Visual Authentication and Identification*

Moni Naor** and Benny Pinkas***

Dept. of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel.

Abstract. The problems of authentication and identification have received wide interest in cryptographic research. However, there has been no satisfactory solution for the problem of authentication by a *human* recipient who does not use any trusted computational device, which arises for example in the context of smartcard–human interaction, in particular in the context of electronic wallets. The problem of identification is ubiquitous in communication over insecure networks.

This paper introduces *visual authentication* and *visual identification* methods, which are authentication and identification methods for human users based on visual cryptography. These methods are very natural and easy to use, and can be implemented using very common "low tech" technology. The methods we suggest are efficient in the sense that a single transparency can be used for several authentications or for several identifications. The security of these methods is rigorously analyzed.

Keywords: authentication, identification, visual cryptography.

1 Introduction

Authentication and identification are among the main issues addressed in Cryptography. In an authentication protocol an *informant* tries to transmit some message to a *recipient*, while an *adversary* controls the communication channel by which the informant and the recipient communicate and might change the messages transmitted through that channel. At the end of the protocol the recipient outputs what he considers to be the message sent to him by the informant. If the adversary does not alter the communication, then this output should be equal to the original message. If however the adversary does change the communication, the recipient should detect this with high probability and report that the communication has been tampered. In an identification protocol, a *user* has to prove his identity to a *verifier*. Any adversary trying to pose as the user should not be able (except with small probability) to convince the verifier that he is communicating with the user.

Authentication and identification protocols have been studied extensively in various setups and under different assumptions on the power of the different parties. This paper concentrates on a scenario in which the recipient in the authentication protocol or the user in the identification protocol is human and as such cannot perform complicated computations or store large amounts of data. We do

* A full version of this paper is available in [9].

** Incumbent of the Morris and Rose Goldman Career Development Chair. Research supported by BSF Grant 32-00032. E-mail: naor@wisdom.weizmann.ac.il.

*** E-mail: bennyp@wisdom.weizmann.ac.il.

not require this human to use any secure computational device except his or her natural capabilities. This case is interesting since a system is as secure as its weakest component, and yet we do not know of any rigorous treatment of the human factor in cryptographic protocols. Here we analyze cryptographic systems in which the human part can be isolated and examined: Authentication by a human recipient is a cryptographic system in which a human has to solve a decision problem – whether to accept or reject the received message. Identification of a human user is a protocol in which an adversary should not be able to replicate the role of the human user, even if this user does not use any computational device. Another motivation to investigate these problems is to construct functional cryptographic protocols in which the human party does not need to use any device except natural human capabilities. The implementation of such protocols may be cheaper since there is need for less hardware.

Although humans cannot perform computations which are easily carried out by computers, the human visual perception can easily perform tasks which may be considered as "complicated computations". The systems we present utilize the visual capabilities of the human user. In our systems the human party and the other party share some secret information, and the human receives, stores and uses this information as an image on a transparency. The systems we suggest are based on the idea of *visual cryptography*, which was introduced in [10]. We describe the basic concepts of visual cryptography in subsection 1.3.

All the systems we suggest are rigorously analyzed. The security of the systems does not depend on any computational assumptions. Instead it is reduced to assumptions regarding human visual capabilities, which can be verified by empirical tests. We therefore present a new framework for proving the security of systems which involve human participants.

1.1 Motivation

The motivation for human identification is clear to anyone who has used a password. Such a system should enable the user to prove his identity to a remote computer, and yet should not enable an adversary who controls the communication to identify himself as the original user. There are systems which perform secure human identification using hand held computing devices or through biometric approaches. Compared to such systems our visual identification system is very "low tech". It does not require special hardware and can actually be independently implemented by anyone who wishes to use it, thus freeing security from being dependent on external hardware suppliers.

Authentication by a human recipient is intended to aid users who receive messages from a remote party through an insecure channel[1]. We will refer to the different parties as follows: the human recipient is Harry (Human), the informant is Sally (since in some applications the informant is a Smartcard), and the adversary is Peggy (in some applications the adversary is the Point of sale). One

[1] It can also be used to authenticate messages that human users send to remote parties, if a second round of communication is used. In this round the remote party answers with an authenticated message which contains the message it received, and the human should acknowledge the correctness of this message using a password.

application can be a user using an a terminal and a network which are insecure to connect to his remote computer. Another application might be the authentication of messages received by facsimile. A major application answers a well known threat to electronic payments: to authenticate the messages sent from an electronic wallet (most commonly a smartcard) to its owner.

