Christoph Busch, Ulrike Korte, Sebastian Abt, Christian Böhm, Ines Färber, Sergej Fries, Johannes Merkle, Claudia Nickel, Alexander Nouak, Alexander Opel, Annahita Oswald, Thomas Seidl, Bianca Wackersreuther, Peter Wackersreuther, Xuebing Zhou

Biometric Template Protection

Ein Bericht über das Projekt BioKeyS

Biometrische Systeme sind zwar technisch weit ausgereift und bieten heute Erkennungsleistungen, die noch vor 10 Jahren unerreichbar waren. Jedoch ist ein weit verbreiteter Einsatz von biometrischen Authentisierungsverfahren durch Bedenken hinsichtlich des notwendigen Schutzes von Referenzdaten gebremst. Eine sichere und datenschutzfreundliche Verarbeitung von biometrischen Daten wird möglich, wenn Template Protection Verfahren zum Einsatz kommen. Diese Verfahren wurden in einer wissenschaftlichen Studie (BioKeyS-Pilot-DB Teil 2) des Bundesamtes für Sicherheit in der Informationstechnik (BSI) untersucht. Dieser Artikel berichtet über die Ergebnisse im Projekt. Er zeigt auf, wie Mechanismen zum Schutz von biometrischen Daten mit Zusatzinformationen z.B. Passwörtern verknüpft und wie die Verfahren auch in Identifikationssystemen eingesetzt werden können.

Projektgruppe BioKeyS

BioKeyS ist ein Forschungsprojekt des BSI, das unter Leitung und im Auftrag des BSI durchgeführt wurde. Die Projektleitung auf Seiten des Auftragnehmers wurde durch die Hochschule Darmstadt durchgeführt. Beteiligt sind ferner das Fraunhofer-Institut für Graphische Datenverarbeitung IGD , die Rheinisch-Westfälische Technische Hochschule Aachen, die Ludwig-Maximilians-Universität München und das Unternehmen secunet Security Networks AG.

Johannes Merkle, Claudia Nickel, Ulrike Korte, Xuebing Zhou, Christoph Busch

Christian Böhm, Bianca Wackersreuther, Annahita Oswald, Peter Wackersreuther

Thomas Seidl, Ines Färber, Sergej Fries

1 Schutz biometrischer Templates

Biometrische Charakteristika können nicht weitergegeben werden. Deshalb wird Biometrie gerne zur Steigerung der Sicherheit von Anwendungen eingesetzt, da so die Zuverlässigkeit der Authentisierung gesteigert werden kann. Ein möglicher Vorbehalt gegen die Verwendung biometrischer Verfahren ist, dass die erreichte erhöhte Sicherheit mit einem verminderten Schutz der Privatsphäre einhergehen kann [CS07]. Darüber hinaus ist es möglich, dass die Einbeziehung biometrischer Verfahren aufgrund von im biometrischen Subsystem vorherrschender Schwachstellen in neuen Sicherheitsrisiken resultiert.

Nach Jain [Jain08] kann das Sicherheitsrisiko eines biometrischen Verifikationssystems in vier Kategorien unterteilt werden:

- Immanente biometrische Fehler aufgrund vom biometrischen Verifikationssystem getroffener falscher Entscheidungen, die häufig durch Wahrscheinlichkeitswerte für Falsch-Akzeptanz und/oder Falsch-Rückweisung ausgedrückt werden.
- Angriff auf die Systemverwaltung aufgrund unzulänglicher Verwaltungsrichtlinien
- Unzulänglich geschützte Infrastruktur, resultierend in Schwachstellen im Zusammenhang mit nicht hinreichend gesicherter Hardware, Software oder Kommunikationskanälen.
- Öffentlichkeit von biometrischen Charakteristika, was die versteckte Gewinnung biometrischer Samples erleichtert und die Erzeugung von Plagiaten zur Beeinflussung des Ergebnisses eines Verifikationssystems ermöglicht.

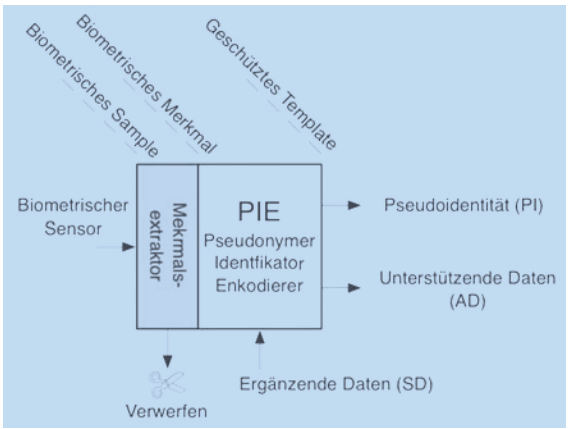
Nicht für alle Risiken lassen sich machbare Gegenmaßnahmen finden. Die Beständigkeit biometrischer Charakteristika ist einerseits eine für die Erkennungsleistung erstrebenswerte Eigenschaft, hat aber auch bedeutende Auswirkungen auf die eingeschränkten Möglichkeiten der Risikominimierung im Bezug auf Identitätsdiebstahl. Sobald ein biometrisches Charakteristikum unberechtigt aufgezeichnet wurde (z.B. in Form eines flüchtigen Latenzfingerabdrucks) und einem potenziellen Angreifer in Form eines Plagiates zur Verfügung steht, ist es so gut wie unmöglich dieses Charakteristikum zu erneuern, wenn man von der theoretischen Option eines chirurgischen Eingriffs absieht. Eine signifikante Verminderung der mit gestohlenen biometrischen Charakteristika einhergehenden Risiken kann jedoch dadurch erreicht werden, dass die Erneuerbarkeit biometrischer Templates, das heißt die Repräsentation eines biometrischen Charakteristikums in einem biometrischen System, sichergestellt wird.

Die Speicherung geschützter biometrischer Templates sollte den folgenden Bedingungen genügen, um die Privatsphäre der betroffenen Person ausreichend sicherzustellen:

- Eine unzulänglich geschützte Infrastruktur, resultierend in Schwachstellen im Zusammenhang mit nicht hinreichend gesicherter Hardware, Software oder Kommunikationskanälen ist zu vermeiden.
- Es ist unmöglich die ursprünglichen biometrischen Samples (z.B. Fingerbilder), die biometrischen Attribute oder irgendeine aus dem biometrischen Sample hergeleitete sensitive Information (wie zum Beispiel Gesundheitsinformationen, Informationen über ethnischen Ursprung, etc.) unmittelbar aus dem geschützten Template zu entnehmen oder das geschützte Template entsprechend zu dekodieren.
- Es ist unmöglich, eindeutige Verbindungen zwischen Personen innerhalb einer Datenbank oder datenbankübergreifend durch das Vergleichen von geschützten Templates herzustellen.
- Ein geschütztes biometrisches Template repräsentiert ausschließlich Daten für einen spezifischen, im Voraus definierten Verwendungszweck oder eine Anwendung.

Geschützte biometrische Templates sollten Mechanismen für deren Widerruf unterstützen. Weiterhin sollte der Kodierungsprozess über Möglichkeiten zur Generierung mehrerer unabhängiger geschützter Templates auf Basis derselben oder sehr ähnlicher biometrischer Charakteristika verfügen. Diesen Prozess der Generierung mehrerer unabhängiger geschützter Templates von denselben biometrischen Charakteristika bezeichnet man als Diversifikation. Die Diversifikationseigenschaft wird zur Vermeidung unerwünschter Verknüpfungen von Personen zwischen Datenbanken und zur Vermeidung der Suche nach Personen mit sehr ähnlichen biometrischen Charakteristika benötigt.

