The optically coded card: general system- and security aspects

UDC: 351.755.6:535.2:535.411

H. Lienhard

In this issue the first products of the application of coherent optics and of Landis & Gyr-card technology are presented: the PHONOCARD® System and the ID 2000 Access Control System. This section will explain some system concepts as well as a number of information theoretical considerations on the security of such card systems.

1. Introduction

The products mentioned each give one example of two application areas with different card types: the voucher card and the identification card, or ID-card for short.

Both cards contain a certain number of optically coded units which in one case represent *value* units that may be consumed and in the other case *information* units. Basiscally we are dealing with two old acquaintances: the ticket (devalued by punched holes) and the identity card (credit card, etc.). Since these documents may only be made or issued by special authorized departments, the authenticity must be easily checkable and the documents must be difficult to counterfeit. Contrary to these classical cases however, we are interested here only in such documents which are to be accepted and examined exclusively by machines. However complex such machines (referred to as acceptors) may be, compared to humans they possess negligibly little ability in recognizing complex patterns. Hence, in order to attain systems of high security, special requirements for the authenticity features must be made; above all, document and acceptor must be well matched.

The acceptor should refuse genuine documents only very rarely; on the other hand it must reject counterfeits with a high probability. In addition to this, for many applications it should also be inexpensive.

2. W- and E-systems

We distinguish between two fundamentally different systems involving machine readable documents:

W-system (rewriting system) – the document can be modified (rewritten) by the acceptor.

E-system (erasing system) – information can at most be erased in the acceptor.

Thus for the voucher system we have in case W "reloadable cards", while in case E the value units are physically destroyed.

A relatively secure acceptor can be realized with a reasonable amount of effort if selection and quality of the parameters which are relevant for the acceptor ensure the following:

- an imitation of the *authentic* units of the documents is improbable, since it would be technologically difficult or too expensive,
- good discrimination against the most common counterfeits is guaranteed (a measure for the discrimination will be introduced in the appendix, together with possible decision procedures).

If the acceptor is to be generally accessible, these conditions can hardly be attained with the W-system: In the first place the "difficult technology" for the production of these units in such an acceptor is practically out of the question. Secondly, the theft of such a device allows the forger to generate such units himself. W-systems are therefore to be considered inherently insecure (a classic example: the magnetic card). For this reason the Landis & Gyr-card systems, the PHONOCARD and the ID 2000, have both been designed as E-Systems.

3. System levels

Although our money-replacement and ID systems serve different purposes, from a technical standpoint they may be dealt with together. In addition to the value units, the value cards also contain certain ID information (e.g. date of issue); the ID cards on the other hand are individualized by "erasing", though this does not take place in the acceptor, as does the devaluation of the value card, but in a (securely kept) special programming device. We shall represent a still valid value unit or a binary one as 1; and an erased value unit or a binary zero as 0. The still unused value cards and the "blank" ID cards both contain a sequence of 1's, which is then changed by erasing:

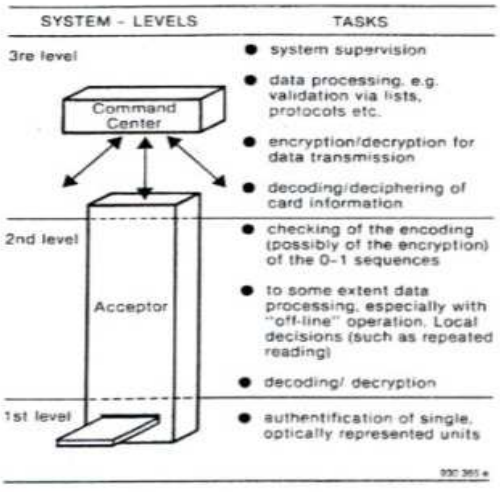
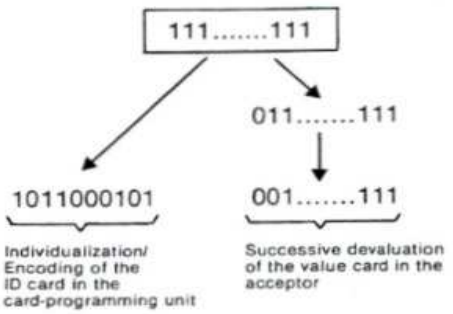


Fig. 1 Logical levels and their tasks in a card system



Essentially three logical levels which are relatively loosely coupled can be identified in the card system (see *fig. 1*):

- *The first level – the optical representation* – stands for the individual information (or value) unit. The unit is *physically represented* on this level.
- *The second level – the code level* – already operates with abstract symbols (0, 1). Sequences of such symbols are combined into words ("bit-strings"). On this level the encoding, perhaps also the encryption (of the card) is examind in the acceptor.
- *The third level – the data processing level* – This level comprises the actual data processing as validation via black-lists, etc. In the case of autonomous acceptors this is ususally done within the microprocessor built into the acceptor; in complete access control systems part of this data processing may be taken over by a central processor.