It should be stressed that a straightforward application of visual cryptography to perform authentication is insecure, as is any straightforward application of a one-time-pad for authentication. In the scheme we suggest Harry is equipped with a (small) transparency. When Harry places the transparency over an image sent to him by Sally, the combination of both images will be the message that is sent to Harry.

The idea of supplying Harry with a transparency to help him in the authentication or to allow him to identify himself might seem strange. However, this procedure has some clear advantages: A transparency is much cheaper than a computing device and the systems we propose use transparencies which can be small enough to be carried in a wallet. Moreover, the production of the transparencies is very simple and so users can build their own authentication or identification systems without having to base their security on external hardware manufactures. The authentication and identification processes are very simple, the user just has to place the transparency on a screen or a printed message and view the result[2], he does not have to key numbers into a computer or consult a codebook. The visual authentication methods we suggest have the additional advantage of being applicable to any kind of visual image, not just for textual messages. The security of the authentication and identification methods does not depend on any computational assumptions and an upper bound on the (small) probability of failure can be computed.

1.2 Previous Work

Human–computer cryptographic interaction has been previously studied in both contexts we examine, authentication and identification. The problem of authentication was previously investigated mostly in the context of electronic payment systems [1, 2, 4] but no satisfactory solution was given for *standard* smartcards. All the suggested solutions require a secure channel between the user (who is the recipient) and his secure hand held computer (the informant). These methods are also only applicable for textual (or even just numerical) messages.

The second problem, human identification which does not require external devices, is very important in the context of access control since it frees the human user from carrying auxiliary computing devices for the identification process. This problem was addressed in [8, 7] but the methods suggested there are not proved to be secure for performing several identifications. Another solution is for the user to carry a list of one-time passwords, such as in [5, 11], but our system offers a much larger "density" for the information that the user carries. That is, it enables a much larger number of identifications for a certain amount of "storage" required from the user. This property enables the user to perform secure identifications with several verifiers, as we describe in subsection 5.2.

[2] The problem of correct alignment between the two images can be solved by providing a solid frame into which the transparency is entered, which fixes it in the right place.

1.3 Visual Cryptography

Visual cryptography was introduced by Naor and Shamir in [10]. It is a perfectly secure encryption mechanism, and the decryption process is performed by the human visual system. The ciphertext is a printed page and the key is a printed transparency of the same size. When the two are stacked together and carefully aligned the plaintext is revealed. Knowing just one of these two shares does not reveal any new information about the plaintext. This encryption scheme can be also considered as a 2-out-of-2 secret sharing scheme (the two shares being the ciphertext and the key), and it can be generalized to a k out of n secret sharing scheme. More information on visual cryptography can be found in the full version of this paper [9] or in [12].

In this paper we will only use the basic 2-out-of-2 visual secret sharing of [10]. In this scheme the plaintext is treated as an image, a collection of pixels. Each pixel in the plaintext is represented by a square of 2×2 real pixels (that is, real dots that are printed on a sheet of paper or on a transparency), these are called subpixles. Each plaintext pixel is divided into two shares such that in each share exactly two of the subpixels are black and the other two are transparent. Suppose that in the first share the two upper subpixels are black. If in the other share the two *lower* subpixels are black, then stacking the two shares together composes an image in which all four subpixles are black. If on the other hand the two *upper* subpixels in the second share are black (as in the first share) then stacking the two shares together yields an image in which only two subpixels are black. The former possibility is used to encrypt a black pixel, whereas the latter one is used to encrypt a white pixel[3]. There are six ways to place two black subpixels in the 2×2 square. For each pixel, one of these options will be chosen randomly for the first share. The second share will be the same as the first one if the pixel is white, or it will contain the complementary subpixels if the pixel is black. Note that since each single share is random, a single share does not add any information to the a-priori information that is known about the shared secret.

A straightforward implementation of visual cryptography for authentication is insecure. For a secure authentication Peggy must have some ambiguity regarding the contents of the share that Harry holds even after knowing the message sent by Sally, as in the case of standard authentication [3].

1.4 Organization of the Paper

In the next section we define the model of the authentication process we investigate, and the exact power of the different parties. Section 3 describes general methods for visual authentication, including efficient methods for performing several authentications using a single transparency. Section 4 defines and section 5 describes methods for secure visual identification of a human user. Section 6 concludes and suggests some open problems.

[3] Note that a white pixel is represented by a square which is not completely white but rather half white. This causes a reduction in the contrast of the image but the image is still easily readable by the human eye.

2 Model and Definitions for Visual Authentication

First we define the *visual authentication scenario*, and based on it we define what is a *visual authentication protocol* which is performed in this scenario. Together they constitute a *visual authentication system*. We then define the security requirements that a visual authentication system should have.

