# Visual Authentication and Identification

Moni Naor [*] Benny Pinkas [†]

**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. The problem of authentication arises for example in the context of smart-card–human interaction, in particular in the context of electronic wallets. The problem of identification is very common 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*. An *adversary* controls the communication channel by which the informant and the recipient communicate and might change the messages transmitted on 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 who tries 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, see [18] for a survey on authentication and [17] for a survey on human identification. This paper concentrates on a scenario in which the recipient in the authentication protocol or the user in the identification protocol is a human being and as such cannot perform complicated computations or store large amounts of data. We do not require this human to use any 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.

[1] Incumbent of the Morris and Rose Goldman Career Development Chair, Dept. of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. Research supported by a grant from the Israel Science Foundation administered by the Israeli Academy of Sciences. E-mail: naor@wisdom.weizmann.ac.il.

[2] Dept. of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: bennyp@wisdom.weizmann.ac.il.

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 have to use anything except natural human capabilities. The implementation of such protocols might 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 can 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 propose are based on the idea of *visual cryptography*, which was introduced in [14]. We will describe basic concepts of visual cryptography in section 1.3.

It should be noted that the authentication systems we propose are suitable for authenticating any type of visual message, be it textual or graphical. In this sense they are better than authentication schemes which require the recipient to consult a small hand-held computing device (such schemes may be used only for textual messages). The size of the transparencies that are used, both for authentication and for identification, is relatively small and it should be possible to carry them in the user's wallet. Some of the visual authentication and all the identification systems we suggest are many-times secure. That is, a single transparency can be used in these systems for several sessions (but not for an unlimited number of sessions).

All the systems we suggest are analyzed rigorously. The security of the systems does not depend on any computational assumptions. Instead, it depends on assumptions regarding human visual capabilities. These assumptions can be verified by empirical tests. We therefore present a new framework for proving the security of systems in which there are human participants. We believe that the security of such systems should be rigorously reduced to assumptions regarding human capabilities, and then empirical tests should verify that a normal human participant has these capabilities.

## 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 tapped the communication of past identifications to identify himself as the original user.

Human authentication is motivated by the needs of users who receive messages from a remote party through an insecure channel. 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 Smart card), and the adversary is Peggy (in some applications the adversary is the Point of sale). One application can be a user ($H$) using an a terminal and a network which are insecure ($P$) to connect to his remote computer ($S$). Another application might be the authentication of messages received by facsimile. A major application which answers a well known threat to electronic payments is discussed in appendix A. In this application a smart card is used as an electronic wallet and the communication from the card to its owner must be authenticated.

It should be stressed that a straightforward application of visual cryptography to perform authentication is insecure, as we shall demonstrate in appendix B. In the scheme we suggest Harry is equipped with a (small) transparency, in addition to his smart card. When Harry places the

transparency over a screen displaying the message sent to him by Sally, the combination of the images on the screen and in the transparency will be the amount that his smart card is requested to pay.

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 schemes without having to rely on or supply their secrets to hardware manufactures. The authentication and identification processes are very simple and effortless, The user just has to place the transparency on the screen and view the result[1]. The user does not have to key numbers into a computer or consult a codebook. The visual authentication methods we suggest have the additional advantage of applying 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 the exact (small) probability of failure can be computed.

We first suggest authentication schemes which are good only for a single transaction. However, even these schemes are efficient enough so that a very small transparency is sufficient to achieve good security. We also present many-times secure authentication and identification systems, which use a single transparency to perform several transactions.

### 1.2 Previous work

Human–computer cryptographic interaction has been previously studied in both contexts we examine: that of the human validating the authenticity of messages he receives from the computer, and that of the computer validating the identity of a human who approaches it.

The first problem, authentication, was previously considered in the literature [2, 4, 19, 8] mostly in the context of electronic payment systems, but no satisfactory solution was given for *standard* smart cards. In all the suggested solutions the recipient had to use a hand-held computer to assist him in the authentication process, and this situation is not optimal because of the costs of the special hand-held device and since the process of the user keying information to or from the device might induce errors. 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 [12, 11] but the methods suggested there are not proven to be secure for performing several identifications. Another solution is for the user to carry a list of one-time passwords, such as in [9], but our system offers a much larger "density" of the information that the user carries. That is, they allow a much larger number of identifications to be performed with the same amount of "area" requirements from the user.

### 1.3 Visual cryptography

Visual cryptography was introduced in [14]. It is a perfectly secure encryption mechanism, and the decryption process is done 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

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

carefully aligned, the plaintext is revealed. Knowing just one of these two shares does not reveal any 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. There has been considerable interest in visual cryptography, including suggestions which improve the contrast of the resulting image [15, 5, 1, 3], or add color to the image [13, 16].