4. Illegal erasing

On the first level the authenticity of the 1's and 0's is examined. While the generation of authentic 1's (of the optically encoded pattern) is very difficult, therefore lessening the probability of an illegal generation, illegal erasing, i.e. the generation of authentic 0's, is likely to be much easier. The illegal erasing of units on a value card can only decrease the card's value; with the ID card attempts may be made to alter the information (e.g. access right) by erasing. Thus an asymmetry exists in the security of the 0's ans 1's. This asymmetry can be eliminated on the second level, however, by appropriate encoding of the bit sequence. For example, we can use an encoding where the legal bit sequences always have a fixed number of 0's an 1's. Upon reading a card, the number of 0's and 1's is immediately checked on level 2. An example: Let the binary sequence contain 96 bits, and let it be accepted only if it contains exactly 48 0's and 48 1's. There are $\binom{96}{48}$ ($\cong 6.4 \cdot 10^{27}$) such sequences as opposed to 2^{96} ($\cong 8 \cdot 10^{28}$) if no restrictions are imposed. With the "48 out of 96 code" only approximately one decimal place is lost as opposed to the unrestricted binary encoding.

5. Encryption

As indicated above *(fig. 1)* we distinguish two types of encryption: The card encryption and the encryption for the data transmission from the acceptor to any control center. In the first case the main point is the protection of the data bank; in the second case it is the protection of the data transmission. This second encryption can be omitted if the transmission line can be made inaccessible for all practical purposes.

By encrypting the cards one attempts to prevent people who are involved with the production and operation of such a system from misusing knowledge about the data bank (e.g. lists of those authorized for access) and the data processing. Of course the programmer who has programmed the decoding and decryption obviously knows these algorithms. If he also has some knowledge of the representation of legal codes in the data bank, he might be able to guess at the card codes themselves. This can be avoided by the use of so-called "trap-door one-way" functions [1] [6]. *Figure 2* shows how a valid card information (CI) for an access control card might be stored in the data bank. Now the card does not contain this information but an encrypted form of it. This is generated in a (securely kept) programming device with a secret special function f (·):

card code CC' = f (CI).

The inverse function f^{-1} () is applied to CC' in the acceptor, or where relevant in the central unit. In an access control system, for example, we check whether f^{-1} (CC') is present on the list of those authorized for access. f must be chosen in such a manner that without knowledge of the special trap-door information, it is practically impossible to derive f from f^{-1}. Thus even knowledge of the "access list" or "black list" is of little use.

If the connecting line between acceptor and control center in an access control system can be tapped outside of the security area, one can gain illegal entry without using the acceptor: It suffices to register a successful signal sequence (one which causes the doors to open), in order to put such a sequence on the line at the appropriate time. A complicated dialogue between acceptor and control center should make such an operation somewhat more difficult, but basically the situation remains unchanged. The following stipulations must be made for such a dialogue: It should be practically impossible to derive a successful signal sequence E_{n+1} from observing the n signal sequences S_r.

$$S_1, S_2, \ldots\ldots S_n \not\longmapsto E_{n+1}$$

If we choose an encryption that is dependent on the time and the date this can in principle be achieved [1].

The card encryption first mentioned, which is concerned above all with the protection of the system operator from the suppliers of the cards and the acceptor, can be realized in several different ways. One interesting possibility is the use of one-way trap-door functions mentioned above. Here the decryption algorithm can be made completely accessible without endangering the system. However, if a classical block encryption [2] is used, the key K must be kept secret by the system operator. A special "key card", with which the key is read into the acceptor, could be used for transmitting the key securely to the acceptor. *Figure 3* illustrates a possible sequence of encodings/decodings in an ID system.

6. Structure of the acceptor

Basically the acceptor represents a control and decision system; *Figure 4* issustrates the general case. The inserted card is illuminated by the light sources, the resulting radiation pattern is collected by the detectors and converted into analogue electrical signals, which are then further processed in the electronic part. Thus the

COMPANY-CODE		123
IDENTIFICATION		321654987
AREA ZONE		05
TIME ZONE		0
ISSUE NUMBER		0
PIN-GENERATION	NBR SECRET CHAR	0
	ALGORITHM	6

930 366 e

Fig. 2 Card information in clear text for an access control card

sensor signals S_i (t) are mapped onto one point in a multi-dimensional decision space, i.e. onto a vector $\underline{Y}$ of digital values. The actual decision alsgorithm is realized in the microprocessor of the acceptor (see also the appendix). The control and regulation algorithms with which the sources, erasing procedures and card transport are controlled are contained in the microprocessor also. In this way not only the authenticity of the optical pattern, but also e.g. the erasing behavior of the card material can be examined. The acceptor concept chosen provides a great deal of flexibility in addition to optimal security.