Definition 1 (visual authentication scenario). There are three entities in the visual authentication scenario: H (Harry), P (Peggy) and S (Sally). H is human and has human visual capabilities. For each protocol the capabilities that are required from H must be stated. These capabilities must include the ability to identify an image resulting from the composition of two shares of a 2-out-of-2 visual secret sharing. Other capabilities might be the ability to verify that a certain area is black, the ability to check whether two images are similar, etc. There is a security parameter n, such that the storage capacities and computing power of S and P are polynomial in n.

In the initialization phase S produces a random string r and creates a transparency T_r and some auxiliary information A_r as a function of r. Their size is polynomial in the security parameter n. S sends T_r and A_r to H through an off-line private initialization channel to which P has no access (this is the only time this private channel is used). S also sends to H a set of instructions that H should perform in the protocol (e.g. checking at a certain point whether a certain area in the image is black, comparing two areas, etc.). These instructions are public and might get known to P, but she is unable to change them.

Following the initialization phase all the communication is done through a channel controlled by P, who might change the communicated messages.

It is hard to rigorously analyze processes which involve humans since there is no easy mathematical model of human behavior. In order to prove the security of such protocols the human part in the protocol should be explicitly defined, thus isolating the capabilities required from the human participant. The security of the protocol must be reduced to the assumption that a "normal" person has these capabilities. This assumption can then be verified through empirical tests. Although we restrict P's power to be polynomial in the security parameter we do not make use of this limitation, the schemes we suggest are secure against an adversary with unbounded computing and memory capabilities.

Definition 2 (visual authentication protocol). S wishes to communicate to H an information piece m, the content of which is known to P.

- S sends a message c to H, which is a function of m and r.
- P might change c before H receives it[4].

[4] In our applications a message c is an image. Therefore it might be possible for P to change it so that it will not be in the form of a black and white image. For instance, m' might contain blinking pixels or, if the resolution is good enough, grey pixels. However, we assume that H either detects such messages as illegal, or assigns each pixel a value of either black or white.

– Upon receiving a message c' H outputs either FAIL or $\langle \text{ACCEPT}, m' \rangle$ as a function of c' and of T_r and A_r. When he outputs ACCEPT he also outputs m', what he considers to be the information sent to him by S.

Next we define the security requirements from visual authentication systems. The first definition ensures that the adversary cannot convince the human recipient to receive *any* message different from the original message. The second definition only ensures that for any a-priori determined message m' the adversary cannot convince the recipient that the received message was m'.

Definition 3 (security). Assume that H has the capabilities required from him for the protocol, that he acts according to the instructions given in the protocol, and that the visual authentication system has the property that when P is faithful then H always outputs $\langle \text{ACCEPT}, m \rangle$. We call the system

– $(1-p)$-authentic if for any message m communicated from S to H, the probability that H outputs $\langle \text{ACCEPT}, m' \rangle$ is at most p (m' should of course be different from m).
– $(1-p)$-single-transformation-secure ($(1-p)$-sts) if for any message m communicated from S to H and any $m' \neq m$ (which was determined a-priori) the probability that H outputs $\langle \text{ACCEPT}, m' \rangle$ is at most p.

A $(1-p)$-sts visual authentication system is obviously less secure than a $(1-p)$-authentic system, but it suffices for many applications and in particular for smartcard payment systems: we can demand that the customer receives the amount of money that his smartcard has to pay (m') directly from the point of sale, and if it does not equal the communicated message then the customer rejects.

In our model the adversary P can change the message sent from S to H at its will. However a legal share of a visual secret sharing scheme should contain exactly two black subpixels in every 2×2 square representing a pixel. There are two types of changes which can be made by P:

1. She can change the position of the two black subpixels in the squares in the image. This change cannot be noticed by the recipient H.
2. She can put more than or less than two black subpixels in a square. This produces an illegal share. However, this deviation will probably go unnoticed by H unless it is done in too many pixels[5]. We will further discuss and quantify this issue in the following section.

We do assume that the image that the human user views does not change after he has placed his transparency. This can be easily achieved if the image is first printed and then used by H (however, this requires the use of a printer which might be

[5] It is not easy to detect such pixels since there is no clear separation between different squares. H can detect these pixels more easily if he is supplied with two "chess board" transparencies: one with the pixels (i, j) with odd $i + j$ blackened, and the other with the even pixels blackened. He will be instructed to put each of these transparencies on the displayed image before putting his "secret" transparency. The first transparency isolates the pixels in the "even" locations and makes it easier to detect illegal pixels in these locations. The second transparency has the same effect for the "odd" pixels.

too expensive for some applications, e.g. for vending machines). We also assume that the contents of H's transparency remain secret. For example, this requires that there is no hidden camera behind H's back that reads the contents of the transparency (a solution against peeping eyes is suggested in [6]).

