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 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 visual authentication methods we suggest are not limited to authenticating textual messages, and can be used to authenticate any image.

An important contribution of this paper is the introduction of a framework for proving the security of protocols in which humans take an active part. We rigorously prove the security of our schemes using this framework.

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, see [20] for a survey on authentication and [18] 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 are

[1] A preliminary version of this paper appeared in Crypto '97.

[2] 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 BSF Grant 32-00032. E-mail: naor@wisdom.weizmann.ac.il.

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

human and as such cannot perform complicated computations or store large amounts of data. We do 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. Such protocols use "low tech" components, like transparencies. Their implementation might therefore be cheaper, and their security is also improved since it does not rely on the integrity of complex hardware components which are bought from external suppliers.

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 [15]. We describe the basic concepts of visual cryptography in subsection 1.3.

It should be noted that the authentication systems we suggest 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 can only authenticate textual messages). The size of the transparencies that are used, both for authentication and for identification, is relatively small and they can even be carried 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 for several sessions.

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 include 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 tapped the communication of past identifications 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 obtaining security which is not dependent on external hardware suppliers.

Authentication by a human recipient is motivated by the needs of users who receive messages from a remote party through an insecure channel[1]. These messages are not necessarily secret and in fact it should be assumed that they are known to the adversary (this knowledge might help the adversary). We will refer to the different parties as follows: the human recipient is Harry (Human),

[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.

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 application can be a user using an a terminal and a network which are insecure to connect to a 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. This application is discussed in detail in appendix A.

In our suggested application of the authentication scheme for electronic wallets Harry is equipped with a (small) transparency, in addition to his electronic wallet (Sally). When Harry places the transparency over the message sent to him from Sally, the combination of the images on the screen and in the transparency will be the amount that his wallet is requested to pay.

It should also 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 (we demonstrate this for the case of electronic payments in appendix B).

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 base their security on external manufactures of security hardware. 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[2]. 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 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 for the (small) probability of failure can be computed.

We first suggest authentication schemes which are good only for a single transaction. Even these schemes are efficient enough so that a 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 investigated [2, 4, 8] mostly in the context of electronic payment systems, but no satisfactory solution was given for *standard* smartcards. All the suggested solutions required 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 [13, 12] but the methods suggested there have not been 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, 19], but our system offers a much larger "density" of the information that the user carries. That is, they allow a

[2] 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.

much larger number of identifications to be performed with the same 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.

1.3 Visual Cryptography

Visual cryptography was introduced by Naor and Shamir in [15]. 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 and aligned together 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. There has been considerable interest in visual cryptography, including suggestions which improve the contrast of the resulting image [16, 5, 1, 3], or add color to the image [14, 17]. For a survey on this subject see [21].

In this paper we will only use the 2-out-of-2 visual secret sharing of [15]. 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 subpixels. 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 stacking the two shares together yields an image in which all four subpixels are black. If, on the other hand, in the second share the two upper subpixels 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 opposite subpixels if the pixel is black. Note that since each single share is random having just one 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, 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 6 compares the different authentication and identification methods 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 initialization private 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 in time 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 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 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 $\langle ACCEPT, m'\rangle$ 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*[5].

In the context of smartcard payments there is another step of communication from H to S. Although this stage does not involve authentication, we describe this final stage in appendix C in order to present a complete solution for this application.

It should be stressed again that although we emphasis the application of visual authentication for smartcard 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.

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 to believe that the received message was m'.

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 $\langle ACCEPT,m'\rangle$ is at most p (where m' is of course 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 $\langle ACCEPT,m'\rangle$ is at most p.*

A $(1-p)$-sts visual authentication system is obviously less secure than a $(1-p)$-authentic system, it only guarantees that it is hard to change the communicated message into a specific message which was determined before the communication started. However, this property 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. 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.

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[6]. We will further discuss and quantify this issue in the following section.

[5] 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.

[6] 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 $i + j$

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 (a solution against peeping eyes is suggested in [11]).