7. The decision problem

The decision procedures used must provide a great deal of security against attempts at fraud. At the same time they must be able to be realized in a simple and efficient manner by microcomputer software. Such decision procedures can be derived with the aid of statistical methods similar to those used in estimation theory and hypothesis tests. As mentioned above, the sensor signals are mapped by an electronic preprocessing (by an operator T_v) onto a vector $\underline{Y}$ of a decision space of dimension N. It is practically impossible, if N > 2, to find a good decision procedure purely intuitively. In the appendix we shall show how the minimalization of the expected risk by erroneous decision (the so-called Bayesian risk) leads to an optimal test (equation (7)). The procedure is too unwieldy for implementation in the acceptor; therefore a *suboptimal*, but stricter, test is derived from the optimal one which produces a simple decision procedure. By assuming a Gaussian statistic, the procedure separates the decision space by hyperplanes into three areas: Γ_0, Γ_1 and the remainder. If the vector $\underline{Y}$ is in Γ_0, a 0 is decided upon; if it is in Γ_1, the decision is for a 1. If it is in neither Γ_0 nor Γ_1, the unit is not accepted. Numerically this means a number of inequalities of the form

$$\underline{\alpha}_{ik}^{T}\,\underline{Y} < D_{ki}$$

must be checked where $\underline{\alpha}_{ik}^{T}$ stands for a precalculated row vector and d_{ki} for precalculated constant (equations (17) to (19)). Together with the decision procedure a discrimination measure, the discrimination information, is also introduced in the appendix. This measure permits a quantitative evaluation of the distinguishability of various alternatives, e.g. the distinguishability of a 1 from a certain counterfeit. This measure can be used as an aid for the specification of the document parameters as well as for the design of the acceptor.

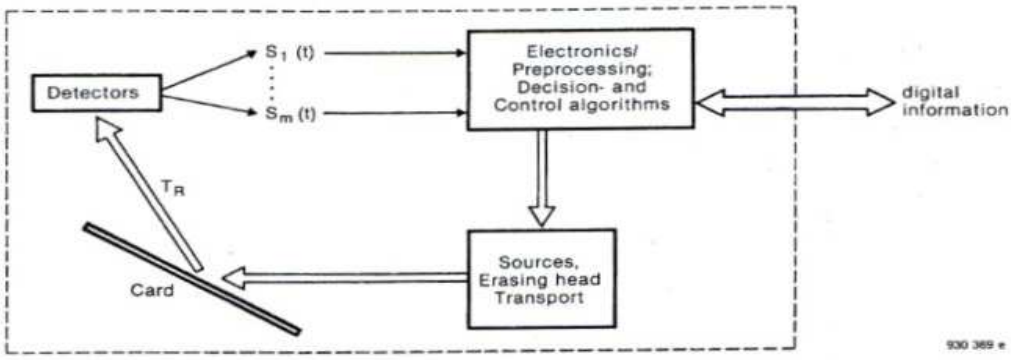


Fig. 4 Structure of the acceptor

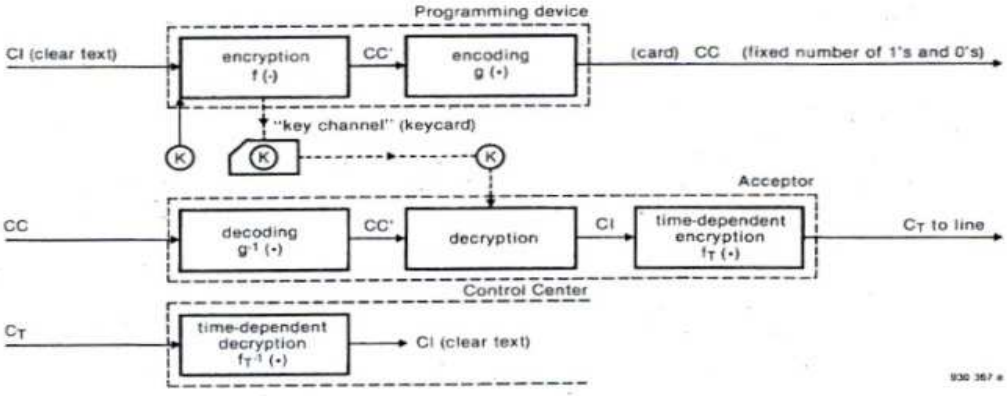


Fig. 3a Card enrcyption by means of block encryption; key K known only to the system operator

