A Discussion on Privacy Needs and (Mis)Use of Biometric IT-Systems

Arslan Brömme

Faculty of Informatics
University of Hamburg
broemme@informatik.uni-hamburg.de

Abstract. This paper discusses aspects of privacy needs and (mis)use of biometric IT-systems along a model for the classification of biometric databases including biometric characteristics, biometric signatures, personal data, and access control mechanisms. A scenario-based discussion of privacy needs, which reveals that a database-organized access to biometric raw data is a main threat to privacy, results in general and technical design requirements for biometric IT-systems.

Keywords. biometrics, privacy, biometric IT-systems, informational privacy, identification, biometric database, biometric characteristic, biometric signature, personal data, iris-biometrics, security

1 Introduction

In a society where techno-economical orientation leads to the construction of biometric IT-systems to measure our physiology and behavior for the purpose of authentication, the discussion considering privacy and (mis)use of biometrics is of particular importance [1–5]. With respect to privacy, design requirements formulated today should affect the implementation of tomorrow's biometric technology.

Privacy is everyone's fundamental human right, which is documented in the *Universal Declaration of Human Rights* by the General Assembly of the United Nations [6]. In this paper a definition of privacy by Alan Westin will be used: "Privacy is the claim of individuals, groups, or institutions to determine for themselves, when, how and to what extent information about them is communicated to others" [7]. In [8] basic privacy principles are formulated which summarize the most essential privacy requirements. Concerning requirements for biometric IT-systems the discussion will focus on the privacy principles of purpose binding and necessity of data collection. The principle of purpose binding limit the subsequent use of personal data to the specified purposes. The principle of necessity of data collection means to avoid or at least to minimize personal data within an IT-system.

Each human being has static and dynamic biological characteristics, which are covered to be private in the above sense of privacy. In this paper, biological characteristics are further subdivided into static and dynamic physiological and behavioral characteristics: e.g. fingerprints, face proportions, hand geometries, iris and retina patterns (static-physiological), pupil contraction and dilatation (dynamic-physiological), and voice, gait, lip movement, keystroke dynamics (behavioral). It is assumed that these types of physiological and behavioral characteristics can be scanned by a special sensor system (digital or analog with an analog-digital-converter) and recorded digitally as time-series of measured values. If a CCD-camera is used as a sensor system, a time-series of single images, which are mathematically represented as multidimensional matrix of raw data, will be archived. Each single RGB (red-green-blue) picture can be handled as a 3-dimensional matrix with three layers for red, green and blue. For a video stream with n single RGB pictures the data structure is a 4-dimensional matrix (3-dimensional matrices as elements in an 1-dimensional array) or three 3-dimensional matrices for each color of the RGB-stream.

A scanned and recorded biological characteristic will be called a biometric characteristic, because the scanning process contains losses of information depending on the physical resolution of the measurement equipment. From images ready for computing, biometric signatures can be calculated [9], which are derived from biometric characteristics by using appropriate one-way calculation functions, e.g. hash functions. The original raw data cannot be reproduced from the hash values. The possibility to generate raw data, which has enough quality to be valid for biometric algorithms depends on the respective algorithm. The space complexity for a biometric signature is lower than the required space for the raw data plus sensor calibration data. Important to note is that one is able to derive any biometric signature if one owns a copy of the raw, scanned data.

The ability to derive any biometric signature from appropriate, high resolution raw data leads to fundamental privacy problems when these data are combined with personal data.

Section 2 gives some explanation of how biometric authentication systems work in principle, followed by an example of a recognition approach based on iris-biometrics. The different phases of the biometric authentication process generate different types of data. These types serve as classifiers for biometric databases as described in section 3. Section 4 discusses aspects of privacy needs for a privacy-fulfilling and -enhancing use of biometric databases resulting in general and technical design requirements for biometric IT-systems.

2 Biometric Authentication Systems - Example 'Iris-Biometrics'

Authentication by IT-systems can be done by means of one or more elements of the following sets of methods for proving the authenticity of a person who seeks access to an IT-system: user possession (e.g. key, chipcard), user knowledge (e.g. user identifier, password), user attribute (e.g. fingerprint, face proportion, iris and retina patterns), or user location (e.g. location by GPS, location in a defined physical area) [8, 10, 11].