The definitions we gave define one-time systems. That is, they do not guarantee the security of the system if it is used to authenticate more than a single message. When we will suggest systems for several authentications we will explicitly define them as n-times secure, i.e. good for securely authenticating n messages.

There are two types of measures for complexity. Physical measures include the size of the information that the user has to carry, the storage and computation requirements from S, and the length of the communication. The second type includes the complexity of the operations that the human user has to perform in the authentication process.

In all the systems we propose the physical requirements are linear in the size of the message and logarithmic in the fault probability p (note also that the communication channel between current smartcards and a host computer runs at 9600 bps, and this throughput is enough for the methods we suggest). The complexity of the operations that the human user has to perform cannot be measured in "number of basic operations" as is done with machine computations. For each scheme we explicitly define what capabilities the human participant should have in order for the scheme to be secure. In some cases these capabilities are quantified (e.g. the human participant notices if the displayed image is different from a "legal" image in more than t pixels), and the other complexity measures are connected to the parameters of this quantification. The assumptions made about human capabilities can be verified through experiments. When these assumptions are verified the protocol is completely proved to be secure.

3 Authentication Schemes

This section describes visual authentication methods which are applicable for any kind of visual data: numerical, textual or graphical. The first three methods are one-time methods that can be used for only a single authentication. We then describe an efficient many-times method which can be used for several authentications. It is also possible to define visual methods which are good only for authenticating textual or numerical messages. Such methods use the fact that such messages are composed of characters which are elements from a small alphabet (i.e. digits or letters). We do not describe these methods since they are of much less interest than methods for general visual messages.

3.1 Method 1 — Content Areas and Black Areas

Initialization: The user H receives a transparency which is a share of a 2-out-of-2 visual secret sharing scheme. It is divided into two areas, one of them (which was chosen at random) is denoted as the *content* area, and the other is denoted as the *black* area.

Authenticated communication: S sends to H a message which is a share of a 2-out-of-2 visual secret sharing scheme. The image which is the combination of the transparency and this share has the message m in the content area and a black area which is completely black (see fig. 1). If the black area is not totally black then H should regard this message as a fraud attempt.

It is easy to prove that the adversary P has success probability at most $1/2$ if the two following assumptions on H's capabilities holds: (a) For any two semantically different messages m and m', H can notice if the share he receives from S has $|m \triangle m'|$ or more pixels in which the number of black subpixels is not two (this assumption seems reasonable since if $|m \triangle m'|$ is too small then the two messages are not semantically different). (b) H is capable of noticing any white subpixel in the black area (since this areas is completely black).

The first assumption prevents P from changing the message using only changes of type 2. The second assumption prevents it from doing any changes of type 1 to the black area. Therefore she must decide which is the content area, and her probability of success is at most $1/2$.

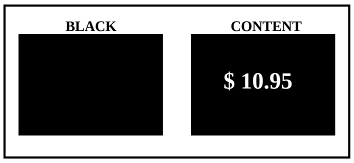

Fig. 1. The result of the composition of the user's transparency and the communicated image, for the "content areas black areas" method.

To reduce P's probability of success we can use k areas: There are $2^k - 1$ possibilities to partition k areas into black areas and content areas such that there is at least one content area. One of these partitions is selected at random and H is told in advance which areas are content areas. The image he observes should have the same message in all the content areas and all the other areas should be black. If P wishes to change the displayed message she must decide exactly which are the content areas, and her probability of success is at most $\frac{1}{2^k-1}$. This is more efficient than repeating the basic scheme to achieve this probability, which would have required k (possibly concurrent) repetitions, using $2k$ areas.

Theorem 4. *There is a $(1 - \frac{1}{2^k-1})$-authentic visual authentication scheme which uses a transparency with k areas such that each is large enough to accommodate the transmitted message. The method assumes H has the capability to detect a white pixel in a black region, to distinguish for every two semantically different messages m and m' between the case that there are more than $|m \triangle m'|$ pixels with more than or less than two black subpixels in the message he receives and the case that there are none, and to compare up to k areas in order to check whether they all contain the same message.*

There is a variation of this method which is slightly less efficient but does not require the user to check the image he receives for illegal pixels before placing his transparency on it. We describe it in the full version of the paper.

