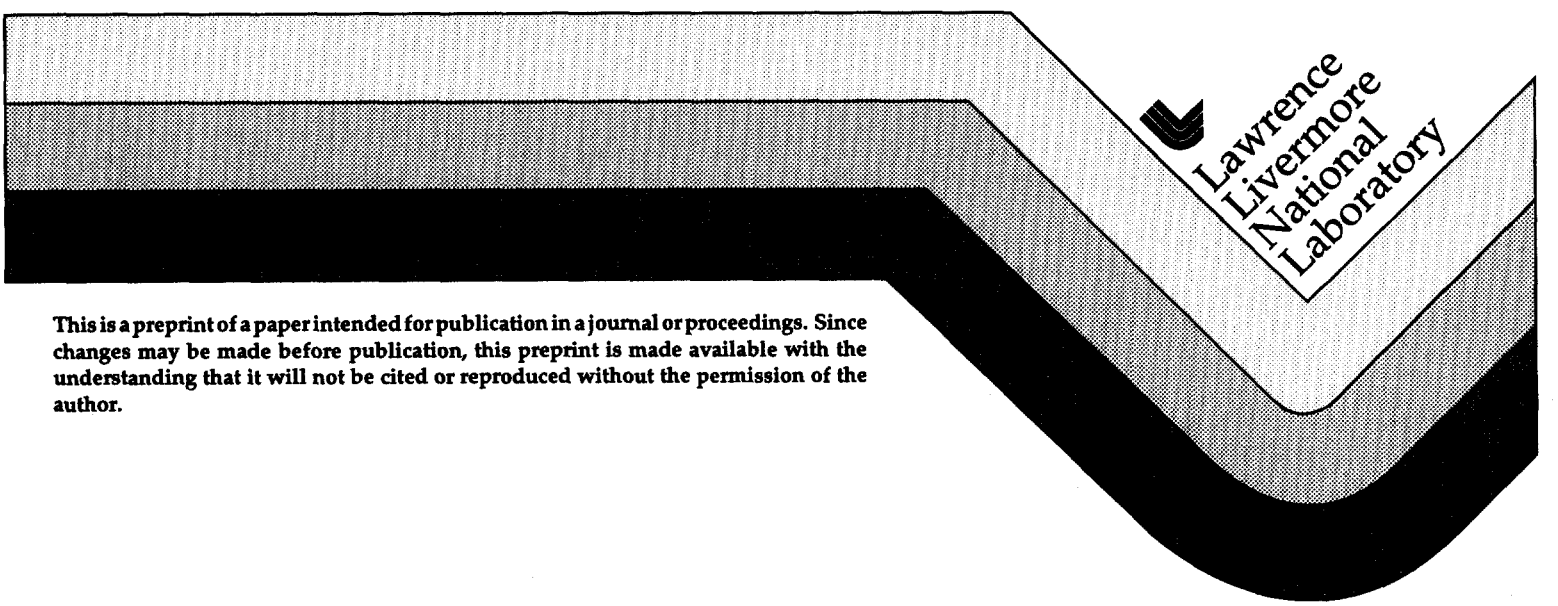
UCRL-JC-123435
PREPRINT

Remote Secure Proof of Identity Using Biometrics

S. K. Sengupta
P. Pearson
R. S. Strait

This paper was prepared for submittal to the
13th American Defense Preparedness Association Symposium & Exhibition
on Security Technology
Virginia Beach, Virginia
June 9-12, 1997

June 10, 1997



Remote Secure Proof of Identity Using Biometrics

Sailes K. Sengupta, Peter Pearson,
and R. Scott Strait

Abstract

Biometric measurements derived from finger- or voiceprints, hand geometry, retinal vessel pattern and iris texture characteristics etc. can be identifiers of individuals. In each case, the measurements can be coded into a *statistically unique* bit-string for each individual. While in electronic commerce and other electronic transactions the proof of identity of an individual is provided by the use of either public key cryptography or biometric data, more secure applications can be achieved by employing both. However the former requires the use of exact bit patterns. An error correction procedure allows us to successfully combine the use of both to provide a general procedure for remote secure proof of identity using a generic biometric device. One such procedure has been demonstrated using a device based on hand geometry.

Introduction

Electronic transactions depend on the proof of identity of an agent at a remote location where the proof is normally provided by the use of public key cryptography or the agent's biometric data. In the former case, the private key, typically a very long bit-string, is normally recorded on a card. The loss of the card is clearly a security risk. Biometric data on the other hand, are subject to statistical variations and to privacy concerns. Furthermore, storage of biometric data in any authentication system risks compromise of the data, and once the agent's biometric data are compromised, they become worthless as a means of generating a private key. Thus prudent security procedure requires that the biometric data should (a) never be stored and (b) be used only in encrypted form.