Biometric authentication gives a higher assurance of a person's identity than the use of a password known to the IT-system and the user, unless both methods are attacked. Classical attacks on passwords and selection criteria for passwords can be found in [12, 13]. According to [14], attacks on biometric IT-systems can be classified using three basic categories: 1. sensor attacks (copy, falsification and similarity attacks), 2. data communication attacks (replay attacks), and 3. database attacks (integrity attacks).

The general process a person is subjected to in order to receive access to system resources by biometric authentication can be divided into the following phases:

1. **Enrollment**: The enrollment of a person is the process of scanning appropriate raw data from this person, the calculation of a biometric signature for the biometric authentication, and the storage of the relevant data in a database.
2. **Biometric authentication (1:n, 1:1)**: During biometric authentication a person's authenticity will be checked by an identification or verification comparison of the signature calculated from the current raw data to the signatures stored in the database of the biometric authentication system (BAS). Identification is a 1:n and verification is a 1:1 access to the table of biometric signatures.
3. **Authorization**: In the authorization phase the system gives implicit authorizations to the user with respect to strong and weak authorizations[1].
4. **Access to system resources**: The access to system resources will be granted by an access management system (AMS), which can be based on the concepts of mandatory access control (MAC), discretionary access control (DAC), or role-based access control (RBAC).

In the following the main aspects of an RBAC-style access management concept for biometric authentication is shown: From an instance of RBAC the set of roles, set of resources, and set of access rights can be derived as $ROLE := \{role_1, role_2, ..., role_i\}$, $RESOURCE := \{res_1, res_2, ..., res_j\}$, and $ACCESS\ RIGHT := \{ar_{\{\}}, ar_1, ar_2, ..., ar_k\}$. $ar_{\{\}}$ means no access right on the respective resource. The access rights can be formulated as sets of tuples for each role: $\{(role_1, res_1, \subset \{ar_{\{\}}, ar_{11}, ..., ar_{1k}\})$,..., $(role_1, res_j, \subset \{ar_{\{\}}, ar_{j1}, ..., ar_{jk}\})$ ... $(role_i, res_1, \subset \{ar_{\{\}}, ar_{11}, ..., ar_{1k}\})$,..., $(role_i, res_j, \subset \{ar_{\{\}}, ar_{j1}, ..., ar_{jk})\}$.

In the enrollment phase the personal data, raw data, and biometric signature of a person will be stored in a table $PERSONAL\ DATA \times RAW\ DATA \times BIOMETRIC\ SIGNATURE$: (($name_1$, $domicile_1$, age_1, sex_1, $income_1$), $raw\ data_1$, $signature_1$) ... (($name_n$, $domicile_n$, age_n, sex_n, $income_n$), $raw\ data_n$, $signature_n$).

[1] Authorizations derived by the access management system are called implicit, otherwise they are called explicit. An authorization is called strong if it cannot be overriden by other authorizations. If it can be overriden in accordance to a rule base, it is called weak.

During authorization an administrator defines sets of pairs $PERSON \times ROLE$ to bind roles to persons: $\{(id_1, role_1)$,..., $(id_1, role_i)\}$,..., $\{(id_n, role_1)$,..., $(id_n, role_i)\}$.

For illustrating the calculation of biometric signatures from recorded raw data the example of iris-biometrics will be used. Iris-biometrics is intended for usage in high security applications, because of the uniqueness of the genotypical iris patterns [15].

Iris-Biometrics. Iris-biometrics is a young field of research which is dedicated to the development of algorithms for measuring, classifying, and recognizing the patterns of the human iris for the usage in biometric authentication systems. For a detailed description of the biological function of the iris please refer to [16].

Figure 1 shows the main steps carried out by an iris-biometrics algorithm in order to receive a biometric signature of the human iris by analyzing its digital image. First, a scan of the eye with respect to lighting conditions (visible and infrared light, cornea reflections) and sensor calibration will be done. Next, the data of the iris will be extracted by finding the edges of the pupil and the iris in the image. With the help of image processing algorithms features of the iris will be found.