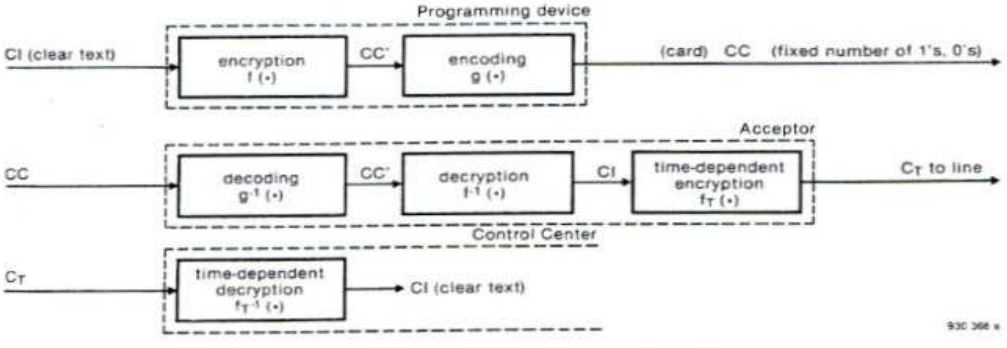


Fig. 3b Card encryption with "one-way trap-door function f", f known only to system operator

8. Connection to the "inverse problem"

(see page 7)

Before a decision procedure is used, we map – as mentioned before – the measured data values with a transformation T_v onto a vector $\underline{Y}$ in the decision space. If the "inverse problem", abbreviated IP, is solvable, i.e. if e.g. the intensity measurements of the far field allow us to identify the optical scatterer, we may use the *representation space* of the scatterer itself as our decision space. We shall limit ourselves here to the parametric case: Let the pattern, or scatterer, be described by a parameter vector $\underline{Y}$ of dimension N. We derive the following mapping from the solution of the "inverse problem".

T_v^*: measurement space → decision space (= representation space)

The IP need not necessarily be solvable for the decision process mentioned in the last paragraph. Instead

it is assumed that the dangerous counterfeits are a priori known. In this case it suffices to protect oneself from these counterfeits; i.e. in the decision space we need sufficient "distance" between authentic patterns and these counterfeits. In order to quantify this "distance" numerically, we use a suitable discrimination measure.

If there is no a priori knowledge of such counterfeits, however, one must proceed differently: One must guarantee that the measurement data, which lead to a decision of "authentic", originate with high probability from only one certain (authentic) pattern. In this case the IP must have a stable, if possible unique, solution. Ambiguity can in principle only be allowed if each of the solutions, i.e. each of the possible patterns, is difficult to produce. If the measurement data are mapped onto a decision space, no discrimination information should be lost in the process. In the decision space the neighbourhoods Γ_0 and Γ_1 can be defined for the acceptance of the legal patterns 0 and 1 by using the discrimination measurement; the environnements are chosen so small that the minimal required acceptance rate for authentic units is just achieved [3]. The requirement of the solvability of the IP calls for extensive measurement data; the procedure is feasible only if the required effort is acceptable. In the acceptor the passive scatterer is illuminated, producing signals in a measurement space $\mathfrak{M}$: let this be the "reading transformation" T_R (*fig. 5*). Further let the representation space of the scatterers (or patterns) be $\mathfrak{D}$, and the decision function be Γ, then the entire acceptance procedure may be shown as in *figure 5*.

9. System separation

We have mentioned two decision procedures:

In one we assume knowledge of dangerous counterfeits, in the other we do not. In reality certain counterfeits are usually known. Thus in a system such as Phonocard (see page 39) the intention is not only protection from actual counterfeits, but also separation from other systems. Naturally, the cards of one country should not be accepted in the Phonocard acceptor of another country. In order to obtain a secure separation of different systems, the units (the 1's on the value cards) are represented differently in each system. The acceptor should be able to recognize cards foreign to its system as known counterfeits and thus to refuse them with high probability.

10. Appendix: Decision procedure and discrimination

At this point we shall try to obtain rational decision procedures by using statistical methods. Intuitive considerations are no longer successful with higher dimensions (N) of the decision space. First we derive a test which is optimal in a certain sense. From this a suboptimal, but stricter, test is developed which is more feasible to be implemented in the accepter.

As illustrated in *figure 6*, the raw sensor signals (vector $\underline{S}(t)$) are preprocessed and transformed into a vector (or point) of a decision space $\mathfrak{G}$ of dimension N.

Transformation

$$T_v: [\underline{S}(t), 0 \leq t \leq T] \longrightarrow \underline{Y} \qquad (1)$$
$$\underline{Y} \in \mathfrak{G}$$

whereby [0, T] represents the observation interval.