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, i.e. 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 (it 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.
- 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 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 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

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.

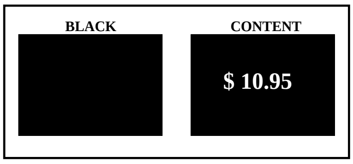

Figure 1: The result of the composition of the user's transparency and the communicated image, for the "content areas black areas" method.

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.

Security is proved under the following two assumptions regarding H's capabilities: (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 different from 2. This assumption seems reasonable since if $|m \triangle m'|$ was too small then the two messages were not semantically different. (b) 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 the adversary 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 and 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 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, this 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.*

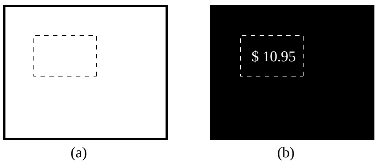

Figure 2: (a) The bounding box depicted on the user's transparency. (b) The composed image.

There is a variation of this method which does not require the user to check the image he receives for illegal pixels before placing his transparency on it, but it is slightly less efficient. 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 to avoid getting caught with good probability, P is forced to display only pixels with two black subpixels.

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.

The following analysis is for the case of preventing P (the adversary who controls the display) 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 $m_d = m \bigtriangleup m' = (m \cap \overline{m}') \cup (\overline{m} \cap m')$. The pixels in this area must be reversed by the adversary (i.e. transformed from black to white and vice versa). Note that changing a pixel from black to white is performed in the same way as changing it 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 subsection that follows.

3.2.1 Sharp Eyed H.

Assume here that if the displayed image differs from m' located in the bounding box by even a single pixel, the adversary fails (either H sees an incomprehensible message or he considers the message as being 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 subpixels is not exactly two.

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) (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 $m_d^{i,j} \bigtriangleup m_d^{i',j'} \neq \emptyset$. In other words, there are $(r - r')(c - c')$ equally likely different sets of pixels (to be reversed) $m_d^{i,j}$, with every two of them having a non empty difference. 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). Denote the hamming distance between vectors v_1 and v_2 as $d(v_1, v_2)$. The image $m_d^{i,j}$ can be viewed as an rc-bit long vector. Define $V^t(m_d^{i,j}) = \{v | d(v, m_d^{i,j}) \leq t\}$, the vectors with hamming distance at most t from $m_d^{i,j}$.

Claim 5 *If P uses (i.e. reverses 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: $\Pr(P$ succeeds by reversing the bits of $v) \leq \frac{|\{(i,j)|v \in V^t(m_d^{i,j})\}|}{(r-r')(c-c')}$ (where the probability is taken over the location (i, j)) since all the possible locations of the bounding box have equal probability.

Claim 6 *$\forall v, m_d, t$, let $S_1 = \{(i, j) \mid v \in V^t(m_d^{i,j})\}$, and let S_2 be a maximal set s.t. $S_1 \cap S_2 \neq \emptyset$ and it holds for every $(i, j), (i', j') \in S_2$ that $d(m_d^{i,j}, m_d^{i',j'}) < 2t$. Then $S_1 \subset S_2$.*

Proof: For every two elements $(i, j), (i', j') \in S_1$ it follows from the triangle inequality that $d(m_d^{i,j}, m_d^{i',j'}) < 2t$. Since S_1 intersects S_2 it is therefore also fully contained in it. □

Thus for every v we can bound the number of locations for which the adversary succeeds 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 using a vector m_d of size $|m'_d| = |m_d| - 1$. 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 deleting 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. That is, 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 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 $t_{m'}$ 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 convince the user to 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.

Note that although the theorem was proved for the hamming distance metric it can be proved to any metric (in which the triangle inequality holds). The only part of the proof which needs changing is counting the number of shifts of the location of the image which result in a distance of less than $2t$.

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.