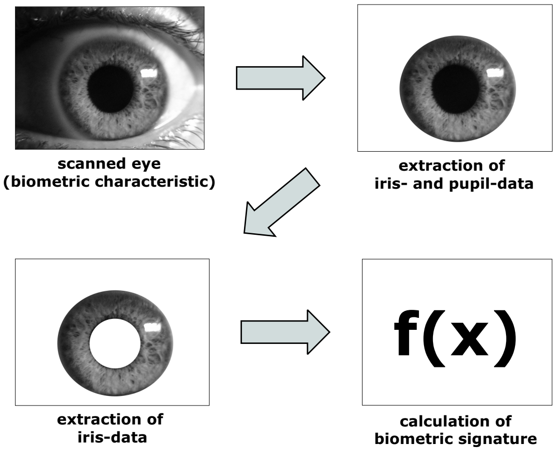

Fig. 1. Digital Iris Analysis

The measured features of the iris can be stored as *iris feature vector sets* (IFVS), which are invariant to the rotation of the iris and the dilatation and contraction of the pupil (figure 2). The angles between the vectors spanned from the center of the pupil are stored in an array. The sum of angle values yields 360 degrees.

For each feature a separate vector set will be created. With given vectors for each feature a matrix can be constructed, which is one possible data structure for the resulting biometric signature.

For instance, three IFVS are constructed on iris features represented here by ellipses (left), rectangles (middle) and triangles (right) in figure 2. The following arrays are derived: 'ellipse'-feature $[160, 10, 120, 30, 40]$, 'rectangle'-feature $[145, 70, 145, 0, 0]$, and 'triangle'-feature $[80, 120, 60, 100, 0]$. To guarantee that all arrays have the same length, an array containing fewer elements is filled up with zeros. The biometric signature is the matrix constructed from these single arrays: $[160, 10, 120, 30, 40; 145, 70, 145, 0, 0; 80, 120, 60, 100, 0]$.

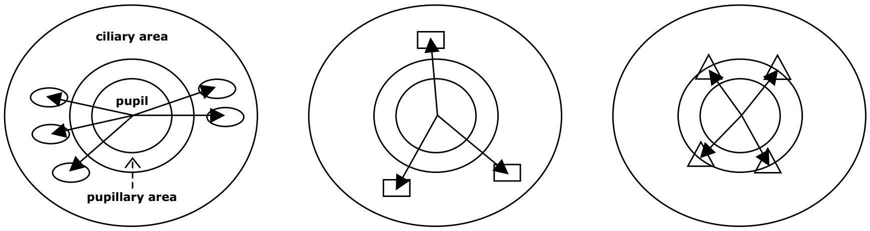

Fig. 2. Iris Feature Vector Sets (IFVS)

For the comparison of a biometric signature with the database, in the best case the complexity of comparisons is $O(n)$ and in the worst case it becomes $O(n^2)$, because cyclic comparisons are necessary if the iris data are rotated, e.g. $[160, 10, 120, 30, 40]$ is the same as $[10, 120, 30, 40, 160]$[2]. If the biometric signatures in the database are organized in a tree structure, the number of comparisons can be optimized to $O(log\ n)$ in the best case. The effect of the dilatation and contraction of the pupil can be neglected in this example, because it only results in a radial transposition of the features, the angles between the vectors remaining the same.

The approach described above, which is one subject of my research, is based on image processing by analyzing the static and dynamic aspects of the morphology of the human iris for biometric authentication. Another class of algorithms for calculating biometric signatures of the human iris is based on frequency analysis methods [15].

3 A Model for the Classification of Biometric Databases

In the following a model for the classification of biometric databases will be described (figure 3). The explanation of biometric databases is followed by the description of the characteristics of a person and of the access management

[2] To scan an iris twice in the same orientation is a difficult task for the cooperative person and the sensor system.

system (broken lined boxes in figure 3). In the absence of rules for access control of finer granularity on biometric data, depending on an instance of an access control concept (MAC, DAC, RBAC), it is possible to generate three types of partial biometric databases.

