Authentication Theory and Hypothesis Testing

Ueli M. Maurer, *Senior Member, IEEE*

Abstract

By interpreting message authentication as a hypothesis testing problem, this paper provides a generalized treatment of information-theoretic lower bounds on an opponent's probability of cheating in one-way message authentication. We consider the authentication of an arbitrary sequence of messages, using the same secret key shared between sender and receiver. The adversary tries to deceive the receiver by forging one of the messages in the sequence. The classical two types of cheating are considered, impersonation and substitution attacks, and lower bounds on the cheating probability for any authentication system are derived for three types of goals the adversary might wish to achieve. These goals are (a) that the fraudulent message should be accepted by the receiver, or, in addition, (b) that the adversary wishes to know or (c) wants to even choose the value of the plaintext message obtained by the legitimate receiver after decoding with the secret key.

Keywords. Cryptography, Authentication, Unconditional security, Hypothesis testing, Impersonation attack, Substitution attack.

I. Introduction

Message authentication is concerned with providing evidence to the receiver of a message that it was sent by a specified legitimate sender, even in the presence of an opponent who can intercept messages sent by the legitimate sender and send a fraudulent message to the receiver. Authenticity (like confidentiality) can be achieved by cryptographic coding based on a secret key shared by sender and receiver.

The security of cryptographic systems can be classified according to the assumed computational resources of an adversary. Security that holds when one assumes a suitable restriction on an adversary's computing power is called computational security while security that holds even when no bound on the adversary's computing power is assumed is called information-theoretic security.

This paper is concerned with information-theoretically secure message authentication, i.e., we consider a scenario in which the opponent has unlimited computing power and knows everything about the system, except for the secret key. We consider bounds on how efficiently a secret key shared by sender and receiver can be used or, more precisely, we derive lower bounds on an opponent's cheating probability that no authentication system with a given key size can overcome.

Compared to the theory of secrecy, definitions in authentication theory are more subtle. For instance, while Shannon's definition of perfect secrecy [15], which means that ciphertext and plaintext are statistically independent, is obviously the strongest possible definition of secrecy, it is not clear how perfect authenticity should be defined; the cheating probability can be made arbitrarily small by using a secret key of sufficient size, but it can never be reduced to zero. Shannon [15] proved the well-known result that for any perfect cipher the secret key must be at least as long as the plaintext or, more precisely, that $H(K) \geq H(M)$ where M and K denote the message and the secret key, respectively. We refer to [8] for a generalization of Shannon's treatment of secrecy.

The first lower bound results in message authentication were purely combinatorial [4], [3]. In 1984, Simmons [16] initiated a sequence of research activities on information-theoretic lower bounds in authentication theory [2], [5], [6], [7], [10], [11], [12], [13], [14], [17], [18], [19], [21].

The results of this paper were presented in part at the 13th Symposium on Theoretical Aspects of Computer Science (STACS'96), Grenoble, France, Feb. 1996, and have appeared in the proceedings.

Author's address: Department of Computer Science, ETH Zürich, CH-8092 Zürich, Switzerland. E-mail address: maurer@inf.ethz.ch

The goal of this paper is give a generalized and considerably simplified treatment of lower bound results in authentication theory. The key to this simplification is the natural observation that authentication, i.e. deciding whether a received message is authentic or not, is a hypothesis testing problem. The receiver is faced with two possible hypotheses: either the message was generated by the legitimate sender knowing the secret key, or by an opponent without *a priori* knowledge of the secret key. The joint probability distribution of the authenticated message and the secret key is different in both cases, and this may allow the receiver to distinguish between the two hypotheses. The goal of the paper is similar in spirit to some of Sgarro's work (cf. [13], [14]) who also investigated a general approach to authentication frauds, based on rate-distortion theory, showing that some of the known bounds follow from a more general result.

Like Shannon's lower bounds [15] on the size of a secret key of a perfect secrecy system, the bounds of this paper show the impossibility of obtaining an arbitrary amount of security with a secret key of fixed size. Constructions of authentication systems are not discussed in this paper, but there exists a substantial body of literature on that topic.