Bild 1 | Erzeugung geschützter Templates (biometrisches Pseudonym PI und unterstützende Daten AD)



2 Referenzarchitektur zum Schutz biometrischer Daten

Mit dem internationalen Standard ISO/IEC 24745 wurde eine Referenzarchitektur zum Schutz biometrischer Daten definiert [ISOtp]. Diese Referenzarchitektur liefert einen Rahmen für die Erzeugung und die Speicherung von geschützten biometrischen Referenzen und fügt sich ein in die Referenzarchitektur eines generischen biometrischen Systems [ISOtp]. Diese Referenzarchitektur erhebt den Anspruch einer Technologieneutralität, d.h. sie soll ein Framework für viele zurzeit existierende Techniken zum Schutz von Templates bieten. Abhängig von einer konkreten biometrischen Anwendung können unterschiedliche technische Anforderungen an die Architektur gestellt werden, weshalb für eine spezifische Implementierung einzelne funktionale Komponenten der Referenzarchitektur möglicherweise nicht zum Einsatz kommen oder zusätzlich hinzugefügt werden müssen.

Statt der üblicherweise gespeicherten Finger- oder Gesichtsbilder wird nach der Referenzarchitektur ein biometrisches Pseudonym (pseudonymer Identifikator – PI) gemeinsam mit unterstützenden Hilfsdaten (Auxilliary Data – AD) als biometrische Referenz hinterlegt. Pseudonyme Identifikatoren sind diversifizierbare, geschützte Identitätsverifikationsstrings (binäre Strings) innerhalb eines vordefinierten Kontexts. Ein pseudonymer Identifikator gibt keine Informationen preis, die Aufschluss über die ursprünglich erhobe-

Bild 2 | Referenzarchitektur eines Template Protection Systems

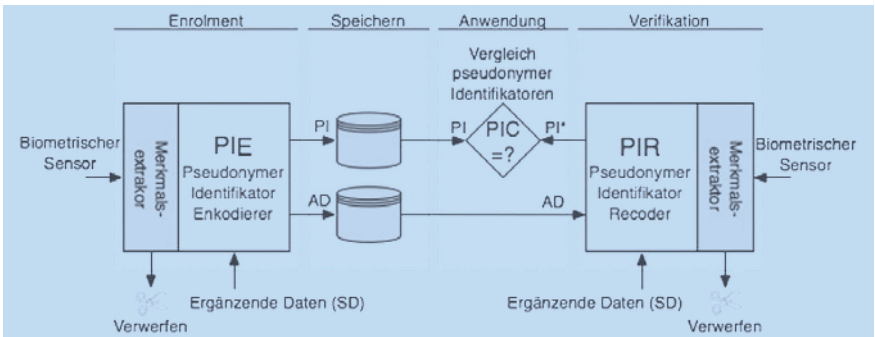
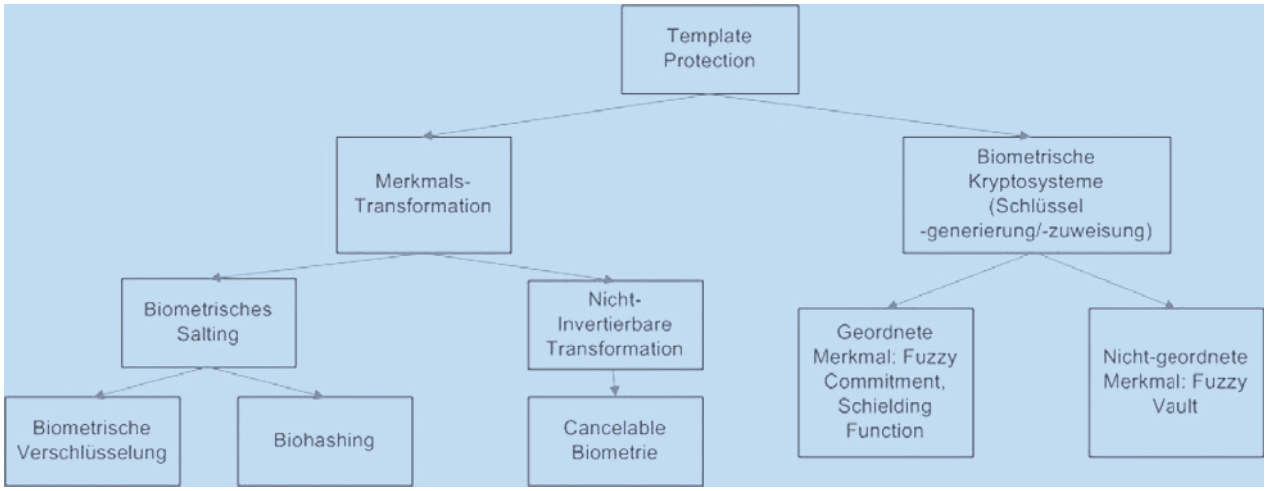


Bild 3 | Übersicht von Template-Protection-Verfahren



nen Daten, die zwischenzeitlich verarbeiteten Daten (z.B. die Fingerprint-Minutien) oder sonstige Identitätsattribute des Besitzers geben.

Der Prozess zur Erzeugung von pseudonymen Identifikatoren wird in Bild 1 dargestellt. Während einer Enrolmentphase wird für ein Individuum eine biometrische Referenz generiert. Innerhalb dieses Prozesses werden von einem biometrischen Datenerfassungsgerät (Sensor) ein oder mehrere biometrische Samples, wie zum Beispiel ein Bild eines Fingerabdruckes oder ein Gesichtsphoto, erzeugt und im Anschluss von einem Merkmals-Extraktor zur Erzeugung biometrischer Merkmale verarbeitet, die in einem proprietären oder einem standardisierten Templateformat zwischengespeichert werden. Abschließend werden von einem Pseudonym-Identifikator-Encoder (PIE) ein pseudonymer Identifikator (PI) sowie möglicherweise benötigte unterstützenden Daten (AD) erzeugt und diese Daten als Referenz hinterlegt.

Der PIE verwendet als Eingabe optional weitere ergänzende Daten (Supplementary Data – SD), die zum Beispiel für folgende Zwecke verwendet werden können:

- Sicherheitsverbesserung durch besitz- oder wissensbasierte Schlüssel, die von der zu erfassenden Person eingegeben werden müssen;
- Verbesserungen durch anwendungs- oder systemspezifische Schlüssel oder Signaturen;
- Verbesserungen durch Limitierung des Wirkbereiches eines pseudonymen Identifikators durch Einbeziehen von digitalen Signaturen oder Zertifikaten oder von zeit- oder ortsabhängigen Informationen für die eine PI Gültigkeit besitzen soll.

Diese verwendeten ergänzenden Daten werden jedoch nicht zusammen mit den geschützten Templates gespeichert, sondern nach dem Generieren des pseudonymen Identifikators verworfen. Einige dieser pseudonymen Identifikatoren könnten darüber hinaus auch zur sicheren Identifizierung genutzt werden und hierdurch zum Beispiel bei der Prüfung auf Duplikate während des Enrolments helfen.