Biometric Databases. In this model a biometric database is defined as a database which holds data about biometric characteristics, biometric signatures, and personal data. The access control mechanism of such a biometric database is assumed to be rule-driven. The rules are derived from the instance of the access control concept being used.

A biometric database which subsumes biometric characteristics (raw data and calibration data), biometric signatures, personal data, and a rule-based access control mechanism is defined to be a complete biometric database. Another kind of biometric database are the partial biometric databases:

A partial biometric database represents a subset of the complete biometric database. Three types of partial biometric databases will be considered in this model. The first one includes biometric characteristics and biometric signatures. The second one consists of biometric characteristics and personal data. And the last one contains biometric signatures and personal data. The access to these three chosen types of partial biometric databases can be done without fine grained rules for access control. In this case only one general rule exists: $\forall persons$: $\exists$ *access to biometric database* $(person, data, access\ right) \wedge$ *access right* $=$ "*all*"

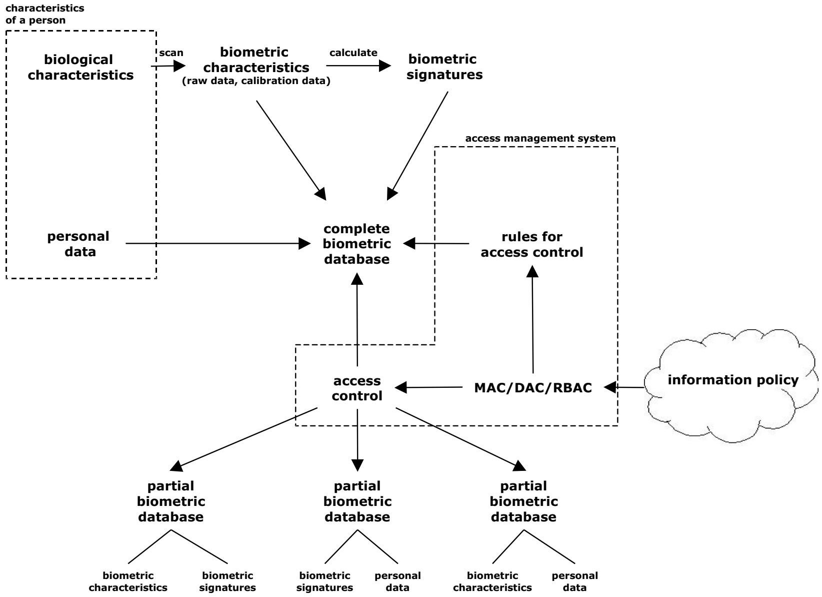
characteristics of a person
biological characteristics
scan
biometric characteristics (raw data, calibration data)
calculate
biometric signatures
access management system
personal data
complete biometric database
rules for access control
access control
MAC/DAC/RBAC
information policy
partial biometric database
partial biometric database
partial biometric database
biometric characteristics
biometric signatures
biometric signatures
personal data
biometric characteristics
personal data

Fig. 3. A Model for the Classification of Biometric Databases

Characteristics of a Person. Each human being has biological characteristics. In a systematic approach biological characteristics can be divided into static-physiological, dynamic-physiological, and behavioral characteristics (section 1). Biological characteristics can be damaged or lost e.g. by accidents and diseases.

From a technical point of view the biological characteristics can be divided into different sensor-dependent classes like visible characteristics (fingerprints, face proportions, hand geometries, iris and retina patterns, and lip movement), audible characteristics (voice), olfactory characteristics (odor of body parts), thermal characteristics and so on.

Usually personal data like given names, surname, age, sex and domicile are assigned to persons. Personal data exceeding basic information about a person are e.g. curriculum vitae, earning capacity and pecuniary circumstances. Very sensitive information about a person is e.g. diseases, consumption of drugs, alcoholism, and criminal record. In general, personal data in the context of informational privacy means any information concerning the personal or material circumstances of an identified or identifiable person [8].