In this paper we will only use the 2-out-of-2 visual secret sharing of [14]. The plaintext is treated as an image, a collection of pixels. Each pixel in the plaintext is represented by a square of $2 \times 2$ 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, one in the ciphertext and the other in the key. 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 by stacking the two shares together we get an image in which all four subpixles are black. If, on the other hand, in the second share the two upper subpixels are the black ones, as in the first share, then stacking the two shares together yields an image in which only two subpixels are black. The former possibility will be used to encrypt a black pixel, whereas the latter one will be used to encrypt a white pixel[2]. There are six ways to place two black subpixels in the $2 \times 2$ 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 opposite subpixels if the pixel is black. Note that having just one share does not leak any information about the other share and hence about the shared secret, each of the two options – black or white – is equally probable.

A straightforward implementation of visual cryptography for authentication is insecure, as we demonstrate in appendix B.

From this discussion it follows that for a secure authentication Peggy must have some ambiguity regarding the contents of the share that Harry holds, as in the case of standard authentication [7].

### 1.4 Organization of the paper

In the next section we define the model of the authentication process we investigate. We define there 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 7 compares the different authentication and identification methods and suggests some open problems.

## 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]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-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 initialization private channel to which $P$ has no access (this is the only time this private channel is used). $S$ also sends $H$ a set of instructions that $H$ should perform in the protocol. The instructions might include checking 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 not able to change them.*

*Following the initialization phase all the communication between $H$ and $S$ is done through a channel controlled by $P$, who might change the transferred 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. Then it is possible to isolate the capabilities required from the human participant (e.g. the ability to verify that a certain image is totally black). 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. In our schemes the computation and storage requirements from $S$ are linear in the size of the message, and are well within the power of current smart cards. In addition, the auxiliary information $A_r$ is relatively short and can be written on the edges of the transparency.

**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 the message $c$ to $c'$ before $H$ receives it*[3]*.*
- *Upon receiving the message $c'$ $H$ outputs either FAIL or $\langle ACCEPT, m' \rangle$ as a function of $c'$ and his secret information $T_r$ and $A_r$. When he outputs ACCEPT he also outputs $m'$, what he thinks to be the information sent to him from $S$*[4]*.*

In the context of smart card payments there is another step of communication from $H$ to $S$. In order to present a complete solution for this application we describe this final stage in appendix C .

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 ensures only that for any particular message $m'$, the adversary cannot convince the recipient to believe that the received message was $m'$.

[3] In our applications the 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.

[4] The structure of the protocol is not the most general possible. In particular, it is possible to define multi-round authentication protocols in which $H$ and $S$ perform several rounds of communication as is done in non-visual authentication [6]. It is also possible to use definitions which allow the recipient a negligible probability of error even when he receives a message which was not altered by the adversary. We do not describe these generalizations in order to simplify the exposition.

**Definition 3 (visual authentication system)** *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 $\langle ACCEPT,m\rangle$. We call the system*

- $(1-p)$-authentic *if for any message $m$ communicated from $S$ to $H$ the probability that $H$ outputs $\langle ACCEPT,m'\rangle$ (and $m' \neq m$) is at most $p$.*
- $(1-p)$-single-transformation-secure *($(1-p)$-sts) if for any message $m$ communicated from $S$ to $H$ the probability that $H$ outputs $\langle ACCEPT,m'\rangle$ ($m' \neq m$) is at most $p$.*

A $(1-p)$-sts visual authentication system is obviously less secure than a $(1-p)$-authentic system, but such a system suffices for many applications and in particular for smartcard payment systems: We can demand that the customer should also receive directly from the point of sale the amount of money ($m'$) that his smart card has to pay. The point of sale then has to change the message $m$ to be exactly equal to $m'$. In this case a system which is single-transformation-secure is secure enough.

It should be stressed again that although we emphasis the application of visual authentication for smart card payments, all the methods we suggest are not limited to authenticating numbers, but can rather be used to authenticate any kind of visual message. However, we only deal with authenticating black and white images.

In our model the adversary, $P$, can change the message sent from $S$ to $H$ at its will. However, if the message is a share of a visual secret sharing scheme it should contain exactly two black subpixels in every $2 \times 2$ 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 inside 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 subpixles inside 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 over time, and in particular it does not change after the user has placed his transparency on the image. 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 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.

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

Several measures of complexity can be examined regarding visual authentication systems:

- The size of the transparency and of the auxiliary information, which is the size of the information that the user has to carry. This can be measured in pixels and bytes. The space and computation requirements from $S$, and the amount of information that $S$ has to communicate to $H$.

[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.

- The complexity of the operations that the human user $H$ has to perform in the authentication process.