A system that uses the agent's biometric data as a private key can provide a higher level of security than either method alone while overcoming some of the limitations of both methods. There will be no need for the agent to carry the card containing the private key since they can be generated by the agent embedded within his body. Also there will be lesser concern about the compromise of biometric data as no such data will ever have to be exchanged.

However, cryptographic applications require exact bit-patterns of data. Thus for the biometric data to be useful in a scenario where cryptographic protocols need to be employed, an error correction mechanism has to be incorporated in some parts of the system. The number of bits that need to be corrected depends on the distributions of the Hamming distance between two biometric measurements from

(a) the same agent, and

(b) that from two different agents.

Clearly, it depends on the particular biometric device used as well. Furthermore, due to an inherent limit on the number of bits that can be corrected by any practical error correcting code (ECC) one must also strive for a stable representation of the biometric data so that bit strings derived from the same agent differ from one another in as few positions in them as possible. In the following sections, we describe the representation scheme and the error correcting code used for the specific device employed to demonstrate the feasibility of our algorithm. This will be followed by the description of a verification system using a generic biometric device. We conclude with indications for future work.

Data Representation

As indicated earlier, this is strictly device dependent. For the hand geometry device[1], we have considered eight possible representations of the 9 bytes of data corresponding to the 9 fields in the measurement data and are described below:

With the stability of the bit-string derived from biometric data as the main objective, we consider for all nine fields in the biometric record, eight possible representation schemes as binary integers depending on the following three levels of dichotomies:

(a) The least significant bit corresponds to the widest single-user variation or to one half of the same.

(b) The mean value for a single user may fall either on a transition or midway between two transitions.

(c) The fields are represented as simple binary integers or as gray-coded integers.

We decided on the representation that gives the maximum normalized separation of the distributions of the Hamming distance between two measurements of the same and of different individuals.

Table 1 shows some statistics of the Hamming distance distribution obtained from 56906 authentic pairs and 3570565 pairs with an impostor in each. Based on a statistical measure of cluster separation, the gray-coded integer representations faired much better than the binary integer representations. Among the Gray-coded representations, Representation 7 was judged the best and was in fact used for our study. It had the added advantage (for coding purposes) of having an exact 4 byte representation.

Table 1. Hamming distance statistics for authentic and impostors

Representation #	Pair	Min	Max	Mean	SD	Bits
1	authen.	0	18	4.7	2.9	30
	impost.	1	25	12.2	3.0	
2	authen.	0	24	8.5	3.6	39
	impost.	2	31	16.4	3.3	
3	authen.	0	21	4.5	2.8	32
	impost.	0	25	11.9	2.9	
4	authen.	0	25	8.4	3.6	39
	impost.	2	32	16.4	3.3	
5	authen.	0	10	2.7	1.5	30
	impost.	1	23	10.5	2.6	
6	authen.	0	16	5.2	2.0	39
	impost.	2	28	14.9	3.0	
7	authen.	0	11	2.6	1.5	32
	impost.	0	22	10.4	2.5	
8	authen.	0	16	5.2	2.1	39
	impost.	2	31	14.9	3.0	

Error Correcting Code

An interesting class of multiple-error-correcting binary codes is the Reed-Muller Code (RMC). Its importance derives from the fact that it has a decoding technique based on the majority logic. This technique is capable of easy implementation 1]. The characteristics for an RMC with parameters r and m, denoted R(r,m) are listed below:

Length of the codewords:	2^m
Number of information symbols:	The number of ways in which r or fewer objects may be selected from m objects.
Minimum Hamming distance:	2^{m-r}
Maximum number of bits corrected:	$2^{m-r-1} - 1$

We have used a R(2,5) code to be applied on the hand geometry data.

Public Key Cryptography Using Diffie-Hellman Protocol

We have used the public key algorithm [2] due to Diffie and Hellman. This cryptographic protocol is based on the difficulty of calculating discrete logarithms in a finite field. It is typically used for key negotiations between two individuals, say A and B. We have adapted it to generate a secure proof of identity scheme. In addition, we have used a one-way hash function called MD5 [3] for additional security. See [2] for the mathematics involved in Diffie-Hellman (DH) protocol. It requires the computation of a large prime p with certain specified properties and a large integer g. A and B have to agree in advance on the use of g and p in the protocol.