Im Verifikationsprozess wird aus dem biometrischen Probesample durch den Pseudonymer-Identifikator-Recoder (PIR) unter Verwendung der bereitgestellten unterstützenden Daten erneut ein pseudonymer Identifikator PI* generiert. Dieser neu erzeugte pseudonyme Identifikator wird mit dem während des Enrolments erzeugten PI verglichen (siehe z.B. [JW99], [DRS04], [ST06], [NJP07], [Rat01]). Wurden dem PIE während der Enrolmentphase ergänzende Daten (SD) zur Verfügung gestellt, so müssen diese Daten auch dem PIR zur Verfügung gestellt werden. Nach dem Erzeugen des neuen pseudonymen Identifikators PI* durch den PIR werden alle Eingabedaten, d.h. das biometrische Sample, die Merkmalsdaten und die ergänzenden Daten gelöscht und PI* wird an einen Pseudonymen Identifikator-Comparator (PIC) übergeben, der PI mit PI* vergleicht. Ausschließlich bei Identität der beiden pseudonymen Identifikatoren ist die Verifikation erfolgreich.

2.1 Verfahren zu Biometric-Template-Protection

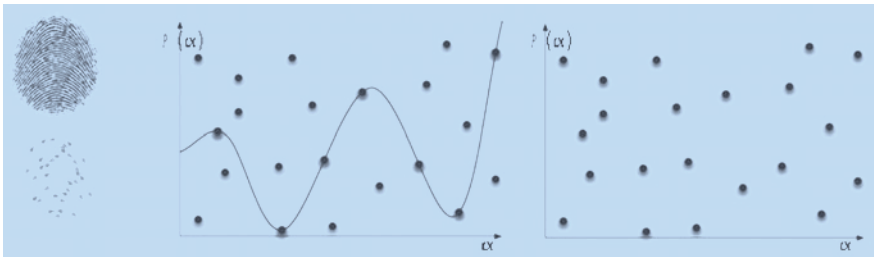
In [Jain08] findet sich eine ausführliche Übersicht zum Stand der Technik von Template Protection Verfahren. Die existierenden Ansätze lassen sich in Transformationsmethoden und Biometrische Kryptosysteme gruppieren. Bei den Transformationsmethoden werden Zufallswerte als Salz (Salt) oder Transformationsparameter verwendet, um die biometrischen Daten zu schützen. In biometrischen Kryptosystemen können unterschiedliche Schemata für geordnete und nicht-geordnete Merkmale verwendet werden. Man unterscheidet hier zwischen den Kryptoverfahren mit Anwendungen zur Schlüsselgenerierung und denen zur Schlüsselzuweisung. Bei der Schlüsselgenerierung kann ein eindeutiger Schlüssel aus den biometrischen Daten erzeugt werden. Schlüsselzuweisungen verwenden unterschiedliche zufällige Schlüssel, um biometrische Daten zu schützen.

2.1.1 Transformationsmethoden

In den Transformationsmethoden werden die pseudonymen Identifikatoren (PI) aus biometrischen Merkmalen mit zufälligem Salt oder einer nicht-invertierbaren Funktion abgeleitet. Sowohl der zufällige Salt als auch die Parameter der nicht-invertierbaren Funktionen sind benutzer- und anwendungsspezifisch. Deshalb müssen diese unterstützenden Daten geheim gehalten werden. Exemplarisch sei hier das Verfahren von Ratha et al. genannt [Rat01]. Für die minutienbasierte Fingerabdruckerkennung werden kartesische und pola-

re Transformationen oder Transformationen mittels Oberflächenfaltung (surface folding) eingesetzt. Die kartesische Transformation bildet die Minutien, die in einer gleichverteilten Zelle eines regulären Gitters über dem Fingerbild liegen, zufällig auf eine neue Zelle ab. In gleiche Weise unterteilt und verwürfelt die Polar-Transformation die Minutien-Unterteilung in einem Polar-Koordinatensystem.

Bild 4 | Stützpunkte und Streupunkte des Vault Sets



2.1.2 Biometrische Kryptosysteme

Biometrische Kryptosysteme kombinieren Kryptografie mit Fehlerkorrekturverfahren oder Quantisierung, um biometrische Daten zu schützen. Kryptografische Verfahren wie etwa Hashfunktionen sind empfindlich gegenüber den Variationen biometrischer Daten, die beispielsweise aufgrund der Alterung, der Änderung der Beleuchtung, der Luftfeuchtigkeit oder durch Sensorrauschen entstehen. Deswegen werden Fehlerkorrektur- bzw. Quantisierungsverfahren zur Kompensation dieser Variationen verwendet. Biometrische Kryptosysteme unterscheiden sich darin, wie die Mechanismen eingesetzt werden und welche Arten von Fehlern toleriert werden können. Auf Quantisierung basierende Verfahren sowie das Fuzzy Commitment sind für geordnete Merkmale geeignet, die eindeutig als Bitstring beschrieben werden können. Der Fuzzy Vault ist dagegen für ungeordnete Merkmale wie Minutien von Fingerabdrücken konzipiert.

Bei einer Quantisierung werden die biometrischen Daten in Komponenten zerlegt, aus denen dann jeweils wenige, robuste Bits extrahiert werden. Oft werden dafür die Werte durch eine Konstante geteilt und gerundet. Für die extrahierten Daten werden dann kryptographische Hashwerte (z.B. mit SHA-2) berechnet und abgespeichert. Die Schwierigkeit liegt darin, die Quantisierung so vorzunehmen, dass zum Einen die resultierenden Daten mit hoher Wahrscheinlichkeit bei jeder Messung konstant sind, zum Anderen aber nicht zuviel von der Information verloren geht, mit der Benutzer unterschieden werden. Dazu müssen die statistischen Verteilungen der biometrischen Daten verschiedener Benutzer (inter-class variation) und des Rauschens (intra-class variation) hinreichend bekannt sein. Zusätzlich werden für jeden Benutzer individuelle Hilfsdaten für die Quantifizierung ermittelt und hinterlegt. Ein Beispiel für Verfahren, die mit Quantisierung arbeiten, sind z.B. die in von Linnartz und Tuyls vorgestellten Shielding Functions [Lin03].

Die auf dem Fuzzy-Commitment basierenden Verfahren konvertieren die biometrischen Daten idealerweise in einen gleichförmig unabhängig verteilten binären Vektor. Eine geheime Zeichenkette wird zufällig generiert, deren Hashwert *h(s)* der PI ist. Aus der Zeichenkette wird unter Verwendung eines Fehlerkorrekturverfahrens ein Codewort *c* gebildet, das die gleiche Länge wie der binäre Merkmalsvektor *t* hat. Durch die Kodierung wird die Zeichenkette mit spezifischer Redundanz ausgestattet, die später die Korrektur von Fehlern ermöglicht. Das Codewort wird mit dem binären Merkmalsvektor mittels XOR-Operation verknüpft. Die resultierende Bitfolge offenbart kaum noch Informationen über seine biometrische Herkunft und kann als Hilfsdaten AD gespeichert werden. Während der Verifikation kann ein „beschädigtes Codewort" aus der biometrischen Probe und den Hilfsdaten gewonnen werden. Wenn die Abweichung zwischen den Enrolment- und Verifikationsdaten nicht zu groß ist, kann der entsprechende Fehlerkorrektur-Dekodierer die Fehler korrigieren und die geheime Zeichenkette rekonstruieren. Ein exakter Vergleich zwischen gespeichertem PI und dem Hashwert der rekonstruierten Zeichenkette liefert das gesuchte Verifikationsergebnis.

Sind die biometrische Daten zu stark verrauscht, kann es passieren, dass bei der Verifikation ein entstandener Fehler nicht mehr mit dem Fehlerkorrekturverfahren korrigiert werden kann. Prinzipiell können Fehlerkorrekturverfahren weniger als 50% der Bits korrigieren; in der Praxis ist die maximal korrigierbare Zahl von Bitfehlern deutlich geringer. Die Rekonstruktion der Zeichenkette ist somit nicht immer gewährleistet.