In this model the above described information about a person is understood as personal data, and the physiological and behavioral characteristics of an individual as biological characteristics. Biological characteristics and personal data of an individual are subsumed as characteristics of a person.

Access Management System (AMS). For this model an access management system can be used which is an instance of any access control concept like MAC and DAC which can be a realization of various security policies like RBAC. The AMS regards to an information policy which influences the way the information from a biometric database needs to be handled in a company or organization.

4 Privacy Needs and Design Requirements of Biometric IT-Systems

Concerning privacy aspects of biometric IT-systems, already today, one is able to develop scenarios describing applications made possible through biometric identification having become a common technology.

In the following, four scenarios are introduced which can be mapped to the different types of (partial) biometric databases which were introduced in section 3. For each (partial) biometric database an adequate access control mechanism is assumed:

- database type 1: biometric raw data, biometric signatures, and personal data
- database type 2: biometric raw data and personal data
- database type 3: biometric signatures and personal data

Scenario #1 'Shoplifter'. In this scenario a supermarket equipped with surveillance cameras is considered. The high rate of shoplifting in the area of the supermarket is leading to cooperations between different supermarkets to minimize losses.

The supermarket managers hand out the biometric raw data from potential shoplifters to a service company. The service company calculates biometric signatures from the potential shoplifters and installs a network-based "shoplifter recognition system" in all participating supermarkets.

The service company maintains a huge database of biometric raw data and biometric signatures of potential shoplifters and identifies the persons entering a supermarket as "potential shoplifter" or "customer" in real-time. If a potential shoplifter enters a supermarket, the manager will be informed immediately.

⇒ A type 1 biometric database is used in this scenario. The service company manages a biometric database which includes biometric raw data and biometric signatures of potential shoplifters. A personal datum is given with the bit that someone is classified as customer or as potential shoplifter.

Scenario #2 'Multimodal Biometric Identification System'. This scenario assumes a biometric IT-system which is based on multimodal identification of persons. This biometric IT-system is the product of a today's typical startup-company and identifies persons by face proportions, lip movement, and voice.

The high rate of developing new algorithms and templates for face recognition leads to a problem with updates of the biometric signature list when new algorithms are used. To minimize acceptance problems concerning usability, the company integrated full raw data and personal data into the system. Every time a new algorithm or template is patched, the system enrolls automatically without the time-consuming participation of the users.

⇒ A type 2 partial biometric database is used in this scenario. The startup-company uses a biometric database with biometric raw data and personal data of the users. An upgrade to a complete biometric database is possible when biometric signatures are included.

Scenario #3 'Research Laboratory Access Control'. Assume a research institute which applies biometric recognition systems to regulate the access to its laboratories. An iris recognition system is used which stores only biometric signatures and personal data for the access control. After the enrollment of a person, the raw data is no longer necessary and will be destroyed.

⇒ A type 3 partial biometric database is used in this scenario. The research facility uses a partial biometric database which includes biometric signatures and personal data. An upgrade to a complete biometric database is not possible in the system, because the biometric raw data will be destroyed after the calculation of the biometric signatures.

Scenario #4 'Public Transportation System'. In this scenario a public transportation system is considered which uses biometric authentication at control gates to grant or deny admittance.

A biometric identification system for face recognition is used to identify people as passengers, who have paid for the usage of the transportation system. The enrollment procedure is part of the payment process. The raw data will be destroyed after the enrollment.

This system is designed for anonymous and pseudonymous purposes. For simple day usage only biometric signatures of passengers will be stored in the database. After exceeding the period of validity the biometric signature will be removed and deleted from the database. For subscribers personal data for bank transfers are stored. Additionally an unique pseudonym will be saved which can be used at the control gate for the faster verification comparison.

⇒ In maximum a type 3 partial biometric database is used in this scenario. In the case of anonymous usage the data fields for personal data are left empty and in the case of pseudonymous usage personal data for the bank transfer are stored additionally. An upgrade to a complete biometric database is not possible, because the raw data will be immediately deleted after the enrollment.