3.2 Method 2 — Position on the Screen

Initialization: Assume the image is composed of $r \times c$ pixels. A "bounding box" of size $r' \times c'$ pixels is drawn with a thin line at a random location on the transparency that is given to H.

Authenticated communication: The combination of the transparency and the communicated share should have the message displayed inside the bounding box, in white on a black background which covers all pixels inside and outside the bounding box. Figure 2 illustrates a transparency with a marked bounding box and a composed image with the message in the bounding box.

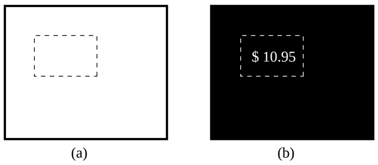

Fig. 2. (a) The user's transparency with the bounding box. (b) The composed image.

It should be shown that for any message $m' \neq m$ the adversary P has small success probability in changing m to m'. The task of P is to reverse the pixels of $m_d = m \bigtriangleup m' = (m \cap \overline{m'}) \cup (\overline{m} \cap m')$ for the image located inside the bounding box. We should prove that her chances in achieving this are small.

It is easy to prove security if we assume H to be very sharp-eyed and to notice if the displayed image is different from m' by even a single pixel: Let $m_d^{i,j}$ be the set of pixels which correspond to the set m_d in the bounding box located at coordinates (i, j). If P does not flip exactly the pixels in $m_d^{i,j}$, she fails. For any two different locations (i, j) and (i', j') it holds that $m_d^{i,j} \bigtriangleup m_d^{i',j'} \neq \emptyset$. There are $(r - r')(c - c')$ equally likely different locations and therefore P's probability of success is at most $\frac{1}{(r-r')(c-c')}$.

A more relaxed assumption on the capabilities of the user is that he can detect differences of t pixels or more between the displayed message and the image with m' in the correct bounding box. If the difference is at least this big then P fails. The following theorem is proved in the full version of the paper (the proof can be applied to other metrics, as is described in the full paper).

Theorem 5. *Let r and c be the number of rows and columns of the image. Let r' and c' be these values regarding the bounding box. Let m be the message communicated by S and let m' be a semantically different message. Assume that the human recipient H has the following capabilities: any image with hamming distance greater than t from m' is not captured by H as being m', and H notices if more than t' pixels in the image displayed to him have more than or less than two black subpixels. Then the authentication system we described is a $(1 - \frac{4(t+t')}{(r-r')(c-c')})$-single-transformation-secure visual authentication system.*

3.3 Method 3 — Black and Grey

The security of the following method is exponential in the hamming distance between the original message and the message that P wishes to display to H. The drawback of this method is that it reduces the contrast of the displayed image.

We previously used the 2-out-of-2 visual secret sharing method in which all four subpixels of a black pixel are black, whereas a white pixel has two black subpixels. We can also define a *grey* pixel as a pixel with three black subpixles. Let the two shares of a pixel be denoted as s_1 and s_2. Given a share s_1 of a black pixel it is easy to construct another share s_1' such that together with s_2 it composes a grey pixel. However, given a share s_1 of a grey pixel the probability of constructing a share s_1' that together with s_2 composes a black pixel is at most $1/4$. When the message m is written in black on a grey background it is therefore hard for the adversary to change a background pixel into a message pixel. Similarly, when the message is written in grey on a black background it is hard for the adversary to "erase" a pixel of the message and change it to a background pixel. The scheme we suggest displays the message in two areas. In one area it is displayed in black on grey and in the other area in grey on black. The user is instructed to verify that the messages on both areas are equal. The following theorem is easily proved using the Chernoff bound.

Theorem 6. *Let t' be an upper bound on the number of pixels of the share sent by S, in which the number of black subpixles is different from two, that still goes unnoticed by the user. For any message m', define $t_{m'}$ as the maximum hamming distance of a displayed message from m' such that a user may accept the displayed message as m'. Let t be an upper bound on $t_{m'}$ over all messages m'. If the message is displayed in the scheme suggested here and the hamming distance between any two semantically different messages m and m' is at least $2 \cdot (t' + \frac{4}{3}(1+\varepsilon)t)$, then this is a $(1-p)$-authentic visual authentication system, where $p = 2e^{-2\frac{\varepsilon^2}{1+\varepsilon}t}$.*

3.4 Many-Times Methods

The three authentication methods we suggested in the previous subsections were all secure for only a single authentication. It is obviously preferable to have methods which are secure for several authentications. A straightforward construction of a many-times scheme is to take any of the previous one-time schemes and store several independent copies of it in different areas of a single transparency. The number of copies in a single transparency depends on the security parameters which define the size of the area that is used by each copy, and on the size of the transparency. This construction is not too bad since the methods we suggested are relatively efficient in the transparency space they use, especially the "black on grey" method of subsection 3.3 which has exponential security. However, we would like to do better than this, since in practice there is great importance for the size of the transparency (which should be minimized) and for the number of possible secure authentications (which should be maximized). Next we define many-times security and demonstrate how to construct an efficient many-times authentication scheme from the "position on the screen" scheme.