In all the systems we propose the requirements from $S$ and from the transparency are linear in the size of the message and logarithmic in the fault probability $p$. Note also that the communication channel between current smart cards 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 General Methods for Authentication

This section describes visual authentication methods which are applicable for any kind of visual data: numerical, textual or graphical. The first method, "content areas and black areas" in subsection 3.1, is very simple and yet secure enough to be practical. The second method, "position on the screen" in subsection 3.2, has greater security. The third method, "black and grey" in subsection 3.3, has security which is exponential in the hamming distance between the message that $S$ sends and the message that $P$ wishes to display. However, the price for the exponential security of this method is a reduction in the contrast of the viewed image. The first three methods are one-time and can be used only for a single secure authentication. We then present an efficient many-times secure 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. These 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

Divide the image into two areas. Define at random one area as the "black" area and the other area as the "content" area. $H$ receives a transparency which is a share of a 2-out-of-2 visual secret sharing scheme. The transparency is divided into two areas and it is written on $H$'s transparency which area is the black area and which is the content area. $S$ 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 figure 1). $H$ is instructed that if the black area is not totally black, then he should regard the message he received as a fraud attempt.

If the adversary $P$ wants to change the message displayed to $H$ she should make some changes to the content area of the image. She can either make changes of type 1 (changing the position of the two black subpixels in the $2 \times 2$ square of a pixel), or changes of type 2 (changing the number of black subpixels in a square). In order to transform a message $m$ into a message $m'$ by changes of type 2, there should be at least $|m \triangle m'| = |(m \cap \overline{m'}) \cup (\overline{m} \cap m')|$ such changes. The security of the system depends on the following two assumptions regarding $H$'s capabilities:

1. 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 different from

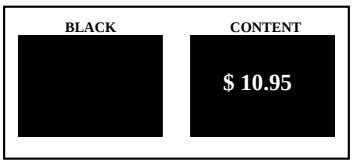


Figure 1: The result of the composition of the user's transparency and the communicated image, for the "content areas black areas" method.

2. This assumption seems reasonable since if $|m \triangle m'|$ was too small then the two messages were not semantically different.

2. $H$ is capable of noticing any white subpixel in the black area. Since all four subpixels of a black pixel are black, the black area is completely black and therefore it seems reasonable to assume that $H$ is capable of detecting any white areas there.

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 make changes of type 1 to this area only. Her probability of success is at most $\frac{1}{2}$.

It is important to note that the method we describe is a one-time method. Each transparency can be used only once. When a transparency is used for the second time $P$ can compare the image that $S$ is sending to the previous image she sent and deduce which area is the black area. $P$ can then change the displayed message to a message of her choice.

To reduce $P$'s probability of success we can use $k$ areas: There are $2^k - 1$ possibilities to partition the areas into black areas and content areas such that there is at least one content area. One of these partitions will be selected at random and will be used. The user 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 for each area whether it is a black area or a content area. 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: this would have required $k$] (possibly concurrent) repetitions, requiring $2k$ 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 catch 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 does not require the user to check the image he receives before he places his transparency. The transparency contains three areas which are ordered at random to be the *black*, the *white*, and the *content* areas. The user knows what is the title of each area, and he expects to see the message in the content area, a black area which is completely black and a white area which is completely white. $P$ does not know the order of the areas. If she displays a pixel with less than two black subpixels, then with probability 1/3 this pixel is in the black area and if we assume that the user can detect a white dot in the black area he can learn that $P$ tries to cheat. The same happens if $P$ displays a pixel with more than two black subpixels. Therefore in order not to get caught with good probability, $P$ is forced to display only pixels with two black subpixels.

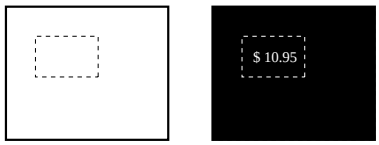


Figure 2: (a) The bounding box depicted on the user's transparency. (b) The composed image.

### 3.2 Method 2 — Position on the Screen

Assume the image is composed of $r \times c$ pixels. In the initialization phase $S$ tells $H$ the location of a bounding box of size $r' \times c'$ pixels in which the message should be displayed. Let $(i, j)$ be the location of the upper left corner of the box. The location of the bounding box is drawn with a thin line on the transparency, so $H$ can easily check whether the message is displayed there. The message is displayed 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 the composed image with the message in the bounding box.

The following analysis is for the case of preventing $P$ (the adversary) who controls the computer screen from displaying a specific message $m'$, and does not consider other messages except $m$ and $m'$. Let $m_c = (m \cap m') \cup (\overline{m} \cap \overline{m}')$. The pixels in this area have the same color in both messages and should not be changed by $P$. Let $m_d = m \triangle m' = (m \cap \overline{m}') \cup (\overline{m} \cap m')$. The pixels in this area must be flipped by the adversary. i.e. transformed from black to white and vice versa. Note that flipping from black to white is performed in the same way as flipping from white to black.

We prove in the next subsection that the solution is secure for a "sharp eyed" user ($H$). This observation is useful mainly to demonstrate the ideas that are used in proving the security theorem for ordinary users, which is given in the following subsection.