Die Erkennungsleistung und die Sicherheit der hinterlegten Templates sind bei diesen Verfahren durch den Informationsgehalt (Entropie) der biometrischen Daten und den verwendeten Fehlerkorrektur-Code limitiert. Wenn die biometrischen Daten relativ robust sind und eine hohe Entropie aufweisen, kann ein sehr hohes Schutzniveau für die Templates realisiert werden. Leider bieten einzelne biometrische Merkmale in der Regel nicht genug Entropie, um ein sehr hohes Sicherheitsniveau zu gewährleisten. Eine mögliche Lösung stellt die Kombination mehrerer biometrischer Merkmale dar.

Beispiele für Template Protection Verfahren, die auf dem Fuzzy Commitment basieren, sind die von Hao et al. [Hao06] und Korte et al. [Kor08] beschriebenen. Es gibt auch Verfahren, die das Fuzzy Commitment mit einer Quantisierung kombinieren, wie das von Tuyls [Tuy05].

Das Fuzzy Vault Verfahren [JS02] wurde für ungeordnete biometrische Merkmale mit unterschiedlicher Länge entwickelt, wie sie zum Beispiel bei der Verwendung von Fingerabdruckminutien auftreten. Weil sich die Anzahl und die Position der detektierten Minutien verändern, wird eine Variante von Shamir's Secret-Sharing-Protokoll [S79] statt des kryptographischen Hash-Werts verwendet. In diesem Secret-Sharing Protokoll können mehrere Stützstellen für geheime Zeichenketten produziert werden und eine Teilmenge der Stützstellen ist ausreichend, um die Zeichenkette zu rekonstruieren. Im Enrolment des Fuzzy-Vault-Verfahrens wird ein zufälliges Polynom

$$P(\alpha)=s_0+s_1\alpha+s_2\alpha^2+\ldots+s_d\alpha^d$$

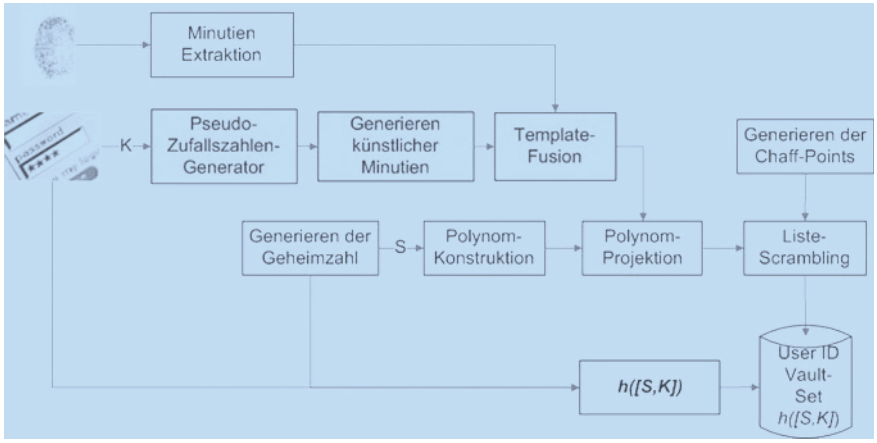
vom Grad *d* generiert. Dessen Koeffizienten bilden die geheime Zeichenkette

$$S=[s_0, s_1, s_2, \ldots, s_d],$$

die – analog zum Fuzzy Commitment Verfahren – mit den biometrischen Daten verknüpft wird. Dazu werden die jeweils zu einer Minutie gehörenden Informationen eines Enrolment-Fingerabdrucks als Zahl M dargestellt. Beispielsweise können die x- und y-Position einer Minutie zu einer 16-Bit langen Zeichenfolge konvertiert

werden, die zu einer Zahl in dem endlichen Körper GF(2^4) korrespondiert. Eine Stützstelle des Polynoms besteht nun aus der Minutieninformation M und deren Projektion $P(M)$ auf das gewählte Polynom. Darüber hinaus wird eine große Anzahl von Spreupunkten (Chaff Points) zufällig generiert. Diese Spreupunkte dienen zur Verschleierung der Stützpunkte. Die Streupunkte und Stützpunkte gemeinsam bilden das „Vault" (dt. „Tresor") als biometrische Referenz. Der Hash-Wert der geheimen Zeichenkette ist der pseudonyme Identifikator PI und wird mit dem Vault Set zusammen gespeichert. Die Funktionsweise des Fuzzy Vault ist in Bild 4 verdeutlicht: Die Kurve im linken Teildiagramm stellt das Polynom $P(\alpha)$ dar, auf der die Minutieninformationen und die entsprechenden Funktionswerte als Stützpunkte eingezeichnet sind. Zahlreiche Spreupunkte liegen außerhalb der Funktionskurve des Polynoms. Wie man im rechten Teildiagramm sieht, ist es ohne Informationen über die echten Minutien kaum möglich, die Stützpunkte, und damit das Polynom, zu identifizieren. Es wäre rechnerisch viel zu aufwändig, alle Kombinationen von potenziellen Stützpunkten zu testen, um das korrekte Polynom zu finden. Bei der Verifikation werden die Stützpunkte anhand der Minutien aus einem Vergleichs-Fingerabdruck identifiziert. Da die Ordnung des Polynoms kleiner ist als die Anzahl der abgebildeten Minutien-Punkte, reicht eine Teilmenge der Minutien-Punkte aus, um das Polynom und damit auch die geheime Zeichenkette zu rekonstruieren. Wenn ausreichend viele echte Stützpunkte gefunden wurden, kann das korrekte Polynom sowie die geheime Zeichenkette rekonstruiert werden. Die Korrektheit der geheimen Zeichenkette kann durch einen Vergleich des gespeicherten Hash-Wertes und des Hash-Wertes der neu berechneten Zeichenkette geprüft werden. In [JS02] wurde dargestellt, dass die Robustheit des Verfahrens verbessert werden kann, wenn zusätzlich ein Fehlerkorrektur-Code für Polynome zum Einsatz kommt. In [UPJ05] wurde eine Lagrange-Interpolation verwendet, um das Polynom zu rekonstruieren. Weiterhin wurde eine CRC-Kodierung benutzt, um die Korrektheit der berechneten geheimen Zeichenkette zu überprüfen. In [NJP07] wird gezeigt, dass die Erkennungsleistung dieses Verfahrens mit zusätzlichen unterstützenden Daten (AD) verbessert werden kann. Als unterstützende Daten werden die Punkte auf den Fingerlinien kodiert, die an Stellen hoher Krümmung liegen. Diese ausgezeichneten Punkte können genutzt werden, um die Ausrichtung der Minutien-Punktewolke in einer Vorverarbeitung zu erreichen.

