technologies*, number 2009 in LNCS, page 1pp, Berlin Heidelberg, 2001. Springer.

[WFNM99] L. Wiskott, J.-M. Fellous, N.Krüger, and C. Malsburg. Face Recognition by Elastic Bunch Graph Matching. In *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, pages 375–373, London New York, 1999. CRC Press.