#### 3.2.1 Sharp Eyed $H$

Assume here that if the displayed image differs from $m'$ by even a single pixel, the adversary fails (either $H$ sees an incomprehensible message, or he sees a message $m''$ which is different from $m'$). $H$ also has the capability to notice if the image sent from $S$ has a pixel in which the number of black subpixles is not exactly two.

Let $m_d^{i,j}$ be the set of image pixels which correspond to the set $m_d$ in the bounding box located at coordinates $(i, j)$ (that is, whose upper left corner is at $(i, j)$). If $P$ does not flip exactly all the pixels in $m_d^{i,j}$, and only them, she fails. For any two different locations $(i, j)$ and $(i', j')$ it holds that $m_d^{i,j} \triangle m_d^{i',j'} \neq \emptyset$. In other words, there are $(r - r')(c - c')$ equally likely different sets $m_d^{i,j}$, with every two of them having a non empty difference. Hence there are this many equally likely different possibilities of sets of pixels to be flipped. If $P$ chooses the wrong set she fails, her probability of success is therefore at most $\frac{1}{(r-r')(c-c')}$.

#### 3.2.2 A Not So Sharp Eyed $H$

Here we only assume that $H$ has the capability to detect differences of $t$ pixels or more between the displayed message and the image with $m'$ in the correct bounding box (actually $t$ might depend on $m'$ and should therefore be denoted as $t_{m'}$, but we omit the subscript to simplify the notation). If the difference is at least this large then $P$, the adversary who changed the communicated image, fails.

As before let $m_d^{i,j} = m \triangle m'$ located at the bounding box in location $(i, j)$. The image containing

$m_d^{i,j}$ can be viewed as an $rc$-bit long vector. Let $V^t(m_d^{i,j})$ be the set of all $rc$-bit vectors which have a hamming distance of less than $t$ from this image.

**Claim 5** *If $P$ uses (i.e. flips the bits of) a vector $v \notin V^t(m_d^{i,j})$ to transform a message displayed in location $(i,j)$ from $m$ to $m'$, she fails.*

**Proof:** Using such a vector $v$ results in the hamming distance between the displayed image and $m'$ being greater than $t$.

**Corollary:** $\Pr(P \text{ succeeds by flipping the bits of } v) \leq \frac{|\{V^t(m_d^{i,j}) | v \in V^t(m_d^{i,j})\}|}{(r-r')(c-c')}$, since all the possible locations of the bounding box have equal probability.

**Claim 6** $\forall v, m_d, t \ \left|\{V^t(m_d^{i,j}) \mid v \in V^t(m_d^{i,j})\}\right| \leq \left|\{m_d^{i',j'} \mid d(m_d^{i,j}, m_d^{i',j'}) < 2t\}\right|$.

**Proof:** Fix one $m_d^{i,j}$ such that $v \in V^t(m_d^{i,j})$. Let $m_d^{i',j'}$ be any other vector for which $v \in V^t(m_d^{i',j'})$. Then from the triangle inequality it follows that the hamming distance between $m_d^{i,j}$ and $m_d^{i',j'}$ is at most $2t$. □

Therefore, for every $v$ we can bound the probability of the success in using $v$ by an upper bound for the number of locations $(i',j')$ that result in a difference of less than $2t$ between $m_d^{i,j}$ and $m_d^{i',j'}$. In order to eliminate boundary problems we allow $(i,j)$ to be in the range $-r' \leq i \leq r + r'$, $-c' \leq j \leq c + c'$. This only increases the upper bound.

Given an $rc$ bits long vector $m_d$ (representing an image of size $r \times c$) and a value $u$, define $S(m_d, u)$ as the number of locations $(i',j')$ such that the hamming distance between $m_d^{i,j}$ and $m_d^{i',j'}$ is less than $u$.

**Claim 7** *For every vector $m_d$ there exists a vector $m'_d$ satisfying $m'_d \subset m_d$ ($m'_d \cap \overline{m_d} = \emptyset$ and $m_d \neq m'_d$), such that $S(m_d, u) \leq S(m'_d, u)$.*

**Proof:** We will prove the claim for the case $|m'_d| = |m_d| - 1$. Consider the black pixel which is the upper most black pixel in the left most column which contains pixels from $m_d$, and call it the "upperleft" pixel. Let $m'_d$ be the image resulting from the deletion of the upperleft pixel of $m_d$. Every shift of the image $m_d$ results in two copies of the original image and of the upperleft pixel. Consider the transposition of these two images together. At least one of the two copies of the upperleft pixel is not covered by a black pixel from the other image. Therefore the difference of the hamming distance between the two copies of $m_d$, from the hamming distance between the two copies of $m'_d$ resulting from the same shift, is either 0, 1, or 2. Therefore every shift of $m_d$ which results in a hamming distance of less than $u$, also results is a hamming distance of less than $u$ between two similarly shifted copies of $m'_d$. □