Bild 5 | Fuzzy Vault Verfahren mit Zusatzinformation (Enrolment)

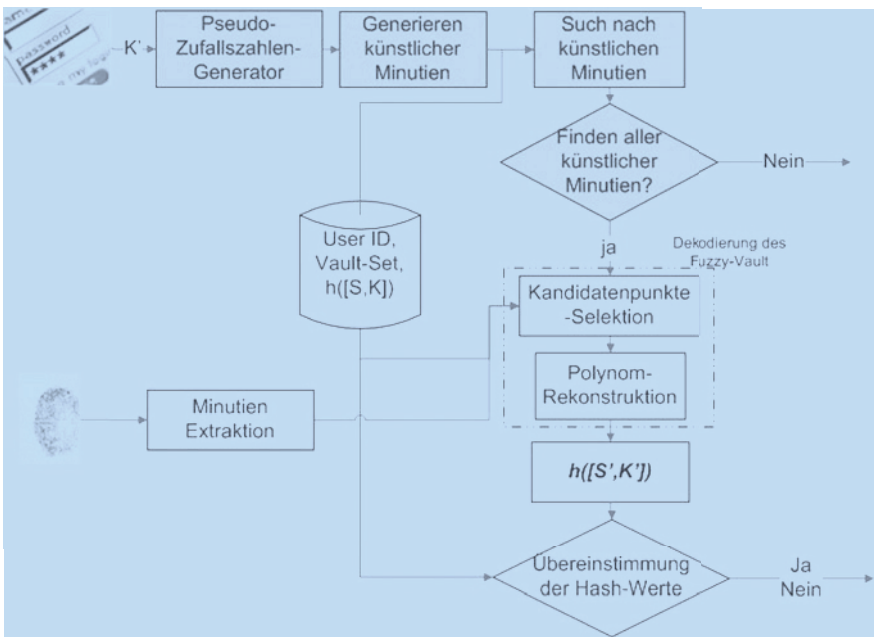


2.2 Integration von Zusatzinformation

Die Integration von Zusatzinformationen in Template-Protection-Verfahren ist sehr wohl geeignet, sowohl die Sicherheit als auch die Erkennungsleistung des gesamten biometrischen Systems zu erhöhen. Unter Zusatzinformation wird dabei im folgenden Text ein Passwort oder eine PIN verstanden.

Ausgehend von minutienbasierten Fingerabdruckerkennungssystemen auf Basis des Fuzzy-Vault-Verfahrens wird die Zusatzinformation als Initialisierungswert (Seed) für die Erzeugung einer Pseudo-Zufallszahl benutzt, um so mehrere unabhängige künstliche Minutien zu erstellen. Somit wirkt sich eine Verlängerung des Passworts nur auf den Initialisierungswert des Pseudo-Zufallszahlen-Generators für die Erzeugung unabhängiger künstlicher Minutien aus. Die Anzahl der künstlichen Minutien ist ein Systemparameter, der unabhängig von der Länge des Passworts ist. Die Anzahl der künstlichen Minutien darf dabei jedoch nicht die Anzahl der Polynom-Koeffizienten übersteigen, da sonst keine biometrische Infor-

Bild 6 | Fuzzy Vault mit Zusatzinformation (Verifikation)



mation bei der Verifikation heran gezogen werden müsste.

Im Enrolment-Prozess werden neben den extrahierten echten Fingerbildminutien auf Basis der Zusatzinformation K mithilfe des Pseudo-Zufallszahlen-Generators künstliche Minutien erzeugt. Dabei muss jedoch ein gewisser Abstand zwischen den unabhängigen künstlichen Minutien und den Minutien im Fingerbild gewahrt werden.

Dies ist notwendig, da Polynomprojektion und -rekonstruktion wie beim Fuzzy Vault-Verfahren in [NJP07] aus Effizienzgründen in einem kleinen endlichen Körper, zum Beispiel in GF(2^4) mit 16 Bit, durchgeführt werden. Der Mindestabstand stellt sicher, dass alle Minutien nach der erforderlichen Quantisierung als verschiedene 16-Bit-Werte dargestellt werden.

Der Mindestabstand wird dabei in x- und y-Richtung eingehalten. Nachdem die künstlichen Minutien (Artificial Points) erzeugt wurden, werden sie mit den echten Minutien des Fingerabdruckbilds (Enrolment Points) fusioniert – und als Eingabedaten für das normale Fuzzy-Vault-Verfahren genutzt. Im zweiten Schritt wird eine geheime Zeichenkette S zufällig generiert und daraus das Polynom des Fuzzy-Vault-Verfahrens gebildet. Die echten und künstlichen Minutien werden nun als Zahlen im endlichen Körper dargestellt (nun als α-Werte bezeichnet), wo sie anschließend auf das Polynom projiziert werden (β-Werte). Dabei wird der α-Wert im endlichen Körper wie folgt als 16-Bit-Wert generiert: Die x- und die dazugehörigen y-Minutienkoordinaten werden jeweils auf 8-Bit diskretisiert und anschließend zu einem 16-Bit-Wert aneinandergereiht. Dies entspricht somit einer Zahl im endlichen Körper. Anschließend werden die zufälligen Spreupunkte generiert, wobei diese ebenfalls einen Abstand zu den echten und künstlichen Minutien wahren müssen. Zu diesen Spreupunkten werden künstliche Polynom-Projektionen im endlichen Körper generiert, die jedoch nicht auf dem Fuzzy-Vault-Polynom liegen. Die nun vorhandenen echten und künstlichen Minutien und sonstige Spreupunkte mit deren zugehörigen Projektionen werden nun zufällig verwürfelt und bilden zusammen ein Vault Set. Weiterhin wird ein aus dem binarisierten Geheimnis S und dem Passwort K gebildeter Hash-Wert $h([S,K])$ zusammen mit einer benutzerspezifischen ID (User ID) und dem Vault Set gespeichert. In diesem Verfahren ist es möglich, dass unterschiedliche Passwörter die gleichen künstlichen Minutien generieren, da die Reihenfolge der Minutien keine Rolle spielt.

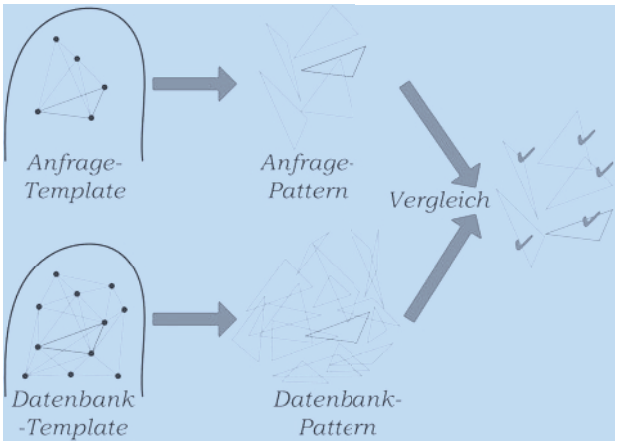
Bei der Verifikation werden anhand der Zusatzinformation K' erneut künstliche Minutien generiert. Diese stimmen bei der Verwendung derselben Zusatzinformation mit den künstlichen Minutien des Enrolments überein. Anhand der User ID wird nun das passende Vault Set geladen und die in den endlichen Körper übertragenen neu generierten künstlichen Minutien mit allen Punkten im Vault Set verglichen. Wenn nicht alle künstlichen Minutien exakt gefunden werden können, wurde die falsche Zusatzinformation für die Erzeugung genutzt und die Verifikation wird vorzeitig als nicht erfolgreich abgebrochen. Verläuft der Vergleich positiv, wird versucht, das Polynom zu rekonstruieren. Dazu werden die Minutien eines Verifikations-Fingerbilds mit den Minutieninformationen im Vault Set (ohne die künstlichen Minutien) verglichen. Die dabei übereinstimmenden Minutieninformationen werden als Index-Liste vom Minutienkomparator zurückgegeben. Die den Indizes entsprechenden Minutieninformationen des Vault Sets dienen als Kandidaten für echte Minutien entsprechend denen im Enrolment-Prozess. Zusammen mit den künstlich erzeugten Minutien wird nun versucht, das Polynom zu rekonstruieren. Dabei muss die Anzahl der übereinstimmenden echten zusammen mit den künstlichen Minutien mindestens der Zahl der Koeffizienten des Fuzzy-Vault-Polynoms entsprechen, um dass Polynom erfolgreich rekonstruieren zu können. Dabei findet die Rekonstruktion im endlichen Körper statt. Dazu wird die Lagrange'sche Interpolation verwendet. Wenn der Minutienkomparator jedoch mehr Kandidatenpunkte als notwendig extrahiert, müssen mögliche Kombinationen der Projektionen in den endlichen Körpern getestet werden, um das Polynom erfolgreich zu rekonstruieren. Konnte das Polynom rekonstruiert werden, wird mit dessen Hilfe das binarisierte Geheimnis S' berechnet. Zusammen mit der binarisierten Zusatzinformation K' wird erneut ein Hash-Wert gebildet und mit dem gespeicherten Hash-Wert verglichen. Stimmen beide überein, wurde das richtige Polynom rekonstruiert und die Verifikation war erfolgreich.