In Section II we discuss the message authentication problem and various cheating strategies. The literature is briefly reviewed in Section III, and hypothesis testing is described in Section IV. Lower bounds on the cheating probability for impersonation and substitution attacks are derived in Sections V and VI, respectively, and these results are combined in the concluding Section VII.

II. Message Authentication and Cheating Strategies

Consider a scenario in which a sender and a receiver share a secret key K. The sender wants to send a sequence of plaintext messages $M_1, M_2, \ldots, M_n$, at some independent time instances, in an authenticated manner to the receiver. Each message M_i is authenticated separately by sending an encoded message or ciphertext[1] C_i which depends (possibly probabilistically) on K and M_i. In order to be fully general, and in contrast to the previous literature, we also allow C_i to depend on the previous plaintext messages $M_1, \ldots, M_{i-1}$ and the previous ciphertexts $C_1, \ldots, C_{i-1}$.

Based on C_i and K, and possibly also on $M_1, \ldots, M_{i-1}$ and $C_1, \ldots, C_{i-1}$, the receiver decides to either reject the ciphertext or accept it as authentic and, in case of acceptance, decodes C_i to a message $\hat{M}_i$. It is assumed that the receiver is synchronized, i.e., he knows the message number i. Without loss of generality we assume that M_i is uniquely determined by $M_1, \ldots, M_{i-1}$, $C_1, \ldots, C_i$ and K and hence, by induction, also by $C_1, \ldots, C_i$ and K alone:

$$H(M_1 \cdots M_i | C_1 \cdots C_i K) = 0.$$

In typical authentication systems, the message M_i is often determined by the ciphertext C_i, i.e., the system provides no secrecy. In such a system without secrecy we have $H(M_i|C_i) = 0$ or, more generally,

$$H(M_1 \cdots M_i | C_1 \cdots C_i) = 0.$$

Authentication schemes without secrecy are often called Cartesian. Our analysis applies to authentication codes that do or do not provide secrecy, be it full or partial.

Following standard practice it is assumed that an adversary knows everything about the system, including the codes used and the plaintext statistics, but that he has no *a priori* information about the secret key. To remove a possible source of confusion it should be pointed out that in the authentication literature plaintext message and ciphertext are also referred to as source state and message, denoted by S and M, respectively.

The adversary is assumed to have full (read and write) access to the communication channel. The adversary's strategy is to select an arbitrary (e.g. optimal) time index i of the ciphertext to be forged. Then he chooses between two different types of attacks.

[1]The term ciphertext, is rarely used in the literature on authentication, but it is justified because our results apply both to systems that do or do not provide secrecy.

• In a so-called *impersonation attack* at time i, the adversary waits until he has seen the ciphertexts $C_1, \ldots, C_{i-1}$ (which he lets pass unchanged to the receiver) and then creates and sends a fraudulent ciphertext $\tilde{C}_i$ which he hopes to be accepted by the receiver as the ith ciphertext C_i.
• In a so-called *substitution attack* at time i, the adversary lets pass ciphertexts $C_1, \ldots, C_{i-1}$, intercepts C_i, and replaces it by a different ciphertext $\tilde{C}_i$ which he hopes to be accepted by the receiver. In a substitution attack, an adversary can of course only be considered successful when $\tilde{C}_i$ is decoded by the receiver to a plaintext message $\hat{M}_i$ different from the message M_i actually sent by the sender.

In order to define what it means for an adversary to be successful in an impersonation or a substitution attack, we can distinguish between three different goals the adversary might want to achieve, which are described below. In the previous literature, only the first of these cases has been considered. The adversary can be considered successful when
(a) the receiver accepts $\tilde{C}_i$ as a valid ciphertext.
(b) the receiver accepts $\tilde{C}_i$ as a valid ciphertext and decodes it to a message $\hat{M}_i$ *known* to the adversary. In other words, an adversary is only considered successful if he also guesses the receiver's decoded message $\hat{M}_i$ correctly.
(c) For a given fixed message x (on which the attack depends), the receiver accepts C_i as a valid ciphertext and decodes it as $\hat{M}_i = x$.
Note that cases (b) and (c) differ from (a) only when the plaintext message is not contained in (or uniquely determined by) the ciphertext, i.e., when the system also provides some degree of secrecy and hence $H(M_i|C_1 \ldots C_i) > 0$. These three cases apply to both the impersonation and the substitution attack.