**Claim 8** *For every image $m_d$, $|m_d| > u/2$, it holds that $S(m_d, u) < (|m_d|^2/(|m_d| - u/2)$.*

**Proof:** The image $m_d$ contains $|m_d|$ pixels, which can be matched to $\binom{|m_d|}{2} < \frac{|m_d|^2}{2}$ couples. Each couple can match in exactly two shifts of the location of the image. If two locations of the image, $m_d^{i,j}$ and $m_d^{i',j'}$, have a hamming distance less than $u$ then they match in at least $(2|m_d| - u)/2 = |m_d| - u/2$ pixels. Therefore $S(m_d, u) < |m_d|^2/(|m_d| - u/2)$. □

Now we can give an upper bound for the adversary's probability of success. First, we limit ourselves for vectors $m_d$ with at least $2t$ black pixels. Otherwise, we claim that the difference between the images $m$ and $m'$ is too small for the user to discern between them. For every image $m_d$ there is an image $m'_d \subseteq m_d$, with $|m'_d| = 2t$. From claims 7 and 8 it follows that $S(m_d, 2t) \leq S(m'_d, 2t) < \frac{4t^2}{2t-t} = 4t$. The adversary's probability of success is at most $\frac{S(m_d,2t)}{(r-r')(c-c')} \leq \frac{4t}{(r-r')(c-c')}$.

The adversary $P$ who controls the communication channel can make changes of type 2 to the image that $S$ sends to $H$. That is, she can transmit an image in which there are pixels with more than or less than two black subpixels. Assume that $H$ can detect if more than $t'$ pixels are changed in this manner, where $t'$ is a parameter. Then to calculate the adversary's probability of success we can certainly do the previous analysis using $t + t'$ instead of $t$.

**Theorem 9** *Let $r$ be the number of rows of the image, and let $c$ be the number of columns. Let $r'$ and $c'$ be these values regarding the bounding box. Let $m$ be the message communicated by the smart card 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 $t_{m'}$ 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,*

*In the authentication system we described the adversary can make the user identify the message as $m'$ with probability at most $\frac{4(t_{m'}+t')}{(r-r')(c-c')}$.*

*Let $t$ be the maximum value of $t_{m'}$ over all messages $m'$. The 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 message that $S$ sends to $H$, and the message that $P$ wishes to display to him. The drawback of this method is that it reduces the contrast of the displayed image.

In the 2-out-of-2 visual secret sharing method that we use 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 hard for the adversary to change a pixel in the background 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 into 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.

To prove the security of the scheme we first analyze the probability of the adversary transforming background pixels to message pixels in the "black on grey" area, and then analyze the whole scheme. Let $t$ be as before the maximum hamming distance from a message that a user accepts, and let $t'$ be the maximum number of pixels in which the adversary can display more than two black subpixels without being detected by $H$.

**Lemma 10** *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 the 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 black on a grey background, and it holds that for any two semantically different messages $m$ and $m'$ we have $|m' \backslash m| > t' + \frac{4}{3}(1 + \varepsilon)t$, then this is a $(1 - p)$-authentic visual authentication system, where $p = 2e^{-2\frac{\varepsilon^2}{1+\varepsilon}t}$.*

**Proof:** By an application the Chernoff bound. Details are omitted from this version.

The analysis for the "grey on black" case is identical, except for considering $|m \backslash m'|$ instead of $|m' \backslash m|$. If two messages have hamming distance $d$ then either $|m \backslash m'|$ or $|m' \backslash m|$ is at least $d/2$. The following theorem is then a straightforward corollary of the previous lemma:

**Theorem 11** *Let the capabilities of $H$ be defined by $t$ and $t'$ as they were defined above. If the hamming distance between any two "semantically" different messages $m$ and $m'$ is greater than $2 \cdot (t' + \frac{4}{3}(1+\varepsilon)t)$, then there exists 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 that we suggested in the previous subsections were all one-time, in the sense that they were secure only for a single authentication. It is obviously much more 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 copies of it in different areas of a single transparency, where each copy depends on different random data. 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 as small as possible) and for the number of possible secure authentications (which should be as large as possible). Next we define many-times security and demonstrate how to construct an efficient many-times authentication scheme from the "position on the screen" method of subsection 3.2.