System Design and Implementation

We now create the design for a system that uses the four entities listed below as building blocks to provide remote secure proof of identity based on a person's biometric data used as his private key :

(a) The preferred representation scheme of the biometric data,

(b) the chosen ECC,

(c) the DH protocol, and

(d) the MD5 hash function

We will refer to the combination of HG reader and the computer with which it communicates as the Field Station (FS). In addition there is the Central System (CS) that communicates with the computer in the FS. In order to save operating time, the integers g and p are hard coded into the software to be used in both the FS and the CS. The two types of transaction, enrollment (E1-E4) and verification (V1-V5) are described next.

Enrollment

E1: The FS selects a 'true' random codeword C to be used as seed to generate a random number to be designated as the user's *secret exponent*.

E2: The FS prompts the user for a biometric device reading and reads the measurement into memory after giving it the chosen representation E.

E3: FS computes the *non-secret* "reference" value say R using C and E and sends the central station
(a) the name to be enrolled,
(b) the corresponding public parameter I obtained by exponentiating i using modular arithmetic.
(c) the reference value R.

E4: The CS stores the record with 3 fields: (Name, I , R)

Verification

V1: On request from a user at the FS, the CS sends the reference value R and a challenge *J* obtained as the result of exponentiating a random j using modular arithmetic

V2: The FS requests and gets a fresh reading V from the person to verified. The reading is denoted by V. Using V and R it then computes an approximate codeword C′

V3: The FS recovers the codeword C from C ′using the chosen error correction code. This in turn is used to regenerate i. Then *J* is exponentiated by i and result hashed to a value z. z is sent to the CS.

V5: CS computes z' the corresponding hashed value obtained from the modular exponentiation of I by j and verifies z' = z. If not, the verification fails.

Conclusions and Future Work

We have successfully demonstrated that, exactly reproducible, and therefore cryptographically useable bit patterns can be extracted from a person's biometric data .

As an application, we have used the bit-pattern derived from an individual's biometric data to provide secure remote proof of identity for him. In this application, no private biometric data ever leave the field station nor are any such data stored in a central station. The device used is the hand geometry reader 3DID built by Recognition Systems, Inc. of Campbell, CA. [4]

Although the data representation scheme and the ECC to be used are typically device-specific, the identification algorithm used is perfectly general . It is important that a good estimate of the number of bits to be corrected be obtained for each device. Future work should be focused in this direction for a comprehensive study of the reproducibility of the bit patterns resulting from any biometric device. At the same time we should study the discriminating power of the bit strings over a large population. To our knowledge, only two works in this direction have been reported in the literature.([5], [6])

Among other advantages, this system

(a) does not require an individual to carry a card,

(b) does not require one to remember, or even note down a password anywhere, and

(c) allows a person to carry his secret individualized password within his own body in the form of statistically unique biometric measurements, yet convertible to a unique bit string on location, never to be revealed to anybody anywhere in an insecure environment.

Among the weaknesses of this system based on the hand geometry are:

1. We have only 16 reproducible bits (for the HG data) allowing for a universe of 65,536 elements. An exhaustive search is clearly possible that reveals A's private key. For a strong system 60-80 reproducible bits are needed. Although precise information is not available, the only devices that come close from EyeDentify Inc. and IrisScan Inc.
2. Surreptitious photographing of the agent's hand clearly poses a risk. Biometrics based on fingerprints arouse similar concerns. Scanning of the retina or iris texture may be the desirable approach here.
3. The random numbers used in the system must be as close to true random numbers as possible. Otherwise, there is a danger of reproducibility of the 'random' number by an adversary rendering the 'secret' a non-secret.

Acknowledgments

We wish to thank Dr. David Sidlowski of Recognition Systems , Inc. Campbell, CA for providing us with the biometrics data based on hand geometry collected at his company.

References

1. Adamek Jiri, Foundations of Coding: Theory and Applications of Error-Correcting Codes with an Introduction to Cryptography and Information Theory, John Wiley, Inc. 336 p., 1991.
2. Schneier, B., Applied Cryptography: Protocols, Algorithms, and Source Code in C, John Wiley and Sons, Inc., New York, 618 p., 1994.
3. Rivest, R., The MD5 Message Digest Algorithm, RFC 1321, Apr. 1992.
4. Recognition Systems, Inc., Campbell, CA 95008.
5. Holmes, James P., Larry J. Wright, Russell L. Maxell, A Performance Evaluation of Biometric Identification Devices, Sandia Report SAND91-0276 . UC-906, June 1991
6. Daugman, J. G., High Confidence Visual Recognition of Persons by a Test of Statistical Independence, IEEE Trans. PAMI, vol. 15, Nov. 11, 1993.
7. EyeDentify Inc., Baton Rouge, LA 70816
8. IriScan, Inc. Mt. Laurel, NJ 08054

Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7405-Eng-48

Technical Information Department • **Lawrence Livermore National Laboratory**
University of California • Livermore, California 94551