For an impersonation attack at time i on a given authentication scheme we will denote the probabilities of success for an optimal attack, for the three described scenarios, by $P_{I,i}$, $P'_{I,i}$ and $P''_{I,i,m}(x)$, respectively. Similarly, for a substitution attack at time i on a given authentication scheme we will denote the probabilities of success for an optimal attack, for the three described scenarios, by $P_{S,i}$, $P'_{S,i}$ and $P''_{S,i,m}(x)$, respectively. However, the same bound will apply to all of these probabilities, and hence they need not be distinguished.

When considering the same success probabilities for a particular observed sequence $C_1 = c_1, \ldots, C_{i-1} = c_{i-1}$ of ciphertexts and, in case of a substitution attack also for a fixed intercepted ciphertext $C_i = c_i$, then we denote the corresponding probabilities by $P_{I,i}(c_1, \ldots, c_{i-1})$, $P'_{I,i}(c_1, \ldots, c_{i-1})$ and $P''_{I,i,m}(x, c_1, \ldots, c_{i-1})$, respectively, for an impersonation attack at time i, and by $P_{S,i}(c_1, \ldots, c_i)$ for a substitution attack at time i. Note that for instance $P_{I,i}$ is the average of $P_{I,i}(c_1, \ldots, c_{i-1})$ over choices of $c_1, \ldots, c_{i-1}$, i.e.

$$P_{I,i} = \sum_{(c_1,\ldots,c_{i-1})} P_{C_1 \cdots C_{i-1}}(c_1, \ldots, c_{i-1}) \cdot P_{I,i}(c_1, \ldots, c_{i-1}).$$

III. Previous Results and Their Limitations

The significance of a lower bound result in authentication theory depends on the generality of the considered model. Instead of reviewing in detail the various papers on the subject, we briefly summarize the restrictions of the existing results and the generalizations achieved in this paper. We refer to [7], [10] and [21] for reviews of the literature on the subject.
• Some papers consider the authentication of only a single message [4], [5], [7], [11], [16], [20]. Most of the papers dealing with the authentication of several plaintext messages $M_1, M_2, \ldots$ consider only schemes that apply the same encoding rule to every plaintext message M_i, thus assuming that all plaintext messages are different and belong to the same message space [3], [10], [21]. The assumption that messages be different is necessary to prevent replay attacks in this model, but it appears to be quite unnatural. The only previous papers considering time-dependent encoding rules are [17], [18], and [19].
• Some papers are restricted to deterministic encoding rules referred to as authentication codes without

splitting [3], [4], [5], [16], [18], [21]. In this paper there is no need to distinguish between deterministic and probabilistic encoding rules.

• Some papers are restricted to authentication without secrecy, i.e. where the ciphertext uniquely determines the plaintext message [3], [16], [20], [18], [21]. Such schemes are sometimes referred to as Cartesian.

• In all previous papers it is assumed that the receiver never makes an error when seeing a valid ciphertext, i.e., that his strategy is to accept a ciphertext if and only if it is consistent with the given secret key K. Our results are more general in that we also provide bounds on an adversary's cheating probability when a certain tolerable non-zero probability of rejecting a valid ciphertext is specified. While this generalization does not appear to be of much practical interest, it is useful because it establishes the link to the standard hypothesis testing scenario.

Our results hold in a general model without any of the discussed restrictions. Moreover, we need not assume that $M_1, M_2, \ldots$ are independent and we can allow the encoding rule for message M_i to depend on the previous plaintext messages $M_1, \ldots, M_{i-1}$. Furthermore, as discussed above, the realistic alternative models in which an adversary is considered successful only when he knows (or can choose) the plaintext message to which the receiver decodes the fraudulent ciphertext, have not been considered previously.

IV. Hypothesis Testing

Hypothesis testing is the task of deciding which of two hypotheses, H_0 or H_1, is true, when one is given the value of a random variable U (e.g., the outcome of a measurement). The behavior of U is described by two probability distributions: If H_0 or H_1 is true, then U is distributed according to the distribution $P_{U|H_0}$ or $P_{U|H_1}$, respectively. For ease of notation we will write $P_{U|H_0} = P_U$ and $P_{U|H_1} = Q_U$.