Definition 7 (n-times security). A visual authentication system is n-times $(1-p)$-single-transformation-secure (n-times $(1-p)$-sts) if the following is true for any n messages $\langle m_1, \ldots, m_n \rangle$. For any message m_i $(1 \leq i \leq n)$ communicated from S to H, and any message m' different from m_i, the probability that H outputs $\langle$ACCEPT,$m'\rangle$ is at most p. If P is faithful then H should always output $\langle$ACCEPT,$m\rangle$.

The many-times authentication scheme we suggest uses the following parameters. The messages to be authenticated are of size $r' \times c'$ pixels, and r_0 and c_0 are the security parameters. The size of the transparency is $r \times c$, where $r = r_0 + n_r r'$ and $c = c_0 + n_c c'$. The transparency is used for $n = n_r n_c$ authentications.

Initialization: A random starting point (i_0, j_0) is chosen s.t. $1 \leq i_0 \leq r_0$ $1 \leq j_0 \leq c_0$. A grid of n areas, each composed of $r'c'$ pixels, is drawn with a thin line on the transparency starting from location (i_0, j_0). The ith area is defined as the area in the intersection of row $\lceil i/n_c \rceil$ and column $(i \bmod n_c) + 1$. Figure 3 illustrates the configuration of the transparency in this scheme.

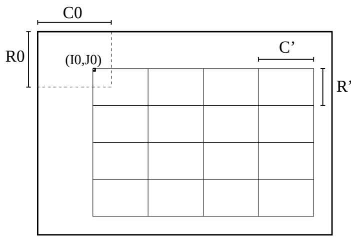

Fig. 3. The user's transparency in the many-times visual authentication scheme.

i-th authentication: S sends her share of the message m_i (written in white over a black background) in the ith area of the grid, and in all the other pixels of the share that she sends there are exactly two black subpixels in two random locations (in the 2×2 square). The human recipient H verifies that the message he sees when he puts his transparency is in the ith area.

Theorem 8. *Assume that if the hamming distance between the displayed image and an image m' is greater than t then the human recipient H does not perceive the displayed image as m'. Also assume that the user notices if in more than t' pixels of the communicated image the number of black subpixels is not two. Then a transparency of size $(r_0 + n_r r') \times (c_0 + n_c c')$ pixels can be used to get an $n_r n_c$-times $(1-p)$-single-transformation-secure visual authentication system, where each message is of size $r' \times c'$ pixels, and where $p = \frac{4(t+t')}{r_0 c_0}$.*

4 Model and Definitions for Visual Identification

The scenario of visual identification is identical to the visual authentication scenario of definition 1. However the goal of the identification protocol is different, to allow the *human user* H to prove his identity to the *verifier* S without consulting any computational device. The objective of the adversary P is to convince the verifier that she (P) is actually the human user. There is no point in constructing visual identification protocols which enable only a single secure identification since this can be achieved by supplying the user with a simple password. We will therefore consider only many-times identification protocols, i.e. protocols in which a single transparency can be used for many identifications. The protocol is a *challenge-response* type protocol in which the verifier sends a challenge to the user, who should answer it based on the secret information he holds.

Definition 9 (visual identification protocol). We define the protocol for the i-th identification of H to S:

- S sends a challenge c_i to H, which is a function of the secret data r.
- Upon receiving c_i the human user H computes a response a_i as a function of c_i and his secret information T_r and A_r, and sends it back to S.
- S decides whether the other party is H based on the messages c_i and a_i, and the secret data r. She then answers either ACCEPT or REJECT.

The adversary P might try to pretend to be H. In this case she might even try to question H by claiming to be S and requiring H to prove his identity. Then she initiates the identification protocol with the verifier S and sends a response which she hopes would convince S that the other party is H.

Definition 10 (ℓ-times $(1-p)$-secure visual identification protocol). A visual identification protocol is ℓ-times $(1-p)$-secure if the following two conditions hold after the adversary P has listened to at most ℓ_1 identifications that were answered by H and has pretended to be the verifier in at most ℓ_2 identifications of H, subject to the constraint $\ell_1 + \ell_2 \leq \ell$.