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 subpixels. 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 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 security is proven by first analyzing the success probability of the adversary in transforming background pixels to message pixels in the "black on grey" area. The analysis uses the Chernoff inequality. The adversary has also the same success probability in transforming message pixels to background pixels in the background area. If m and m' have hamming distance d then either $|m \backslash m'|$ or $|m' \backslash m|$ is at least $d/2$. The following theorem then follows:

Theorem 10 *Let t' be an upper bound on the number of pixels of the share sent by S, in which the number of black subpixels 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 that we suggested in the previous subsections were all one-time in the sense that they were 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 copies of it in different areas of a single transparency, where each copy depends on independent 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 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" method of subsection 3.2.

Definition 11 (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 $m_1, \ldots, m_i$. 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 messages that should be authenticated be of size $r' \times c'$ pixels. The parameters r_0, c_0 are the 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 is used for $n = n_r n_c$ authentications in the following way:

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 $n_r n_c$ areas, each composed of $r'c'$ pixels, is drawn with a thin line on the transparency starting from location (i_0, j_0). The 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.

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 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.

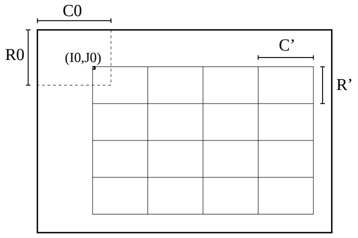

Figure 3: The user's transparency in the many-times visual authentication scheme.

Security: Each of the possible r_0c_0 starting points is chosen with equal probability. Given the messages $\langle m_1, \ldots, m_i \rangle$ and the shares that were sent by S for these messages the probability that the user's transparency was created using a certain starting point is proportional to the number of possible transparencies which correspond to these messages and shares and to this starting point. In these transparencies the values of the pixels which have a black or a white value in the communicated messages are fixed, and the other pixels can have any value. Since there is no pixel in which non-random data is displayed more than once, the same number of transparencies correspond to any starting point and all starting points are still equiprobable. 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 success probability of the adversary is at most as in the previous scheme. We get the following theorem:

Theorem 12 *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 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 needing to consult 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 some secret information he holds.

Definition 13 (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 14 (ℓ-times $(1-p)$-secure visual identification protocol) *A visual identification 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 with H, subject to the constraint $\ell_1 + \ell_2 \leq \ell$.*

- *S always accepts when H answers 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 ℓ communications $c_{i_1}, b_{i_1}, \ldots, c_{i_\ell}, b_{i_\ell}$ (where ℓ_1 of them were initiated by S and ℓ_2 by P) 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. It is required that no coalition of at most k verifiers can 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 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 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 the 2-out-of-2 visual secret sharing schemes of [15] 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 then 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 process 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).

In the next subsection we describe an identification system for a single verifier and in subsection 5.2 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 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 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 pseudorandom 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 S can send his response using a point-and-click interface. Also, in order to assure that different displays do not distort the size of the image, the user can be asked at the beginning of the identification to put his transparency on the screen and click on two fixed locations which are marked on the transparency. Then the displayed image can be resized to match the transparency). The verifier S accepts only if H's answer is correct for all the d squares.

Theorem 15 *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.*

Proof: It is clear that H can always identify himself successfully. Consider the situation after ℓ identifications. The best strategy for the adversary P is to use these identifications to query the user ℓ times and learn the color of $d\ell$ squares. When S queries the user she chooses squares randomly and then the expected success probability of P is $\sum_{i=0}^{d}\binom{d}{i}(d\ell/N)^i(1-d\ell/N)^{(d-i)}10^{-(d-i)} = (\frac{1}{10}+\frac{9d\ell}{10N})^d$. A transparency with N squares can therefore be definitely used for $\ell = \frac{N}{9d}$ identifications and the security is still greater than $1-5^{-d}$.