**Definition 12 ($n$-times $(1-p)$-single-transformation-secure visual authentication system)** *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 $m_1, \ldots, m_n$. For any message $m_i$ $(1 \leq i \leq n)$ communicated from $S$ to $H$, and any message $m'$ different from $m$, the probability that $H$ outputs $\langle ACCEPT, m' \rangle$ is at most $p$, even if the adversary $P$ knows the communication that was passed in the authentication of messages $m_1, \ldots, m_{i-1}$. If $P$ is faithful then $H$ should always output $\langle ACCEPT, m \rangle$.*

The many-times authentication scheme we suggest is as follows. Let the message that should be authenticated be of size $r' \times c'$ pixels. The parameters $r_0, c_0$ will be used as security parameters. Let the size of the transparency be $r \times c$, where $r = r_0 + n_r r'$ and $c = c_0 + n_c c'$. The transparency will be used for $n_r \cdot n_c$ authentications in the following way:

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

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

*Security (Proof sketch):* Each of the possible $r_0 c_0$ starting points has equal probability, and so the $r_0 c_0$ possibilities for the location of the bounding box in which the message $m_i$ is displayed are

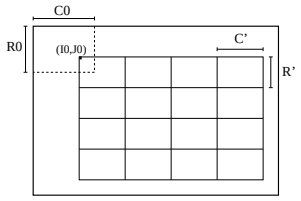


Figure 3: The user's transparency in the many-times visual authentication scheme.

equiprobable. For every starting point $((i_0, j_0)$ the values $m_1, \ldots, m_i$ sent in the first $i$ messages correspond to the same number of possible values for the contents of the user's transparency, since there is no pixel in which a non-random data is displayed more than once. The situation is therefore as in the "position on the screen" method of subsection 3.2, for a screen size of $(r_0 + r') \times (c_0 + c')$, except that previously it was known that all pixels were either black or white, and here some of the pixels might contain random data. The probability for the success of the adversary is at most as in the previous scheme. We get the following theorem:

**Theorem 13** *Assume that if the hamming distance between the displayed image and an image $m'$ is greater than $t$ then the human recipient $H$ does perceive the displayed image as $m'$. Also assume that the user notices if in more than $t'$ pixels the number of black sub-pixels 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 purpose of the identification protocol is different, to allow the *human user $H$* to prove his identity to the *verifier $S$* without needing to consult any computational device. The purpose 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, and he should answer it based on some secret information he has in order to prove his identity.

**Definition 14 (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 $r$ which is common to $S$ and $H$, and of additional random bits $r_S$ which are known only to $S$.*
- *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 first 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 15 ($\ell$-times $(1-p)$-secure visual identification protocol)** *A visual identification is $\ell$-times $(1-p)$-secure if the following two conditions hold after the adversary $P$ has listened to at most $\ell$ identifications that were answered by $H$ or even after $P$ has pretended to be the verifier in at most $\ell$ authentications with $H$.*

- *Whenever $H$ answers $S$ according to the protocol, $S$ accepts with probability 1.*
- *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 $\ell$ communications $c_{i_1}, b_{i_1}, \ldots, c_{i_\ell}, b_{i_\ell}$, 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. However, no coalition of at most $k$ verifiers can pretend to be the user in a conversation with a verifier which is not in the coalition. The visual identification scenario against coalitions of $k$ 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, except for requiring security even when the coalition members use the secret information $r_i$ each of them has and also the information they gathered while tapping to at most $\ell$ identifications of the user.

# 5 Visual Identification Methods

The methods we suggest for visual identification do not use the 2-out-of-2 visual secret sharing schemes of [14], or any other 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 one another: black, white, green, blue, red, yellow, purple, brown, pink and orange. A different set of colors can be used, and the security parameters then change as a function of 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 or even to connect a smart card into a special port in the remote computer. 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 the visual identification process that is performed when a user connects from a remote host can use a web browser to display the image that is sent from the verifier to the user, and in this case no special software should be installed on the remote computer for the purpose of identification.

**Visual identification with a human verifier:** 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 require a remote computer to prove its identity to him before he sends it some confidential information (e.g. his credit card number).

In the next subsection we describe an identification system for a single verifier and in subsection 6 we describe a system secure against coalitions of corrupt verifiers.

### 5.1 A secure visual identification scheme for a single verifier

Here the basic unit we use in the transparency is not a pixel but rather a *square*, which is a collection of a few pixels (for example, a square of $4 \times 4$ 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 with which are used 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 pseudo random number generator and $S$ should know its seed. The latter option requires considerably less memory from $S$).

Let $N$ be the number of squares in the transparency, and let $d$ be the number of squares which are queried about in the protocol. 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 the image received from $S$ 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, it can allow $S$ to 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. On the other hand, an adversary $P$ might query $H$ at most $\ell$ times and learn the color of only $d\ell$ of the squares. $P$ does not have any information about the colors of the other squares. Therefore her probability of success is $(\frac{1}{10} + \frac{9d\ell}{10N})^d$ at the most[6]. 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 can be summed up in the following theorem:

**Theorem 16** *A transparency with $N$ squares colored with 10 colors can be used for an $\ell$-times $(1-(\frac{1}{10} + \frac{9dn}{10N})^d)$-secure visual identification scheme, such that in each identification the use should send to the verifier the colors of $d$ squares.*

## 6 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 $(1-q)N$ of the $N$ squares in the transparency that the user $H$ holds. The identification protocol which is secure against coalitions of $k$ verifiers 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.