- S always accepts when H answers according to the protocol.
- If an adversary P receives the message c_i sent from S and answers it with a message b_i which is a function of c_i and any previous ℓ communications (where ℓ_1 of them were initiated by S and ℓ_2 by P, and all were answered by H), then S accepts with probability at most p.

A stronger definition is security against coalitions of k corrupt verifiers. That is, there are many verifiers and the user might need to prove his identity to any one of them. No coalition of at most k verifiers should be able to pretend to be the user in a conversation with a verifier which is not a member of the coalition. The visual identification scenario against coalitions of size k is identical to the single verifier visual identification scenario, except for the creation and distribution of the random data r and its derivatives: a central trusted authority generates r, sends each verifier S_i some secret data r_i which is a function of r and of i, and as before sends H the transparency T_r and the auxiliary information A_r. The visual identification protocol against coalitions of size k is as in the single verifier case except for S_i basing her operation on the data r_i and not on r. The definition of security is identical for the former security definition, but security is required even when the coalition members use all the secret information r_i they have and the information they gathered while tapping to or initiating at most ℓ identifications of the user.

5 Visual Identification Methods

The methods we suggest for visual identification do not use any visual secret sharing scheme since there is no need to construct an image to be viewed by H. Instead H has to prove to the verifier S that he knows some property of the transparency. We use colored transparencies, or more concretely ten different colors which we assume to be easily discernible from each other: black, white, green, blue, red, yellow, purple, brown, pink and orange. A different set of colors can be used and the security depends on the number of colors in the set.

A very attractive property of our methods is that they are very "low tech" in comparison to current secure identification methods that require the user to consult a hand held computing device, to connect a smartcard into a special port in the remote computer, or even to use biometric identification devices. Visual identification methods enable everyone with access to a color printer (or even to a black and white printer) to build a secure identification scheme which can be used for example to permit access to certain areas or to identify parties for communication. Furthermore, since the world-wide-web introduces a universal graphic interface a visual identification can be performed when a user connects from a remote host, and use a web browser to display the image that is sent from the verifier to the user. In this case no special software should be installed on the remote computer for the purpose of identification.

The visual authentication methods we suggest demand very little of the verifier. Therefore the roles of the verifier and prover can be reversed, i.e. the verifier is human and he verifies the identity of a computer with which he communicates. The human can then demand a remote computer to prove its identity to him before he sends it some confidential information (e.g. his credit card number).

5.1 A Secure Visual Identification Scheme for a Single Verifier

Here the basic unit we consider in the transparency is not a pixel but rather a *square*, which is a collection of a few pixels (for example, a square of 4×4 pixels). At the initialization phase the user H receives a transparency which is divided into many squares, and each square is randomly colored with one of the ten possible colors. The order of the colors is kept secret and is known only to H and to the verifier S (S either knows the order explicitly, or alternatively the order can be determined by the output of a pseudorandom number generator and S should only store its seed).

Let N be the number of squares in the transparency, and let d be the number of squares which are queried about in a single identification. The identification protocol goes as follows: S chooses d random squares. She sends H an image which is completely black, except for the locations of the d squares which are white. The user H puts his transparency over this image and sends back to S the colors in the locations of the white squares, by some predefined order (to make the system easier to use H can send his response using a point-and-click interface). The verifier S accepts only if H's answer is correct for all the d squares.

It is clear that H can always identify himself successfully. The best strategy for P is to query the user ℓ times and learn the color of $d\ell$ squares. P does not have any information about the colors of the other squares. Her probability of success is expected[6] to be $(\frac{1}{10}+\frac{9d\ell}{10N})^d$. A transparency with N squares can therefore be definitely used for $\ell=\frac{N}{9d}$ identifications and the security is still greater than $1-5^{-d}$. This result is summed up in the following theorem:

Theorem 11. *A transparency with N squares colored with 10 colors can be used for an ℓ-times $(1-(\frac{1}{10}+\frac{9d\ell}{10N})^d)$-secure visual identification scheme, such that in each identification the user should send to the verifier the colors of d squares.*

5.2 A Visual Identification Scheme Secure Against Coalitions of Verifiers

In this scheme the secret information r_i that each verifier S_i receives contains the colors of a random subset of $(1-q)N$ squares in the transparency that the user holds (where $0<q<1$). The identification protocol is identical to the previous identification protocol except for the verifier questioning the user about the colors of random squares from the set of squares whose colors the verifier knows. The "density" of the visual identification scheme, i.e. the large number of squares which can be stored in a single transparency, enables this scheme to be secure against relatively large coalitions.