Im Projekt wurde neben der allgemeinen Evaluierung auch die Sicherheit dieses hier vorgestellten neuen Verfahrens überprüft. Es zeigte sich, dass neben der Polynomrekonstruktion auch der Aufwand, das Passwort richtig zu schätzen, die Sicherheit erhöht. Ebenfalls wurde gezeigt, dass das Verfahren eine verbesserte Resistenz gegen Verknüpfungsangriffe bietet.

3 Anwendung in Identifikationssystemen

Dieser Abschnitt betrachtet die Nutzung geschützter Templates in der Anwendung von Identifikationsverfahren im Kontext zentraler Datenbanken. Die Pseudonymisierung der Biometriedaten wird dabei durch das Template Protection Verfahren Fuzzy Vault [JS02] gewährleistet. Da beim Einsatz von Template Protection Verfahren der Vergleich zweier biometrischer Referenzen komplexer im Rechenaufwand ist, wurden verschiedene Verfahren zur effizienten Unterstützung des Identifikationsprozesses in großen Datenbanken analysiert. Zwei ausgewählte Ansätze werden hier exemplarisch erläutert. Bei den biometrischen Daten handelt es sich dabei wie im vorherigen Abschnitt um Referenzen, die aus Fingerabdrücken bzw. den daraus extrahierten Minutienmengen abgeleitet wurden. Im Rahmen des Vorgängerprojektes BioKeyS-Multi [KMN09] wurde bereits ein auf dem Fuzzy-Vault Ansatz basierendes Verifikations-System implementiert. Eine direkte, naive Übertragung dieses Ansatzes auf die Aufgabe der Identifikation bewirkt im schlechtesten Fall ein Durchsuchen der gesamten Datenbank, bis eine Entscheidung getroffen werden kann. Diese Vorgehensweise ist bei großen Datenbeständen von mehreren Millionen Referenzen nicht zielführend. Die im Folgenden vorgeschlagenen Verfahrensweise sollen nun dazu dienen, diesen Prozess dahingehend zu beschleunigen, dass die zu untersuchenden Datenbankreferenzen entsprechend einer gewissen Präferenz betrachtet werden. Der Ansatz verfolgt das Ziel, dem anfragenden Anwender eine bestimmte Reihenfolge der gespeicherten Datenbankreferenzen anzubieten, in der anschließend der Verifikationsabgleich durchgeführt werden soll. Dabei kommt es darauf an, dass trotz vielfältiger Approximationstechniken die tatsächliche Datenbankreferenz in dieser

Bild 7 | GeoMatch Ansatz anhandvon Dreiecken



Rangliste eine möglichst frühe Position einnimmt.

Im Allgemeinen liegen für die extrahierten Minutien eines Fingerabdrucks eine Vielzahl von Informationen, wie beispielsweise die Richtung, der Typ oder gegebenenfalls der Ridge-Count zu benachbarten Minutien vor. Im Folgenden wird für die Minutieninformationen eine Beschränkung auf die Ortskoordinate vorgenommen. Ausgangssituation für die Verfahren ist somit eine Menge von Datenbankreferenzen $R \in DB$ (im Folgenden Datenbanktemplates genannt) sowie eine Anfrageprobe Q, welche jeweils eine Menge von 2-dimensionalen Objekten $m=(m_x,m_y)$ beinhalten. Im Falle der Anfrageprobe Q handelt es sich bei diesen 2-dimensionalen Objekten ausschließlich um Minutien. Im Gegensatz dazu enthält ein Datenbanktemplate $R \in DB$ neben den Minutien zusätzlich eine große Menge an zufällig eingestreuten Chaff-Points, sodass dieses im Allgemeinen um einen gewissen Faktor k größer ist, als eine Anfrageprobe. Für jedes enrolte Subjekt S existieren Templates zu allen 10 Fingern $S_i=\{R_{i,1},...,R_{i,10}\}$. Da für jedes Template der zugehörige Fingertyp $T=\{1,...,10\}$ bekannt ist, kann angenommen werden, dass alle Fingertypen ($t \in T$) separat verwaltet werden.

3.1 Das GeoMatch Verfahren

Bei dem Verfahren GeoMatch handelt es sich um eine Filterstruktur, welche lediglich approximierte Ähnlichkeiten zwischen einer Anfrageprobe und einem Datenbanktemplate berechnet. Um gegen globale Rotationen sowie Verschiebungen des Anfragesamples robust zu sein, werden bei diesem Ansatz keine globalen Lagepositionen der Minutien für den Vergleich verwendet, sondern lediglich relative Positionen der Minutien zueinander. Für diese relativen Lagepositionen eignen sich geometrische Figuren wie n-Ecke, durch welche die relative Anordnung von n Punkten zueinander eindeutig beschrieben wird. Dieser Ansatz sieht vor, aus den jeweils zu vergleichenden Daten (Templates versus Probe) n-Ecke zu extrahieren und anschließend einen Vergleich lediglich basierend auf diesen n-Ecken durchzuführen (vergleiche Bild 7). Werden genügend viele n-Ecke einer Anfrageprobe im Vergleichsobjekt der Datenbank wiedergefunden, so kommt das betreffende Datenbanktemplate für den genauen Abgleich in Frage. Je größer die n-Eck-Übereinstimmug ist, desto wahrscheinlicher ist eine tatsächliche Templateübereinstimmung und entsprechend höher ist daher eine vorzunehmende Priorisierung für den genauen Abgleich (Ranking).

Die Anzahl aller möglichen n-Ecke in einem Template mit m 2-dimensionalen Objekten beträgt;

$$\binom{m}{n} = \frac{m!}{n! * (m-n)!}$$

Diese Anzahl wächst bei konstantem m exponentiell in n. Um den Platzbedarf für ein Datenbankpattern, sowie die benötigten Vergleichsoperationen gering zu halten, sollte n daher eher klein gewählt werden. Für steigendes n besitzt ein n-Eck indessen eine bessere Beschreibungsgenauigkeit, was tendenziell ein größeres n befürwortet. Für diesen Ansatz wurden Dreiecke (n=3) gewählt, welche sich mit 3 Werten, z.B. zwei Seitenlängen und einem Winkel, exakt beschreiben lassen. Ein Anfrage- sowie ein Datenbankpattern ist somit eine Menge von Dreiecken, wobei jedes Dreieck wiederum ein dreielementiges Tupel ist. Für einen exakten Vergleich von Anfrageprobe mit einer Datenbankreferenz ist eine diskriminative Beschreibung entsprechend der eines vollständigen Graphen für jedes Pattern ausreichend.

Um einen vollständigen Graphen zu rekonstruieren, sind jedoch nicht alle vorkommenden n-Ecke notwendig. Da durch diesen Ansatz ferner kein exakter Vergleich, sondern lediglich ein approximativer Vergleich angestrebt wird, kann die Menge der Dreiecke im Anfrage-, sowie im Datenbank-Pattern im Allgemeinen stark eingeschränkt werden. Dies ist z.B. durch Festlegen einer oberen, sowie einer unteren Schranke für die Seitenlängen eines Dreiecks möglich.