5.2 A Visual Identification Scheme Secure against Coalitions of Verifiers

In the multi-verifier scheme the secret information r_i that each verifier S_i receives contains the colors of a random subset T_i of squares in the transparency that the user holds. The squares are chosen to this subset independently at random with probability $1-q$, where q is a parameter between 0 and 1. The identification protocol is identical to the previous identification protocol, except for each verifier S_i only questioning the user about colors of random squares from the subset T_i 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.

Imagine that there is a coalition of k corrupt verifiers $S_1, \ldots, S_k$ who intend to impersonate the user H to another verifier S_0. The coalition members have listened to ℓ_1 identifications of H to

S_0, and there have also been ℓ_2 identifications of H to members of the coalition. Assume also that $\ell_1 + \ell_2 \leq \ell$.

The probability that the color of a square in the set T_0 is not disclosed in the secret information of any of the coalition members (i.e. the square is not in $\cup_{i=1}^{k} T_i$) is q^k. The coalition members can tap to identifications of H to S_0 and then they can learn colors of squares in T_0. The drawback in listening to such identifications is that since S_0 chooses the queries at random they contain many squares whose colors are already known to the coalition. The coalition members can also learn colors of squares from identifications of H in which they operate as the verifiers. Then they can query H about squares from $\overline{\cup_{i=1}^{k} T_i}$, whose colors are unknown to them. However, the coalition members do not know which squares are in T_0 and are useful in order to impersonate H to S_0, and therefore they cannot create queries which contain only squares from T_0.

A straightforward (but cumbersome) probability calculation can compute the best strategy for the coalition members (i.e. deciding on the values of ℓ_1 and ℓ_2 subject to the constraint $\ell_1 + \ell_2 \leq \ell$), and the success probability of the coalition. These are functions of N (the size of the transparency), d (number of squares which are queried about in a single identification), $(1 - q)$ (the probability that the color of a certain square is known to a certain verifier) and ℓ (the number of identifications that can be performed or listened to by the coalition).

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-time 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 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.

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 (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.

References

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

[2] Abadi M., Burrows M., Kaufman C. and Lampson B., *Authentication and delegation with smart-cards*, Science of Computer Programming, 21 (2), Oct. 1993, 93-113.

[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 Waidner M., *The esprit project cafe – high security digital payment system*, in Computer Security – ESORICS 94, Springer-Verlag LNCS Vol. 875, 1994.

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

[6] Gemmell P. and M. Naor, *Codes for interactive authentication*, Crypto '93, Springer-Verlag LNCS Vol. 773, 1003, 355-367.

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

[8] Gobioff H., Smith S., Tygar J. D. and B. Yee, *Smartcards 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 Society Symposium on Network and Distributed System Security, 1994.

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

[11] Kobara K. and H. Imai, *Limiting the visible space visual secret sharing schemes and their application to human identification*, Asiacrypt '96, Springer-Verlag LNCS Vol. 1163, 185–195.

[12] Matsumoto T., *Human-computer cryptography: an attempt*, in ACM Conf. on Computer and Communication Security, ACM Press, March 1996, 68-75.

[13] Matsumoto T. and Imai H., *Human identification through insecure channel*, in Eurocrypt '91, Springer-Verlag Springer-Verlag LNCS Vol. 547, 1991, 409–421.

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

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

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

[17] 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/`.

[18] 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`.

[19] Rubin A. D., Independent one-time passwords, *Computing Systems*, The USENIX Association, Vol. 9, No. 1996, 15–27.

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

[21] Stinson D. R., *An introduction to visual cryptography*, presented at Public Key Solutions '97. Available at `http://bibd.unl.edu/~stinson/VCS-PKS.ps`.

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 smartcards. Harry, a customer, has a private smartcard 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 smartcards 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, 8], and has not received yet a solution applicable for standard smartcards: Suppose that Harry buys something which costs \$1. The smartcard (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.

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

B A Direct Application of Visual Cryptography 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 smartcard knows the value 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 smartcard 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 smartcard will send Harry a message containing \$10 and therefore when she gets the share that the smartcard 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 Electronic Payments — Communication 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.

[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.