Imagine that a coalition of $k$ verifiers has performed $\ell$ runs of the identification protocol with the user. Therefore it might have learned the values of at most $d\ell$ squares of the user's transparency which were not previously known to the coalition members. Now the coalition attempts to identify itself as $H$ to another verifier $S_j$. The verifier $S_j$ queries about $d$ squares which are chosen at random, independently from the information that the coalition has. For each such square, the probability that its color is known to the coalition members is at most $1-q^k+d\ell/N$. Therefore, as in the previous scheme, the expected success probability of the coalition is at most $(\frac{1}{10} + \frac{9}{10}(1 - q^k + d\ell/N))^d = (1 - \frac{9}{10}(q^k - d\ell/N))^d$. On the other hand, the members of the coalition could have listened to $\ell$ identifications of the user to the verifier $S_j$, and they might have learned the colors of $d\ell$ of the squares whose colors $S_j$ knows. We require that $d\ell < (1-q)N/4$, to keep the error probability smaller than $2^{-2d}$.

[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 $\sum_{i=0}^{d} \binom{d}{i} (d\ell/N)^i (1 - d\ell/N)^{(d-i)} 10^{-(d-i)} = (\frac{1}{10} + \frac{9d\ell}{10N})^d$.

**Theorem 17** *When $d\ell < (1-q)N/4$, a transparency with $N$ squares colored with* 10 *colors can be used for an $\ell$-times $\left(\min(1-2^{-2d}, 1-(1-\frac{9}{10}(q^k - d\ell/N))^d)\right)$-secure against $k$-verifiers visual identification scheme.*

Note that the result for a single verifier is a private case of this theorem, for $k = 0$ and $q = 1$.

# 7 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.

Comparing the one-time visual authentication methods, the advantage of the first method ("black area content area") is that its security depends of 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 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.

The visual identification methods we suggested demonstrate a secure many-time identification method, which is very "low tech" and can be implemented with almost no investment.

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. Our methods are secure against infinitely powerful adversaries but this security is not needed in practice. 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. Also, if we allow two-way interaction then it is possible to use the visual authentication methods we suggested in order to allow a human informant to authenticate a message he sends to a recipient which is a computer. It should be interesting to devise such a method which does not require two-way interaction. A related problem is to devise a one-way function which is easily computable by humans.

# References


[1] Ateniese C., Blundo C., De Santis A. and D. R. Stinson, *Visual cryptography for general access structures*, accepted for publication in *Inf. and Comp.*. Also available at `http://www.eccc.uni-trier.de/eccc` as TR096-012.

[2] Abadi M., Burrows M., Kaufman C. and B. Lampson *Authentication and delegation with smart-cards*, Technical report 67, DEC Systems Research Center, October 1990. Revised July 1992.

[3] Blundo C., De Santis A. and D. R. Stinson. *On the contrast in visual cryptography schemes.* Manuscript. 1996. Available at `ftp://theory.lcs.mit.edu/pub/tcryptol.96-13.ps`.

[4] 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 M. Waidner, *The esprit projecSpringer-Verlag, 1991,*

*pp. 409–421. t cafe – high security digital payment system*, in *Computer Security – ESORICS 94*, LNCS Vol. 875, Springer-Verlag, 1994.

[5] S. Droste, *New results on visual cryptography*, Crypto '96, LNCS Vol. 1109, Springer-Verlag, 1996, pp. 401-415.

[6] Gemmell P. and M. Naor, *Codes for interactive authentication*, Crypto '93, LNCS Vol. 773, pp. 355-367, 1993.

[7] Gilbert E., MacWilliams F. and N. Sloane, *Codes which detect deception*, *Bell Sys. Tech. J.*, Vol. 53, No. 3, pp. 405–424, 1974.

[8] Gobioff H., Smith S., Tygar J. D. and B. Yee, *Smart cards in hostile environments*, in Proceedings of The Second USENIX Workshop on Electronic Commerce, November 1996. Also available in `http://www-cse.ucsd.edu/users/bsy/pub/hostile.ps`.

[9] N. M. Haller, *The S/KEY one-time password system*, in Internet Soc. Symp. on Network and Dist. Sys. Sec., 1994.

[10] ISO/IEC 7816-3:1989 *Identification cards – Integrated circuit(s) cards with contacts – Part 3: Electronic signals and transmission protocols.*

[11] T. Matsumoto, *Human-computer cryptography: an attempt*, in ACM Conf. on Comp. and Comm. Sec., 1996.

[12] Matsumoto T. and H. Imai, *Human identification through insecure channel*, Eurocrypt '91, LNCS Vol. 547, 1991.

[13] D. Naccache, *Colorful Cryptography – a purely physical secret-sharing scheme based on chromatic filters*, in *Coding and Information Integrity*, French-Israeli workshop, December 1994.

[14] Naor M. and A. Shamir, *Visual Cryptography*, Eurocrypt '94, LNCS Vol. 950, Springer-Verlag, 1995, pp. 1–12.