What is Misuse and What is Use of Biometric IT-Systems from the Point of View of Privacy? All four scenarios described have a strong individual motivation. In scenario #1 the supermarket managers are interested in protecting their property. In scenario #2 a company needs to sell products which are still under development, otherwise the company is not in the position to develop anything. In scenario #3 a research facility wants to control who enters its laboratories. In scenario #4 only passengers who have paid for the usage are allowed to enter the public transportation system.

Daily practice of privacy depends mainly on the cultural and constitutional reality of a state. From the *right of informational self-determination* point of view a classification of each scenario as a 'misuse'- or 'use'-case of biometric IT-systems can be derived. The term *right of informational self-determination* was defined by the German Constitutional Court in 1983 in connection with the Census Decision and can be translated as "informational self-determination, meaning the right of an individual to determine the disclosure and use of his personal data on principle at his discretion" [8].

Not only the German legislation represents a basis for this classification. In [17] a discussion of the technical aspects of biometrics with regard to the American, the European and the German legislation is given. In [3] the Canadian legislation is discussed in the context of privacy and biometrics.

By using the above definition it is immediately clear that a type 1 partial biometric database cannot be used without the consent of the person who is classified as "potential shoplifter" or "customer". The respective biometric IT-system is designed in a way that the consent is not a part of the system.

⇒ Scenario #1 is a misuse-case

The situation in scenario #2 is a different one. Like in scenario #1 the system is not designed in a way to integrate the consent of the person. But it is implicitly given within a working contract that the employee agrees to access control systems of the company if the collected data is only used for access control. Yet, an employee does not need to agree to a system which is not equipped with adequate algorithms for biometric authentication. The system collects more personal data than necessary for the specified purpose of biometric authentication for access control (privacy principle of necessity of data collection). Being a startup-company is no reason to run development cycles of the system by the cost of the right of informational self-determination.

⇒ Scenario #2 is a misuse-case

A very privacy-friendly case is given in scenario #3. No biometric raw data is collected in the database. For the biometric authentication it is necessary to produce biometric raw data, but only for the aim of calculating a biometric signature. After the calculation the raw data is deleted immediately.

⇒ Scenario #3 is a use-case

The most privacy-friendly case is given in scenario #4. There is a choice between an anonymous and pseudonymous usage of the system. No biometric raw data and personal data are collected in the database when the system is used in an anonymous way. For the pseudonymous usage the system needs personal data for faster admittance and bank transfer. In both cases the raw data is deleted immediately after the calculation of the biometric signature.

⇒ Scenario #4 is a use-case

Against the background of this scenario-based discussion it is concluded that the access to biometric raw data is a main threat to privacy and therefore the use of type 1 and type 2 partial biometric databases for biometric IT-systems should be avoided. Type 3 partial biometric databases conform to privacy needs and can be integrated into biometric IT-systems.

General and Technical Design Requirements for Biometric IT-Systems. From the preceding discussion we can derive the following general and technical design requirements for biometric IT-systems with regard to privacy:

- General Requirement: A biometric IT-system should be designed in a way to use biometric data in accordance to privacy needs.
- Technical Requirement #1: A biometric IT-system should not rely on a (partial) biometric database which stores biometric raw data.
- Technical Requirement #2: A biometric IT-system should not enable the usage of a biometric database outside the specified purpose of biometric authentication (privacy principle of purpose binding)[3].
- Technical Requirement #3: A biometric IT-system should not collect personal data which are unnecessary for the specified purpose (privacy principle of necessity of data collection).
- Technical Requirement #4: A biometric IT-system should use adequate algorithms for computing biometric signatures.

Further technical design requirements for biometric IT-systems can be derived from cryptographic biometrics [18, 19]. [18] describes a cryptographic implementation based on the "wallet with observer" architecture by Chaum and

[3] This includes the combination of a biometric database with other databases and the interpretation of the biometric data for medical or psychological purposes.

Pedersen with an observer-implanted tamper resistant biometric verification facility.