A decision rule assigns one of the two hypotheses to each possible value u that U can assume. There are two types of possible errors in making a decision. Accepting hypothesis H_1 when H_0 is actually true is called a type I error, and the probability of this event is denoted by α. Accepting hypothesis H_0 when H_1 is actually true is called a type II error, and the probability of this event is denoted by β. The optimal decision rule is given by the famous Neyman-Pearson theorem which states that, for a given maximal tolerable probability β of type II error, α can be minimized by assuming hypothesis H_0 if and only if

$$\log \frac{P_U(u)}{Q_U(u)} \geq T \tag{1}$$

for some threshold T depending on α. Here and in the sequel, logarithms are understood to be taken to the base 2. Note that only the existence of T, but not its value is specified by this theorem. The term on the left of (1) is called the log-likelihood ratio. We refer to [1] for an excellent treatment of hypothesis testing.

Let P_U and Q_U be arbitrary probability distributions over the same finite or countably infinite set $\mathcal{U}$. The expected value of the log-likelihood ratio with respect to P_U is called the discrimination and is defined by

$$L(P_U; Q_U) = \sum_{u \in \mathcal{U}} P_U(u) \log \frac{P_U(u)}{Q_U(u)}.$$

The discrimination is non-negative and is equal to zero if and only if the two distributions are identical.

A well-known result in hypothesis testing (cf. [1], Theorem 4.4.1[2]) provides a relation between the error probabilities α and β and the discrimination $L(P_U; Q_U)$. Let the function $d(\alpha, \beta)$ be defined by

$$d(\alpha, \beta) \triangleq \alpha \log \frac{\alpha}{1-\beta} + (1-\alpha) \log \frac{1-\alpha}{\beta}$$

[2]Note that in our formulation of this result we have exchanged α and β as well as P_U and Q_U.

$$= \quad -h(\alpha) - \alpha \log(1-\beta) - (1-\alpha)\log\beta.$$

where

$$h(\alpha) \stackrel{\triangle}{=} -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$$

is the binary entropy function.

Lemma 1: The type I and type II error probabilities α and β satisfy

$$d(\alpha, \beta) \;\leq\; L(P_U; Q_U).$$

In particular, for $\alpha = 0$ we have $-\log\beta \leq L(P_U; Q_U)$ which is equivalent to

$$\beta \geq 2^{-L(P_U;Q_U)}.$$

Consider the special case of hypothesis testing where $U = [S, T]$ consists of a pair of random variables S and T, where $P_U = P_{ST}$ is the actual joint distribution of this pair and where $Q_U = P_S P_T$ is the product of the two marginal distributions. This special case will be important in the analysis of impersonation attacks. Note that P_{ST} and $P_S P_T$ are both probability distributions over the same set $\mathcal{S} \times \mathcal{T}$ (where S and T take on values in $\mathcal{S}$ and $\mathcal{T}$, respectively). We have

$$\begin{aligned} L(P_{ST}; P_S P_T) &= \sum_{s,t} P_{ST}(s,t) \log \frac{P_{ST}(s,t)}{P_S(s) P_T(t)} \\ &= H(S) + H(T) - H(ST) \\ &= I(S;T) \end{aligned} \tag{2}$$

where the standard definitions of the entropy $H(S)$ of a random variable S, of the mutual information $I(S;T)$ between two random variables S and T, and (below) of conditional entropy and conditional mutual information are used. We also refer to [1] for an excellent introduction to information theory. The second and third step of (2) follow from these definitions. We have

$$L(P_{ST}; P_S P_T) = 0$$

if and only if the two distributions P_{ST} and $P_S P_T$ are identical, i.e., if and only if S and T are statistically independent. This fact is needed for deriving the conditions for equality in the lower bounds which, however, we consider not sufficiently interesting to state in this paper.