Theorem 12. *When $\ell \le \frac{Nq^k}{2d}$ a transparency with N squares colored with 10 colors can be used for an ℓ-times $1-(1-\frac{9}{20}(1-\frac{d}{(1-q)N})^\ell)^d$-secure against k-verifiers, visual identification scheme, in which the user has to send the values of d colors in each identification.*

6 Conclusions and Open Questions

We have suggested methods for visual authentication and identification, and have given rigorous analysis of their security. All methods are secure regardless of the computational capabilities of the adversary. We also demonstrated a secure many-times visual identification method which is very "low tech" and can be implemented with almost no investment.

Comparing the one-time visual authentication methods, the advantage of the first method ("black area content area") is that its security depends on relatively easy requirements from the human user. Its disadvantage is the loss in area which implies that the security may not be as small as we would like. The advantage of the "position on the screen" method is that the error probability is proportional to the number of pixels and not to the redundancy in area. Its disadvantages are that the probability might not be small enough, and more capabilities are required of the human user. The advantage of the "black and grey" method is that the probability of non-detection is exponentially small in the distance between semantically different messages. Its disadvantages are the loss in contrast, and the additional capabilities required of the user. In comparison to the one-time methods

[6] This follows since P knows the colors of at most $d\ell$ squares. S chooses squares randomly and the expected success probability of P is $\sum_{i=0}^{d}\binom{d}{i}(d\ell/N)^i(1-d\ell/N)^{(d-i)}10^{-(d-i)} = (\frac{1}{10}+\frac{9d\ell}{10N})^d$.

the many-times authentication method has the advantage of substantially reducing the amount of transparency area that is needed per authentication in order to achieve a certain security level.

There are many open questions left. It should be interesting to find an authentication method whose security is exponential in the *size* of the message, or a method which does not reduce the contrast and whose security is exponential in the hamming difference between the messages. Another open problem is to devise more efficient methods which are secure only against polynomial adversaries. An important issue is to check which human capabilities can be easily verified and to base the security of the visual methods on these capabilities (in particular a better measure than hamming distance can be used to define similarity between images). It should also be interesting to design a method that enables a human informant to authenticate a message it sends, *without* requiring two-way interaction. A related problem is to devise a one-way function which is easily computable by humans.

7 Acknowledgments

We thank Omer Reingold for his careful reading and valuable suggestions, and the anonymous referees for their helpful comments.

References

1. Abadi M., Burrows M., Kaufman C. and Lampson B., Authentication and delegation with smart-cards, *Sci. of Comp. Prog.*, 21 (2), Oct. 1993, 93-113.
2. Boly J., Bosselaers A., Cramer R., Michelsen R., Mjolsnes S., Muller F., Pedersen T., Pfitzmann B., de Rooij P., Schoenmakers B., Schunter M., Vallee L. and Waidner M., The esprit project cafe – high security digital payment system, in *Computer Security – ESORICS 94*, Springer-Verlag LNCS Vol. 875, 1994.
3. Gilbert E., MacWilliams F. and Sloane N., Codes which detect deception, *Bell Sys. Tech. J.*, Vol. 53, No. 3, 1974, 405–424.
4. Gobioff H., Smith S., Tygar J. D. and Yee B., Smartcards in hostile environments, in *Proc. of The 2nd USENIX Workshop on Elec. Commerce*, Nov. 1996, 23-28.
5. Haller N. M., The S/KEY one-time password system, in *Internet Soc. Symp. on Network and Dist. Sys. Sec.*, 1994.
6. Kobara K. and Imai H., Limiting the visible space visual secret sharing schemes and their application to human identification, in *Asiacrypt '96*, Springer-Verlag LNCS Vol. 1163, 1996, 185–195.
7. Matsumoto T., Human-computer cryptography: an attempt, in *ACM Conf. on Comp. and Comm. Sec.*, ACM Press, March 1996, 68-75.
8. Matsumoto T. and Imai H., Human identification through insecure channel, in *Eurocrypt '91*, Springer-Verlag LNCS Vol. 547, 1991, 409–421.
9. Naor M. and Pinkas B., Visual authentication and identification, in *Theory of Cryptography Library*, `http://theory.lcs.mit.edu/~tcryptol`.
10. Naor M. and Shamir A., Visual cryptography, in *Eurocrypt '94*, Springer-Verlag LNCS Vol. 950, 1995, 1–12.
11. Rubin A. D., Independent one-time passwords, *Computing Systems*, The USENIX Association, Vol. 9, No. 1996, 15–27.
12. Stinson D. R., An introduction to visual cryptography, presented at *Public Key Solutions '97*. Available at `http://bibd.unl.edu/~stinson/VCS-PKS.ps`.