3.2 Das BioSimJoin Verfahren

Alternativ zum GeoMatch-Ansatz wurde im Projekt das BioSimJoin-Verfahren entwickelt. Ausgangssituation für dieses Verfahren ist ein Datenraum, in dem die Minutien aller Personen verschleiert, also mit zusätzlichen Chaff-Points gespeichert sind. Dabei werden die Minutien und Chaff-Points verschiedener Fingerinstanzen, also beispielsweise Daumen und Zeigefinger, in separaten Datenräumen gespeichert. Die Chaff-Points repräsentieren zufällige Punkte aus dem Raum der potenziellen Minutien, die zusammen mit den echten Minutien abgespeichert werden und diese damit verschleiern. In einem ersten Schritt wird nun zu jeder Minutie der angefragten Person eine Bereichsanfrage mit Radius r durchgeführt. In diesem Fall soll also z.B. eine Anfrage der Form „Finde alle Minutien bzw. Chaff-Points, deren Position im Bereich [225...250] liegt." beantwortet werden. Dieses Vorgehen ist schematisch in Bild 8 dargestellt. In diesem Beispiel enthält der angefragte Fingerabdruck drei unterschiedliche Minutien m_1, m_2 und m_3. Für jede dieser Minutien m_i werden diejenigen Minutien, aber auch Chaff-Points bestimmt, die sich innerhalb des Bereiches mit Radius r um m_i befinden. Die räumlichen Koordinaten (hier X und Y) zweier Minutien m_a und m_b unterscheiden sich höchstens um einen euklidischen Abstand r. Dieses Distanzmaß definiert sich wie folgt:

$$dist(m_a - m_b) = ((m_{a\_X} - m_{b\_X})^2 + (m_{a\_Y} - m_{b\_Y})^2)^{0,5}$$

Bild 8 | Verfahren BioSimJoin

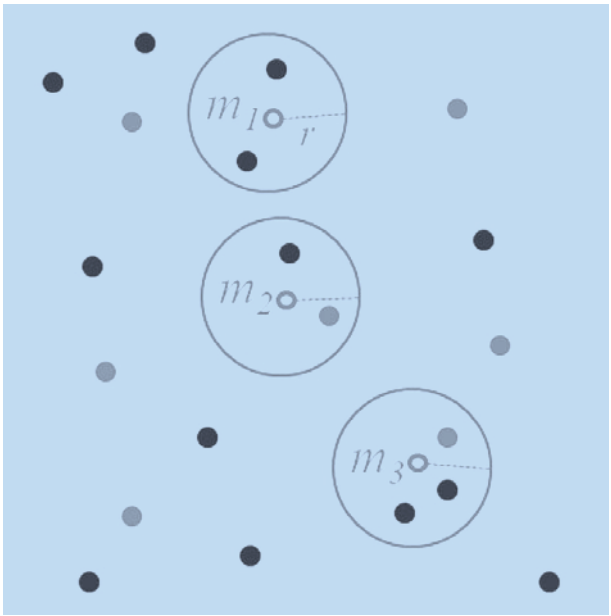
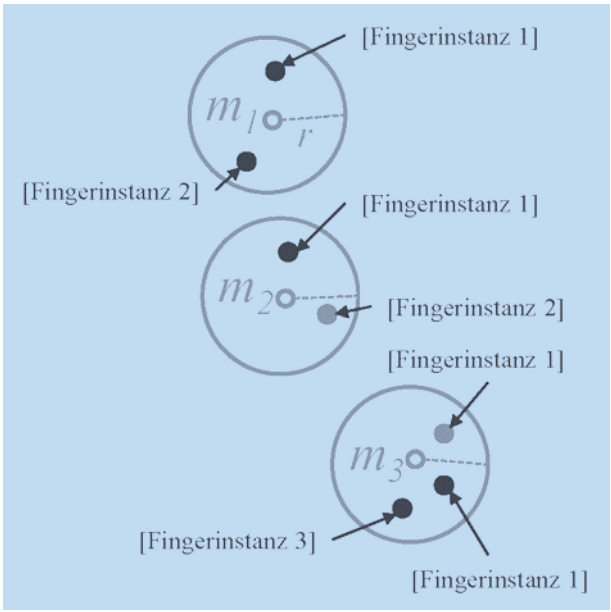


Bild 9 | Information der Treffer für BioSimJoin



Dieser Bereich ist durch die Kreise um m_i dargestellt. „Echte" Minutien sind durch schwarze Punkte dargestellt, graue Punkte repräsentieren Chaff-Points. Um die Sicherheit der biometrischen Merkmale nicht zu gefährden, liegt diese Information dem Verfahren BioSimJoin allerdings nicht vor. Vielmehr werden sowohl Minutien, als auch Chaff-Points in gleicher Weise behandelt. Für jede Minutie m_i kann anschließend eine Liste von Minutien bzw. Chaff-Points erstellt werden, die sich innerhalb des Bereichs um m_i befinden. Für jede solche Minutie bzw. jeden Chaff-Point ist dabei bekannt, von welchem Subjekt sie bzw. er stammt. Noch einmal sei darauf hingewiesen, dass diese Informationen sowohl für Chaff-Points als auch für echte Minutien vorliegen, da Chaff-Points lediglich dazu dienen, die biometrischen Merkmale zu verschleiern. Sie dürfen sich daher nicht in ihrer Darstellung von echten Minutien unterscheiden. Ansonsten wäre für einen potenziellen Angreifer ohne weiteres erkennbar, um welche Art von Merkmal es sich handelt. Diese Repräsentation wird in Bild 9 genauer erklärt.

Für die Minutie m_3 des angefragten Subjektes (natürliche Person) konnten beispielsweise zwei echte Minutien und ein Chaff-Point innerhalb eines Radius r als Treffer identifiziert werden. Der graue Chaff-Point stammt von einer Fingerinstanz der Person 1. Die linke Minutie bezieht sich auf den Fingerabdruck der Person mit der Identifikationsnummer 3 und die rechte Minutie gehört, wie auch der Chaff-Point zu Person 1. Insgesamt ergeben sich über alle drei Minutien der angefragten Person hinweg folgende Kandidaten: Vier Treffer für Person 1, zwei Treffer für Person 2 und ein Treffer für die Person 3. Diese Reihenfolge entspricht schließlich dem Ergebnis des Verfahrens BioSimJoin. Es dient dazu, dem Nutzer eine effiziente Identifikation zu ermöglichen, da es sich in diesem Beispiel sehr wahrscheinlich bei Person 1 um die angefragte Person handelt. Unsere Testergebnisse zeigen, dass der Suchraum durch BioSimJoin stark eingeschränkt werden kann. Bild 10 fasst den algorithmischen Ablauf nochmals schematisch zusammen. Die Methode `filter` von BioSimJoin erhält zwei Parameter; zum einen alle Minutien des angefragten Subjektes und zum anderen den Radius r, der die Bereichsanfrage spezifiziert. Zu Beginn des Algorithmus wird eine Datenstruktur `candidates` erstellt, die für alle Subjekte der Referenzdaten die Anzahl der Treffer mit den Minutien des angefragten Subjektes speichert. Diese Datenstruktur wird mit jeweils 0 Treffern je Subjekt initialisiert. Des Weiteren liegt dem Algorithmus eine Datenstruktur `minutiaeDB` vor, in der sowohl Minutien als auch Chaff-Points aller Subjekte der