The above considerations can be generalized to a scenario where the testing person knows a random variable V, which can be considered as side information. In other words, we consider a collection of pairs $(P_{U|V=v}, Q_{U|V=v})$ of distributions, each pair occurring with probability $P_V(v)$. The hypothesis testing strategy may depend on the value v of V, and for each v we can define $\alpha(v)$ and $\beta(v)$ as the error probabilities of type I and II, respectively, given that $V = v$. An alternative form of Lemma 1 is hence

$$d(\alpha(v), \beta(v)) \;\leq\; L(P_{U|V=v}; Q_{U|V=v}). \tag{3}$$

Equation (2), conditioned on $V = v$, becomes:

$$L(P_{ST|V=v}; P_{S|V=v} P_{T|V=v}) = I(S;T|V=v) \tag{4}$$

The following lemma provides a lower bound similar to Lemma 1, where α and β are taken as the average (over values of V) error probabilities.

Lemma 2: The average error probabilities of type I and II,

$$\alpha = \sum_v P_V(v)\alpha(v)$$

and

$$\beta = \sum_v P_V(v)\beta(v),$$

respectively, satisfy

$$d(\alpha, \beta) \ \leq \ \sum_v P_V(v) L(P_{U|V=v}; Q_{U|V=v}).$$

Proof: The function $d(\alpha, \beta)$ is a convex-$\cup$ function in both its arguments and hence one can apply Jensen's inequality (cf. [1]) to conclude that

$$d(\alpha, \beta) \ \leq \ \sum_v P_V(v) d(\alpha(v), \beta(v)).$$

Now using (3) completes the proof. $\square$

In analogy to above, consider the special case of hypothesis testing where $U = [S, T]$ consists of a pair of random variables S and T whose distribution depends on a random variable V, and consider the collection of pairs of distributions

$$(P_{U|V=v}, Q_{U|V=v}) = (P_{ST|V=v}, P_{S|V=v} P_{T|V=v}),$$

each pair occurring with probability $P_V(v)$. Then the expression on the right side of the inequality in Lemma 2 becomes

$$\begin{aligned} \sum_v P_V(v) L(P_{U|V=v}; Q_{U|V=v}) &= \sum_v P_V(v) L(P_{ST|V=v}, P_{S|V=v} P_{T|V=v}) \\ &= \sum_v P_V(v) I(S; T|V = v) \\ &= I(S; T|V). \end{aligned} \tag{5}$$

V. Impersonation Attacks

We now return to the analysis of message authentication, in particular the impersonation attack. The problem of deciding whether a received ciphertext $\tilde{C}$ is authentic or not can be viewed as a hypothesis testing problem. H_0 corresponds to the hypothesis that the ciphertext is authentic, and H_1 corresponds to the hypothesis that the ciphertext has been generated by an adversary. Referring to Section IV, we are interested in proving lower bounds on β, for a given tolerated upper bound on α. Such a result is stated in the form

$$d(\alpha, \beta) \leq B$$

for some bound B which for $\alpha = 0$ implies $-\log \beta \leq B$ or, equivalently,

$$d(\alpha, \beta) \leq B \quad \Longrightarrow \quad \beta \geq 2^{-B}. \tag{6}$$

Consider an impersonation attack on the i-th message M_i. The receiver knows K and the ciphertexts $C_1 = c_1, \ldots, C_{i-1} = c_{i-1}$, and sees a ciphertext $\overline{C}_i$, which could either be a correct ciphertext $\overline{C}_i = C_i$ sent by the legitimate receiver (hypothesis H_0) or a fraudulent ciphertext $\overline{C}_i = \tilde{C}_i$ inserted by an adversary (hypothesis H_1). A potential adversary would choose $\tilde{C}_i$ depending on the observed particular ciphertexts $C_1 = c_1, \ldots, C_{i-1} = c_{i-1}$, but without further knowledge about the secret key.

In its most general form, an adversary's strategy for impersonation at time i can hence be described by an arbitrary probability distribution $Q_{\tilde{C}_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}$, where we have used the symbol Q instead of P to make explicit that this is a different random experiment than that induced by a legitimate use of the system. Note that in a deterministic (non-splitting) strategy, $Q_{\tilde{C}_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}(c_i)$ is equal to 1 for one particular value c_i, and zero otherwise.