Although we wish to concern ourselves in the following with the decision problem only, some remarks regarding the preprocessing, i.e. the transformation $T_v(\cdot)$, are in order:

- As we shall show, the *discrimination information* I (i:j) plays a major role between different hypotheses H_i and H_j in the decision algorithm. Ideally this I should not be decreased by the transformation $T_v(\cdot)$; according to Kullback [3], $\underline{Y}$ would then be a "sufficient statistic" for the discrimination.
- Normally the decision procedure D (see *fig. 6*) is realized in software; on the other hand the preprocessing usually calls for carefully designed analogue hardware.

10.1 The optimal test

In the space $\mathfrak{G}$ subsets Γ_j, which are in a certain sense optimal, are to be defined, so that the hypothesis H_j follows reliably from $\underline{Y} \in \Gamma_j$. To achieve this the Bayesian risk (see [4]) is introduced, which will be minimized.

With E { } = mathematical expectation and Prob (), P () = probability, this risk (i.e. the expected cost) can be written as follows:

$$E\{\text{Cost}\} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} c_{ij} \,\text{Prob}\,(\text{Decision for } H_j, \text{ in reality } H_i)$$

$$= \sum_{i,j} c_{ij} \cdot P(\underline{Y} \in \Gamma_j \text{ and } H_i); \qquad (2)$$

c_{ij} = cost factors

Altogether we consider M different hypotheses.

To illustrate we shall consider a more concrete problem: Let the hypotheses H_0 and H_1 represent *legitimate alternatives* (e.g. 0 or 1 on an access permit or credit card), and let $H_2 \ldots H_{M-1}$ represent counterfeits. With more experience M can later be increased, which merely means a change in the acceptor software.

Theoretically there are of course innumerable false alternatives imaginable. Experience shows however that only those counterfeits which can be produced with relatively little expense can really be dangerous to the

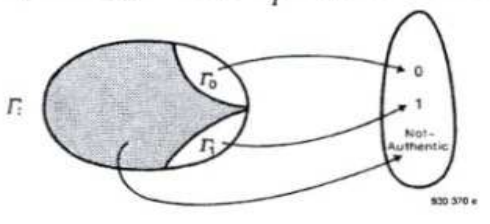


Fig. 5 Acceptance procedure

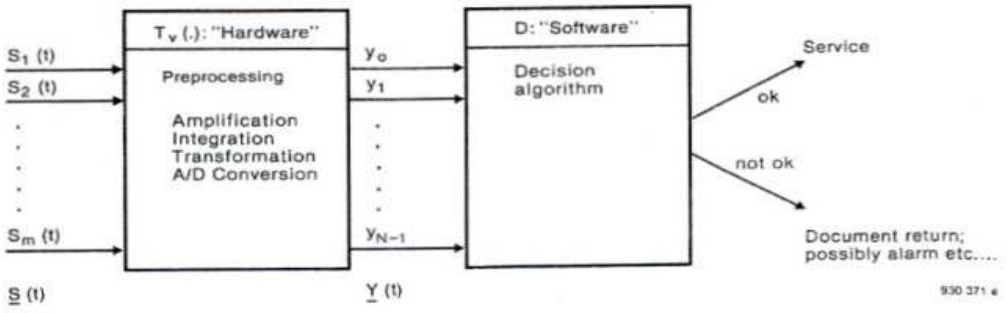


Fig. 6 Structure of the signal processing in the acceptor

system. Thus the $H_2 \ldots H_{M-1}$ should represent the (M-2) most dangerous alternatives. Accordingly we now choose *the cost factors* c_{ij} in (2).

$c_{ii} = 0$:
No loss if the decision is correct.

$c_{ij} = 0$, for $i, j \geq 2$:
No loss if different *false* patterns are confused.

$c_{ij} \ll c_{ji}$, for $i = 0,1$; $j \geq 2$.
We set $c_{ij} = \underline{c}$; $c_{ji} = \overline{c}$

i.e. low cost factor, if the good pattern is refused; but high cost, if the false pattern is accepted.