Bild 10 | Algorithmischer Ablauf des Verfahrens BioSimJoin

```
Algorithmus filter_BioSimJoin(minutiae_query, r)

Eingabe: Minutien der Anfrageperson
         Radius der Bereichsanfrage

candidates        = [(p_1,0), (p_2,0), …, (p_n,0)]
minutiae_DB       = [(x_1,y_1,p_1), (x_2,y_2,p_2), …, (x_m,y_m,p_m)]

FOR EACH Minutia m_q IN minutiae_query DO{
    FOR EACH Minutia m_DB IN minutiae_DB DO{
        IF m_q.fingertype = m_DB.fingertype
        AND dist(m_q,m_DB) ≤ r DO{
                candidates.increment(m_DB.p)
        }

    }
}
Ausgabe: Kandidatenliste sortiert nach Treffer-Werten
```

Referenzdaten gespeichert sind. Für jede solche Minutie sind die Angaben über X- und Y-Koordinate, sowie die Identifikationsnummer des entsprechenden Subjektes verfügbar, zu der die Minutie gehört. Für jede Minutie des angefragten Subjektes wird anschließend der euklidische Abstand zu allen Minutien bzw. Chaff-Points der Referenzdaten des passenden Fingertyps bestimmt. Falls dieser den Wert von *r* nicht überschreitet, wird ermittelt, von welchem Subjekt diese Minutie bzw. der Chaff-Point stammt. Die Anzahl der Treffer dieses Subjektes kann somit um 1 erhöht werden. Nachdem die äußere Schleife vollständig durchlaufen wurde, speichert `candidates` die Anzahl der tatsächlich ermittelten Treffer für jedes Subjekt der Referenzdaten. Abschließend wird diese Liste nach Treffern absteigend sortiert und von dem Verfahren BioSimJoin als Ergebnis zurückgeliefert.

4 Schlussfolgerung

In diesem Beitrag wurde zunächst eine Übersicht über mögliche Template-Protection-Verfahren zum Schutz biometrischer Merkmale in Verbindung mit einer Referenzarchitektur gegeben. Dann wurde vorgestellt, wie das bekannte und etablierte Fuzzy-Vault-Verfahren um Zusatzinformationen erweitert werden kann und zudem eine Anpassung für Identifikationslösungen gestaltet werden könnte. Die für die Identifikation in diesem Beitrag vorgestellten Verfahren GeoMatch und BioSimJoin versuchen durch eine approximierte Ähnlichkeitsberechnung einer Anfrageprobe mit Elementen der Datenbank ein Ranking der Datenbanktemplates zu erstellen. Auf diese Weise kann der genaue, aber aufwändige Authentifikationsvergleich zunächst auf Datenbanktemplates mit einer hohen Trefferwahrscheinlichkeit ausgeführt werden, wodurch die Anzahl der insgesamt zu betrachteten Datenbanktemplates sehr stark eingeschränkt wird. Bei beiden Verfahren ist ein weiterer Geschwindigkeitsvorteil zu erzielen, sofern eine geringere Approximationsgenauigkeit akzeptabel ist. Da in diesen Verfahren dennoch alle paarweisen Vergleiche zwischen Anfrageprobe und Datenbanktemplates durchgeführt werden müssen, um ein Ranking zu erstellen, ist hier immer eine linear wachsende Laufzeit bei steigender Datenbankgröße zu erwarten. Um eine solche linear steigende Laufzeit zu vermeiden, sind Indexstrukturen unverzichtbar. Für Bereichsanfragen mehrdimensionaler Objekte, wie sie durch die unscharfe Suche auf Minutien erforderlich sind, eignen sich insbesondere R- bzw. R*-Bäume. Einen entsprechenden Ansatz, sowie entsprechende experimentelle Beobachtungen, zeigt hier die indexierte Variante von BioSimJoin, der Ansatz BioSimJoin* [BK2]. Für den Fall, dass eine Identifikation des Anfragesubjektes nicht möglich ist, wird, unabhängig vom zugrunde liegenden Identifikationsverfahren, ohne weitere Randbedingungen das gesamte Ranking und damit die gesamte Datenbank für den Authentifikationsvergleich betrachtet. Es ist daher empfehlenswert ein Schwellwertkriterium zu definieren, durch welches ein vorzeitiger Abbruch der Identifikationslösung möglich ist, falls die durch die vorgestellten Verfahren approximierte Ähnlichkeit zu gering ist. BioSimJoin ermittelt ein Ranking an Position 44 bei einer Datenbankgröße von 100 Subjekten in 1.465 ms. Da bei BioSimJoin keine Rotationen bzw. Verschiebungen gezielt berücksichtigt werden, ist es dem Verfahren GeoMatch hinsichtlich Laufzeit überlegen, kann jedoch bezüglich Effektivität keine [BK2] vergleichbaren Ergebnisse liefern. In den Identifikationslösungen wurden keine zusätzlichen biometrischen oder personenbezogenen Informationen verwendet. Die vorgestellten Suchverfahren ermöglichen somit den Einsatz des FuzzyVault in Identifikationslösungen mit einem einer Verifikationslösung vergleichbarem Schutz der Privatsphäre.

Literatur

[BK2] Bundesamt für Sicherheit in der Informationstechnik: BioKeyS PilotDB Teil 2 (Projekt Template Protection). Abschlussbericht, in Vorbereitung.

[Rat01] N.K. Ratha, J.H. Connell und R. Bolle: Enhancing security and privacy of biometric-based authentication systems. IBM Systems Journal, Vol. 40, Issue 3, 2001.

[Lin03] J.-P. Linnartz und P. Tuyls: New shielding functions to enhance privacy and prevent misuse of biometric templates. In Proc. Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA), LNCS 2688, Springer, 2003.

[Hao06] F. Hao, R. Anderson und J. Daugman. Combining crypto with biometric effectively, IEEE Trans. Comp. 55, 2006.

[Kor08] U. Korte, M. Krawczak, J. Merkle, R. Plaga, M. Niesing, C. Tiemann, H. Vinck, und U. Martini: A cryptographic biometric authentication system based on genetic fingerprints. In Proc. Sicherheit 2008, LNI 128, GI, 2008.

[Tuy05] P. Tuyls, A. Akkermans, T. Kevenaar, G.J. Schrijen, A. Bazen, R. Veldhuis: Practical biometric template protection system based on reliable components. In Proc. Audio- and Video-Based Biometric Person Authentication (AVBPA), LNCS 3546, Springer, 2005.

[ISOtp] ISO/IEC FDIS 24745 Information technology – Security techniques – Biometric information protection, 2010.

[JW99] A. Juels und M. Wattenberg: A fuzzy commitment scheme. In Proc. ACM Conf. Comp. and Comm. Security, 1999.

[DRS04] Y. Dodis, L. Reyzin, A. Smith: Fuzzy extractors: How to generate strong keys, In Proc. Advances in Cryptology – EUROCRYPT 2004, LNCS 3027, Springer, 2004.

[NJP07] K. Nandakumar, A.K. Jain, S. Pankanti: Fingerprint-based fuzzy vault: Implementation and performance. IEEE Trans. Information Forensics and Security, Vol. 2, No 4, 2007.

[S79] A. Shamir: How to share a secret. Commun. ACM 22, 1979.

[JS02] A. Juels und M. Sudan: A fuzzy vault scheme, In Proc. IEEE Int. Symp. Information Theory, 2002.

[UPJ05] U. Uludag, S. Pankanti und A.K. Jain: Fuzzy vault for fingerprints. Proc. Int. Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA), LNCS 3546, Springer, 2005.

[KMN09] U. Korte, J. Merkle und M. Niesing: Datenschutzfreundliche Authentisierung mit Fingerabdrücken, in DuD – 1/2009.