In the sequel, consider probability distributions conditioned on the event that $C_1 = c_1, \ldots, C_{i-1} = c_{i-1}$. Under hypothesis H_0, the pair $[\overline{C}_i, K]$ (seen by the receiver) is generated according to the probability distribution

$$P_{C_iK|C_1=c_1,\ldots,C_{i-1}=c_{i-1}},$$

whereas under hypothesis H_1, $[\overline{C}_1, K]$ is generated according to the distribution

$$Q_{\tilde{C}_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}} \cdot P_{K|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}.$$

The following theorem generalizes results of several papers, including those by Walker [21], Rosenbaum [10], and Smeets [18]. Recall the definitions of $P_{I,i}$ and $P_{I,i}(c_1, \ldots, c_{i-1})$ from Section II.

Theorem 3: For every authentication scheme and for every particular values $c_1, \ldots, c_{i-1}$ of observed ciphertexts, we have

$$d(\alpha, P_{I,i}(c_1, \ldots, c_{i-1})) \ \leq \ I(C_i; K|C_1 = c_1, \ldots, C_{i-1} = c_{i-1}).$$

Moreover,

$$d(\alpha, P_{I,i}) \ \leq \ I(C_i; K|C_1 \cdots C_{i-1}).$$

In particular, for $\alpha = 0$ we have

$$P_{I,i}(c_1, \ldots, c_{i-1}) \ \geq \ 2^{-I(C_i;K|C_1=c_1,\ldots,C_{i-1}=c_{i-1})}$$

and

$$P_{I,i} \ \geq \ 2^{-I(C_i;K|C_1\cdots C_{i-1})}.$$

It it worth discussing this last bound briefly. For the special case of authenticating only a single message M by the ciphertext C, it is Simmon's bound [16] which states that the impersonation probability is lower bounded by $2^{-I(C;K)}$. This means that unless C leaks a substantial amount of information about the secret key, authenticity is not achievable. Of course, C should not give all information about K since this would allow an adversary to guess K and successfully launch a substitution attack. For the general case of n messages, the theorem states that every ciphertext C_i must give additional information about the key, otherwise the cheating probability in an impersonation attack is small.

Proof. It suffices to describe one strategy that allows the adversary to achieve the cheating probabilities stated in the theorem. One admissible (but generally not optimal) strategy is for the adversary to choose $\tilde{C}_i$ according to the probability distribution

$$Q_{\tilde{C}_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}} = P_{C_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}.$$

Observe that the distribution $P_{C_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}$ does not depend on the particular choice of the secret key and hence is known to the adversary.

The first inequality of the theorem then follows from equations (3) and (4) by letting $V = [C_1, \ldots, C_{i-1}]$, $v = [c_1, \ldots, c_{i-1}]$, $U = [C_i, K]$, $S = C_i$, $T = K$, and hence

$$P_{U|V=v} = P_{C_iK|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}$$

and

$$Q_{U|V=v} = P_{C_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}} \cdot P_{K|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}.$$

The second inequality follows from Lemma 2, using equations (4) and (5). The third and fourth inequalities follow by application of equation (6). □

Consider now scenario (b) mentioned in Section II, i.e., in addition to having $\tilde{C}_i$ accepted by the receiver the adversary also wants to know the message $\hat{M}_i$ the receiver decodes it to.

Theorem 4: For every authentication scheme and for every particular values $c_1, \ldots, c_{i-1}$ of observed ciphertexts, we have

$$d(\alpha, P'_{I,i}(c_1, \ldots, c_{i-1})) \;\leq\; I(M_iC_i; K|C_1 = c_1, \ldots, C_{i-1} = c_{i-1}).$$

Moreover,

$$d(\alpha, P'_{I,i}) \;\leq\; I(M_iC_i; K|C_1 \cdots C_{i-1}).$$

In particular, for $\alpha = 0$ we have

$$P'_{I,i}(c_1, \ldots, c_{i-1}) \;\geq\; 2^{-I(M_iC_i;K|C_1=c_1,\ldots,C_{i-1}=c_{i-1})}$$

and

$$P'_{I,i} \;\geq\; 2^{-I(M_iC_i;K|C_1\cdots C_{i-1})}.$$

Proof. The proof is along the lines of the proof of Theorem 3. Again, it suffices to describe one strategy that allows the adversary to achieve the cheating probabilities stated in the theorem. One admissible strategy is to choose the pair $[\hat{M}_i, \tilde{C}_i]$ according to the distribution

$$Q_{\hat{M}_i\tilde{C}_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}} = P_{M_iC_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}.$$