5 Summary

A model for the classification of biometric databases was presented to enable a scenario-based discussion of privacy needs regarding different types of (partial) biometric databases. The discussion reveals that a database-organized access to biometric raw data is a main threat to privacy. From a general point of view of privacy, a biometric IT-system should reduce or avoid the misuse of biometric databases. Technical design requirements for biometric IT-systems comprise the avoidance of biometric databases which contain biometric raw data, the avoidance of the usage of biometric databases outside the specified purpose, the avoidance of collecting unnecessary data for the specified purpose, and the usage of adequate algorithms and biometric signatures.

References

1. Woodward, John D.: Biometrics: Privacy's Foe or Privacy's Friend?, Proceedings of the IEEE, Vol. 85, No. 9, 1997
2. Cavoukian, Ann: Privacy and Biometrics - An Oxymoron or Time to Take a 2nd Look?, Information and Privacy Commissioner(IPC)/Ontario, Computers, Freedom and Privacy 98, Austin, Texas, 1998
3. Information and Privacy Commissioner (IPC)/Ontario: Privacy and Biometrics, Canada, September, 1999
4. George Tomko (Ed.): Biometrics as a Privacy-Enhancing Technology: Friend or Foe of Privacy?, Privacy Laws & Business 9th Privacy Commissioners' / Data Protection Authorities Workshop, Hotel Reyes Catolicos, Santiago de Compostela, Spain, September 15th, 1998
5. Woodward, John D.: Biometric Scanning, Law & Policy: Identifying the Concerns - Drafting the Biometric Blueprint, University of Pittsburgh Law Review, 1997
6. General Assembly of the United Nations: Universal Declaration of Human Rights, http://www.un.org/Overview/rights.htm, December 10th, 1948
7. Westin, Alan F.: Privacy and Freedom, Atheneum, New York, 1967
8. Fischer-Hübner, Simone: Privacy-Enhancing Design and Use of IT-Security Mechanisms, habilitation, Faculty of Informatics, University of Hamburg, 1999
9. Jain, Anil K., Bolle, Ruud and Pankanti, Shrath (Eds.): Biometrics - Personal Identification in Networked Society, Kluwer Academic Publishers, 1999
10. Silberschatz, Abraham and Galvin, Peter B.: Operating System Concepts, 4th edition, Addison-Wesley Publishing Company, 1994
11. Denning, Dorothy E. and MacDoran, Peter F.: Location-Based Authentication: Grounding Cyberspace for Better Security, Computer Fraud & Security, Elsevier Science Ltd., 1996
12. Pfleeger, Charles P.: Security in Computing, 2nd edition, Prentice-Hall International Inc., 1997
13. Eckert, Claudia: IT-Sicherheit - Konzepte, Verfahren, Protokolle, Oldenbourg Verlag, 2001

14. Bundesamt für Sicherheit in der Informationstechnik (BSI): Vergleichende Untersuchung biometrischer Identifikationssysteme - BioIS, http://www.bsi.org/, 2000
15. Daugman, John G.: Recognizing Persons by their Iris-Patterns, in: Jain, Anil K., Bolle, Ruud and Pankanti, Shrath (Eds.): Biometrics - Personal Identification in Networked Society, Kluwer Academic Publishers, 1999
16. Hart, William M. (Ed.): Adler's Physiology of the Eye - Clinical Application, 9th edition, Mosby-Year Book Inc., 1992
17. Gundermann, Lukas and Köhntopp, Marit: Biometrie zwischen Bond und Big Brother - Technische Möglichkeiten und rechtliche Grenzen, Datenschutz und Datensicherheit (DuD) - Recht und Sicherheit in Informationsverarbeitung und Kommunikation, Schwerpunkt: Biometrie, Volume 23, No. 3, 1999
18. Bleumer, Gerrit: Biometric yet Privacy Protecting Person Authentication, in Aucsmith, David (Ed.): Information Hiding 1998, LNCS 1525, 1998
19. Bleumer, Gerrit: Biometrische Ausweise - Schutz von Personenidentitäten trotz biometrischer Erkennung, Datenschutz und Datensicherheit (DuD) - Recht und Sicherheit in Informationsverarbeitung und Kommunikation, Schwerpunkt: Angewandte Biometrie, Volume 24, No. 6, 2000