$c_{01} = c_{10} = \underline{c}$
Low cost factor if 0 and 1 are confused.

or in matrix form:

$$\left[\begin{array}{cc|ccc} 0 & \underline{c} & \underline{c}\ \underline{c} & \cdots\cdots\cdots & \underline{c}\ \underline{c} \\ \underline{c} & 0 & \underline{c}\ \underline{c} & \cdots\cdots\cdots & \underline{c}\ \underline{c} \\ \hline \overline{c} & \overline{c} & & & \\ \overline{c} & \overline{c} & & & \\ \vdots & \vdots & & 0 & \\ \vdots & \vdots & & & \\ \overline{c} & \overline{c} & & & \end{array}\right] = [c_{ij}]$$

By introducing conditional probability densities $p\,(\underline{y} \mid H_i)$ we can write the following for (2):

E {Cost}

$$= \sum_{i,j} \int_{\Gamma_j} c_{ij}\, P(H_i)\, p\,(\underline{y} \mid H_i)\, d\underline{y}$$

$$= \sum_{j} \int_{\Gamma_j} \sum_{i} c_{ij}\, P(H_i)\, p\,(\underline{y} \mid H_i)\, d\underline{y} \qquad (3)$$

In order to minimize the risk, the subsets Γ_j are chosen in such a manner that the integrands in (3) become as small as possible:

$$\underline{Y} \in \Gamma_j \longleftrightarrow \text{all } k \neq j:$$
$$\sum_i c_{ij} P(H_i)\, p(\underline{y} \mid H_i) < \sum_i c_{ik} P(H_i)\, p(\underline{y} \mid H_i) \qquad (4)$$

For the a priori probabilities $P\,(H_i)$ we set:

$$P(H_o) = P(H_1) = p \qquad (5)$$

As already mentioned, we assume zero probability for the hypothese with $i \geq M$. Thus:

$$\sum_{i=2}^{M-1} P(H_i) = 1 - 2\,p \qquad (6)$$

From equations (4), (5) and the assumptions about the coefficients c_{ij} it follows that

$$\underline{Y} \in \Gamma_0 \longleftrightarrow \left\{ \begin{array}{l} p(\underline{y} \mid H_1) < p\,(\underline{y} \mid H_0) \\ \displaystyle\sum_{i=2}^{M-1} P(H_i) \frac{p(\underline{y} \mid H_i)}{p(\underline{y} \mid H_0)} < (\underline{c}/\overline{c})\, p \end{array} \right\}$$

$$\underline{Y} \in \Gamma_1 \longleftrightarrow \left\{ \begin{array}{l} p(\underline{y} \mid H_0) < p\,(\underline{y} \mid H_1) \\ \displaystyle\sum_{i=2}^{M-1} P(H_i) \frac{p(\underline{y} \mid H_i)}{p(\underline{y} \mid H_1)} < (\underline{c}/\overline{c})\, p \end{array} \right\}$$

— otherwise $\longrightarrow$ not accepted. (7)

10.2 The suboptimal test

A simpler, stricter test can be realized with (8):

$$\underline{y} \in \Gamma_0 \longleftrightarrow \left\{ \begin{array}{l} p(\underline{y} \mid H_1) < p(\underline{y} \mid H_0) \\ i \geq 2: \dfrac{p(\underline{y} \mid H_i)}{p(\underline{y} \mid H_0)} < \gamma_i \end{array} \right\}$$

$$\underline{y} \in \Gamma_1 \longleftrightarrow \left\{ \begin{array}{l} p(\underline{y} \mid H_0) < p(\underline{y} \mid H_1) \\ i \geq 2: \dfrac{p(\underline{y} \mid H_i)}{p(\underline{y} \mid H_1)} < \gamma_i \end{array} \right\}$$

— otherwise not accepted. (8)

$$\text{with } \gamma_i = (\underline{c}\,/\,\overline{c}) \cdot \frac{p}{(M\text{-}2)\, P(H_i)} \;;\; (i \geq 2)$$

$$\text{or } \gamma i = (\underline{c}\,/\,\overline{c}) \;\; \frac{p}{1 - 2p} \;;$$

$$\text{if } P(H_i) = \frac{1 - 2p}{M\text{-}2} \qquad (i \geq 2) \qquad (9)$$

(i.e. counterfeits have the same probability).
We set $\gamma_0 = \gamma_1 = 1$ and define:

$$L_{Ki} = \log \frac{p(\underline{y} \mid H_K)}{p(\underline{y} \mid H_i)} = \log \frac{p_K\,(\underline{y})}{p_i(\underline{y})} \;;$$

$$p_i(\underline{y}) = p(\underline{y} \mid H_i) \qquad (10)$$

then instead of (8) we can write:

$$\underline{y} \in \Gamma_0 \longleftrightarrow L_{oi} > \log \gamma_i^1 \text{ for } i \neq 0.$$
$$y \in \Gamma_1 \longleftrightarrow L_{1i} > \log \gamma_i^1 \text{ for } i \neq 1. \qquad (11)$$

We now define the following entity:

$$I(0{:}i) = E\,\{L_{oi} \mid H_o\} =$$

$$\int p_o\,(\underline{y}) \log \frac{p_o(\underline{y})}{p_i(\underline{y})}\, d\underline{y}$$

$$(= I(p_o : p_i)) \qquad (12)$$

We call I(0:i) the *discrimination information* of the hypothesis H_o versus hypothesis H_i [3].