Note that although in reality the plaintext distribution may not even be known to the legitimate users of the system, we must for the analysis nevertheless assume that the distribution $P_{M_iC_i|C_1=c_1,\ldots,C_{i-1}=c_{i-1}}$ is (at least approximately) known to the adversary. The proof is now completed as in the proof of Theorem 3, replacing all occurrences of C_i by M_iC_i. □

We observe that since

$$I(M_iC_i; K|C_1 \cdots C_{i-1}) \geq I(C_i; K|C_1 \cdots C_{i-1}),$$

the bounds in Theorem 4 are generally smaller than those in Theorem 3. This means, as can be expected, that cheating is less likely to succeed if the adversary also needs to know the message to which his fraudulent ciphertext is decoded to. In a system without secrecy, i.e. when M_i is determineed by C_i, there is no difference between the two settings.

Consider now scenario (c) mentioned in Section II, i.e., in addition to having $\tilde{C}_i$ accepted by the receiver, the adversary also wants the decoded message $\hat{M}_i$ to be equal to a particular value x.

The proof of Theorem 4 also holds when everything is conditioned on the further event $M_i = x$. The bounds of the following corollary depend on the particular value x. Some x may allow a higher success probability than others, and of course the maximum over all values x is also a lower bound on $P_{I,i}(c_1, \ldots, c_{i-1})$.

Corollary 5: For every authentication scheme, for every particular values $c_1, \ldots, c_{i-1}$ of observed ciphertexts, and for every message x chosen by the adversary, we have

$$d(\alpha, P''_{I,i,m}(x, c_1, \ldots, c_{i-1})) \;\leq\; I(C_i; K|C_1 = c_1, \ldots, C_{i-1} = c_{i-1}, M_i = x).$$

Moreover,

$$d(\alpha, P''_{I,i,m}(x)) \;\leq\; I(C_i; K|C_1 \cdots C_{i-1}, M_i = x).$$

In particular, for $\alpha = 0$ we have

$$P''_{I,i,m}(x, c_1, \ldots, c_{i-1}) \;\geq\; 2^{-I(C_i;K|C_1=c_1,\ldots,C_{i-1}=c_{i-1},M_i=x)}$$

and

$$P''_{I,i,m} \;\geq\; 2^{-I(M_iC_i;K|C_1\cdots C_{i-1},M_i=x)}.$$

Some lower bounds of this section for which $\alpha = 0$ could be strengthened further along the lines of [5] where it is suggested to optimize the bounds over choices of the distribution of the message, resulting in expressions involving the infimum of the mutual information between ciphertext and key, where the infimum is taken over choices of the plaintext distribution not changing the set of acceptable ciphertexts.

VI. Substitution Attacks

The bounds on the success probability in an impersonation attack show that in order for an authentication system to be secure, the ciphertexts $C_1, \ldots, C_i$ must give substantial information about the secret key K, thus reducing the adversary's uncertainty about K. One particular type of substitution attack is to try to guess the secret key K. If this succeeds, then any of the three forms of the substitution attack succeeds.

In this section we therefore first derive lower bounds on an adversary's probability of guessing the correct value of K, given certain side information, which immediately yields a lower bound on the cheating probability for any of the substitution attacks.

Let S be a random variable. The entropy $H(S)$ is the expected value of $-\log P_S(S)$. Because the minimum of the values occurring in the averaging, namely $\min_s(-\log P_S(s))$, is upper bounded by the average, it is straight-forward to prove that

$$\min_s(-\log P_S(s)) = -\log(\max_s P_S(s)) \leq H(S)$$

and hence that the probability of guessing S correctly when knowing only P_S is lower bounded by

$$\max_s P_S(s) \geq 2^{-H(S)}.$$

Similarly, when considering side information given in form of the random variable T, where $P_{S|T}$ is known, we have

$$\max_s P_{S|T=t(s)} \geq 2^{-H(S|T=t)}$$

for all t, and the average (over values of T) success probability of guessing S when given T is

$$\sum_t P_T(t) \max_s P_{S|T}(s,t) \geq 2^{-H(S|T)}$$

as can easily be shown by application of Jensen's inequality. This leads to the following theorem:

Theorem 6: For all $c_1, \ldots, c_i$ we have

$$P_{S,i}(c_1, \ldots, c_i) \geq 2^{-H(K|C_1=c_1,\ldots,C_i=c_i)}. \tag{7}$$

Moreover,

$$P_{S,i} \geq 2^{-H(K|C_1 \cdots C_i)}. \tag{8}$$

These bounds also hold for the other two types (b) and (c) of substitution attacks.