[15] Naor M. and A. Shamir, *Visual Cryptography II: improving the contrast via the cover base*, Cambridge Workshop on Cryptograhic Protocols, 1996. A full version is available at `ftp://theory.lcs.mit.edu/pub/tcrypto/96-07.ps`

[16] Rijmen V. and B. Preneel, *Efficient colour visual encryption or 'shared colors of Benetton'*. Presented at the rump session of Eurocrypt '96. Also available at `http://www.esat.kuleuven.ac.be/~rijmen/vc/`.

[17] R. Rivest, Class notes of lecture 9 in *6.915 Computer and Network Security*, available at `http://www.theory.lcs.mit.edu/~rosario/6.915/lecture9.ps`.

[18] G. Simmons, *A survey of information authentication*, in *Contemporary Cryptography – The Science of Information Integrity*, IEEE Press, 1991, pp. 379–419.

[19] Yee B. and J. D. Tygar, *Secure coprocessors in electronic commerce applications*, in Proceedings of the First USENIX Workshop on Electronic Commerce, July 1995.

# Appendices

# A Visual Authentication of Electronic Payments — Motivation

In this appendix we examine in detail an electronic payments application and describe how visual authentication can protect it from a possible fraud attempt. Consider an electronic cash system based on smart cards. Harry, a customer, has a private smart card which serves as his electronic wallet and contains a certain amount of money. When Harry wishes to buy something from Peggy, a merchant, he pays her using his electronic wallet. A transaction should take place between Peggy's POS (Point of sale) and Harry's wallet, after which a certain amount of money is transferred from the wallet to Peggy. The POS might also be a vending machine, and should handle transactions of small sums of money.

Common smart cards do not have any device for directly displaying or receiving information, and all their input and output is transferred through a host computer[7]. An analog to this situation in terms of conventional commerce might be that of customers being required to give their wallets to the merchant, who would take by herself the proper amount of money for the purchase. Most customers would obviously not agree for such a process.

Consider the following possible fraud attempt by Peggy, which was observed in [2, 4, 19, 8], and has not received yet a solution applicable for standard smart cards: Suppose that Harry buys something which costs \$1. The smart card (Sally) is asked by Peggy to pay her some money, but she might try to ask for any some of money (e.g. \$10). Sally might try to ask Harry for an acknowledgment for this payment, but all the communication between them goes through Peggy who might change its contents. It was most preferable if Harry could send Sally an acknowledgment which is a function of the sum he is willing to pay and of some secret information which only he and Sally know. However, this is difficult since Harry is human and cannot compute complicated functions. We will solve this problem by enabling trusted communication from Sally to Harry with *visual authentication*. Sally can then send Harry the sum that she is required to pay, and Harry can answer her (if he is approves) with a predetermined secret password that only the two of them know.

# B A Direct Application of V.C. Cannot be Used for Authentication

It is not secure to use a straightforward implementation of visual cryptography for authentication. We exemplify this for the case of electronic payments: Suppose that Harry has a key, i.e. a transparency, and his smart card knows the contents of this key. Before making a payment the smart card asks the POS to display an image which is a share of a 2-out-of-2 visual secret sharing. The result of the composition of this image with Harry's transparency is the sum that the smart card is required to pay. The card then waits for a predetermined password from Harry before it makes the payment. Suppose that Peggy tells Harry that he has to pay her \$1 but demands \$10 from the smart card. She knows that the smart card will send Harry a message containing \$10 and therefore when she gets the share that the smart card asks her to display she can deduce the contents of Harry's transparency. Then she can display an image which will result in the message \$1 appearing when Harry places his transparency on it.

# C Visual Authentication of Elec. Payments — Comm. from $H$ to $S$

If $H$ accepts, he sends back a message $d$ to $S$. This message is a function of $m'$,$c'$, $T_r$ and $A_r$, and should be easy to compute by $H$ (most conveniently, this message is fixed regardless of the value of $m'$). The message can be altered by $P$ to be $d'$. Upon receiving $d'$, $S$ carries a computation based on $d'$, $r$ and $m$. The output of this computation is either ACCEPT or FAIL, and is the final output of the authentication protocol. We will disregard this step for now, since we use a secret password or PIN, which $H$ sends to $S$ if $H$ accepts. This password is part of the secret information $A_r$ and does not depend on the contents of the message $m$. $S$ will accept if and only if she receives this password[8]. Therefore, the probability that $S$ accepts is equal to the probability that $H$ accepts plus the probability that $P$ can guess the password. Since the latter probability is very small, it is enough to investigate whether $H$ accepts or fails.

[7] The current standard for smart cards[10] does not include a specification for a direct input or output device connecting a smart card to its owner, so the situation of all communication between them going through the host computer will most probably continue.

[8] Note however that the secure visual identification methods which we suggest in the sequel can be used instead of the password, and $S$ would then require the recipient to prove his identity instead of sending a password.