It is indicated in *figure 6* how the acceptance rate (i.e. decision in favor of H_o, if H_o is given) is influenced by the discrimination information. In particular with large values of $\log \gamma_i^1$ (e.g. with a relatively great probability of counterfeit) we need a suitably greater I(o:i) for a reasonable acceptance rate.

10.3 The discrimination information

This is a generalization of Shannon's "mutual information".

It follows immediately from Jensen's inequality (e.g. [5]) that

$$I(p{:}q) \geq o \qquad (13)$$

with $I = o \longleftrightarrow p = q$ (with probability 1 for generalized densities p,q).

In addition the following holds true for statistically independent features $(y_1, y_2, \ldots y_N)$:

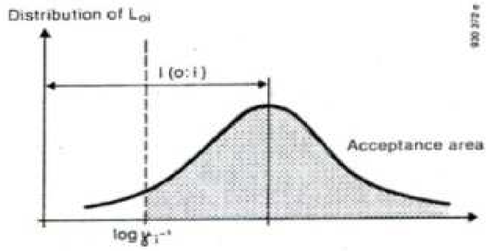
$$I(k{:}i; y_1, y_2 \ldots y_N) = \sum_{l=1}^{N} I(k{:}i; y_l) \qquad (14)$$

(Factorization of densites
$p_k(y_1, \ldots y_N)$; $p_i()$).

The discrimination information is *invariant* under non-singular transformation [3]:

$$\underline{X} = T(\underline{Y}) \longrightarrow I(k{:}i\,; \underline{X}) = I(k{:}i\,; \underline{Y}) \qquad (15)$$

It follows from (13) through (15) that $(I(k{:}i))^{1/2}$ can be interpreted as a *geometric distance* between the hypotheses H_i and H_k after a cor-



Assumption

$p(\underline{y} \mid H_o) = n(r_o, \sigma^2)$; i.e. normal distribution with mean r_o and variance σ^2
$p(\underline{y} \mid H_i) = n(r_i, \sigma^2)$

and let

$$\Phi\,(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2} t^2}\, dt.$$

$$P\{L_{oi} > \log \gamma_i^1 \mid H_o\} = \Phi \left\{ \sqrt{I/2} - \frac{\log \gamma_i^1}{\sqrt{2I}} \right\}$$

$I = I(o{:}i) = E\{L_{oi} \mid H_o\}$

Fig 7: Influence of I(o:i)

responding transformation (rotation) of the vector $\underline{Y}$. The situation becomes particularly transparent in the case of normal distributions with identical covarence:

Let $p_i(\underline{y}) = n(\underline{r}_i, (\overline{II})$ (16)

i.e. normal distributions with mean r_i and the unit matrix as covariance matrix.

From (8) or (11) it immediately follows that

$$\alpha_{ik}^T (\underline{y} - \underline{r}_k) < d_{ki}$$

$$\text{with } \alpha_{ik}^T = (\underline{r}_i - \underline{r}_k)^T / (2\, I(k:i))^{1/2} \quad (17)$$

$$\sqrt{2}\, d_{ki} = (I(k:i))^{1/2} - \log \gamma_i^{-1} / (I(k:i))^{1/2}$$

$$\text{where } 2\, I(k:i) = (\underline{r}_k - \underline{r}_i)^T (\underline{r}_k - \underline{r}_i) =$$

$$\| \underline{r}_k - \underline{r}_i \|^2$$

with $\| \underline{r}_k - \underline{r}_i \|$: Euclidean Vector Distance.

The inequality in (17) simply means that the projection of vector $(\underline{y} - \underline{r}_k)$ onto vector $(\underline{r}_i - \underline{r}_k)$ must be smaller than d_{ki};
in a two dimensional example:

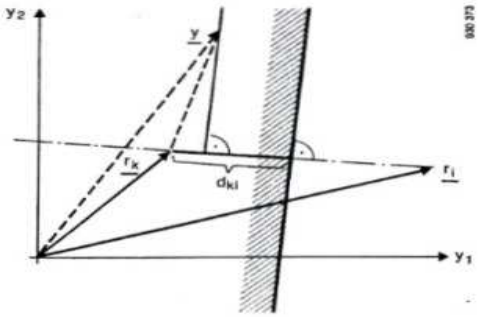


Fig. 8 A two dimensional example for procedure (17)

i.e. $\underline{y}$ must lie within the shaded area. In the N dimensional space the acceptance set is bounded by hyperplanes. This holds true in general for probability distributions which belong to the "exponential" family, that is for many of the frequently occurring distributions such as normal distribution, Poisson's distribution, etc. If we assume normal distribution with a covariance matrix Σ instead of $\overline{II}$ in (16), we get the following in place of (17):

$$\underline{\beta}_{ik}^T (\underline{y} - \underline{r}_k) < d_{ki} \quad (18)$$

$$\text{with } \underline{\beta}_{ik}^T = (\underline{r}_i - \underline{r}_k)^T \Sigma^{-1} / (2\, I(k:i))^{1/2}$$

$$\text{and } 2\, I(k:i) = (\underline{r}_i - \underline{r}_k)^T \Sigma^{-1} (\underline{r}_i - \underline{r}_k) \quad (19)$$

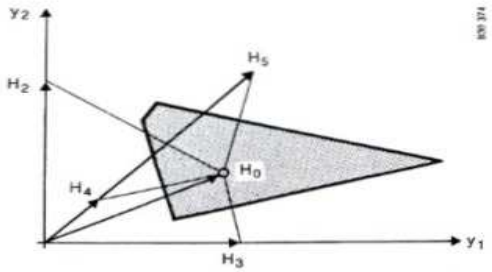


Fig. 9 Example with only one permissible hypothesis H_0. Decision space for H_0: boundaries against H_2, H_3, H_4, H_5

In this case the discrimination information is thus given by the relative position of vectors $\underline{r}_i$, $\underline{r}_k$ and the covariance matrix Σ.

In *figure 9* the simplified case (reduced to 2 dimensions) of a decision situation is shown.

The most dangerous cases are:

- Consideration of only one feature (hypotheses H_2, H_3)
- equal, small values in Y_1 and Y_2 (H_4)
- equal, maximal values in Y_1 and Y_2 (H_5)

All γ_i are set equal to 1 in this case; i.e. the cost ratio in (9) is set equal to the ratio (counterfeit probability) / $P(H_o)$.

This only makes sense if these counterfeit cases already have little probability.

Figure 10 illustrates the case of a 0/1 detection.

10.4 Design-criteria

If all other parameters are fixed, the acceptance probability is changed in relation to the "power of rejection" by variation of $(\underline{c}/\overline{c})$:

By decreasing $(\underline{c}/\overline{c})$ the acceptance rate is decreased; however, the test for counterfeits becomes stricter. In actual applications one must compute a number of cases in order to get a sensible compromise.

The probability ratios $\dfrac{p}{(M\text{-}2)\,P(H_i)}$ (see (9)) are totally unknown in reality. Here we shall have to take an "intelligent guess", dependent upon the technology used.

The *discrimination information* is not only dependent upon the design of the acceptor, but also upon the *nature of the document:* If the document exhibits much dispersion (large Σ) in the parameters to be tested, this information will accordingly be small (see (19)).

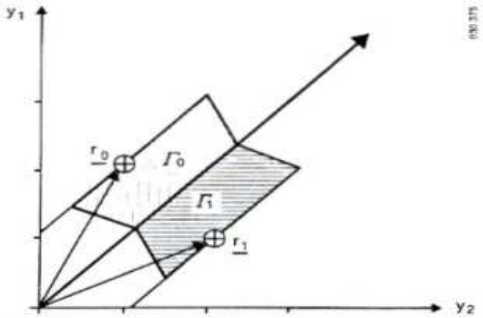


Fig. 10 Example of a 0/1-detection
$Y \in \Gamma_0 \rightarrow$ "0"
$Y \in \Gamma_1 \rightarrow$ "1"
otherwise: counterfeit

How well an acceptor can be designed (with finite cost!) is strongly dependent on the production quality of the documents to be tested.

In conclusion we can summarize:

The *choice* and *quality* of the document parameters should
(1) lead to *minimal imitation probability* (difficult technology);
(2) bring about *large values for the discrimination information* against the most common counterfeits.

11. Bibliography

[1] Diffie, W., Hellman, M.E.: New directions in cryptography; IEEE Transactions on Information Theory, Vol IT-22, No. 6, Nov. 1976.

[2] Feistel, H.; Notz, W.A., Smith, D.L.: Some cryptographic techniques for machine-to-machine data communications. Proceedings of the IEEE, Vol. 63, No. 11, Nov. 1975.

[3] Kullback, S. Information Theory and Statistic (1959), Dover, 1968.

[4] Fukunaga, K. Introduction to Statistical Pattern Recognition, Academic Press NY, 1972.

[5] Feller, W. An Introduction to Probability Theory and its Applications; Vol. II. John Wiley, 1966.

[6] Rivert, R.L.: Shavric, A.: Adleman, L.: A method for obtaining digital signatures and public-key cryptosystems.

Author: Heinz Lienhard
LGZ Landis & Gyr Corporation
CH-6301 Zug (Switzerland)

Translator: Carol Schild
CH-6340 Baar (Switzerland)