VII. Conclusions

The results of this paper can be combined as described in the following. When a sequence of n messages, $M_1, \ldots, M_n$, is to be authenticated using the same secret key K, an adversary could choose the type of attack with the highest success probability. A secret key K is used optimally when all these probabilities are (roughly) equal. When $\alpha = 0$ is required in all of these possible attacks, we obtain

$$\begin{aligned} -\sum_{i=1}^{n} \log P_{I,i} - \log P_{S,n} &\leq \sum_{i=1}^{n} I(C_i; K|C_1 \cdots C_{i-1}) + H(K|C_1 \cdots C_n) \\ &= H(K). \end{aligned}$$

This follows from

$$I(C_i; K|C_1 \cdots C_{i-1}) = H(K|C_1 \cdots C_{i-1}) - H(K|C_1 \cdots C_i).$$

The following theorem follows from the fact that $-log(.)$ is a convex-$\cup$ function. It generalizes results of Walker [21] and Rosenbaum [10].

Theorem 7: For every authentication scheme for authenticating n messages $M_1, \ldots, M_n$ in which the legitimate receiver never rejects a valid ciphertext, we have

$$\max(P_{I,1}, , \ldots, P_{I,n}, P_{S,n}) \geq 2^{-H(K)/(n+1)}.$$

This theorem states that if the adversary can choose between an impersonation attack at any time i, for $1 \leq i \leq n$, or a substitution attack at time n, then no authentication system can prevent him from being successful with probability at least $2^{-H(K)/(n+1)}$. Note that the lower bound for the cheating probability in a substitution attack at time i increases with i. Therefore it makes no sense to include the terms $P_{S,1}, \ldots, P_{S,n-1}$ in the maximization. This can be interpreted as follows: In an optimal scheme, the key is split into $n+1$ parts, where the first n parts are used to protect against an impersonation attack at times $1 \leq i \leq n$, and the last part is used to protect against a substitution attack at the end.

For the special case of authenticating only a single message with the key K (i.e. $n = 1$), this means that half of the key is used for protecting against an impersonation attack, and the other half for protecting against a substitution attack. The success probability is at least one over the square root of the size of the key space:

$$\max(P_{I,1}, P_{S,1}) \geq 2^{-H(K)/2} \geq \frac{1}{\sqrt{|\mathcal{K}|}},$$

where $|\mathcal{K}|$ is the key space. This special case of our bounds confirms a combinatorial result already known Gilbert et al. [4].

Acknowledgments

The author is grateful to Jim Massey for introducing him to authentication theory, to Andrea Sgarro for providing certain references, and to the anonymous referees for their comments.

References

[1] R. E. Blahut, *Principles and practice of information theory*, Addison-Wesley, 1987.
[2] E.F. Brickell, A few results in message authentication, *Congressus Numerantium*, vol. 43, pp. 141–154, 1984.
[3] V. Fåk, Repeated use of codes which detect deception, *IEEE Trans. on Information Theory*, Vol. 25, No. 2, 1979, pp. 233–234.
[4] E. N. Gilbert, F. J. MacWilliams, and N. J. A. Sloane, Codes which detect deception, *Bell Syst. Tech. J.*, Vol. 53, No. 3, 1974, pp. 405–424.
[5] R. Johannesson and A. Sgarro, Strengthening Simmons' bound on impersonation, *IEEE Trans. on Information Theory*, Vol. 37, No. 4, 1991, pp. 1182–1185.
[6] T. Johansson, Lower bounds on the probability of deception in authentication with arbitration, *IEEE Transactions of Information Theory*, Vol. 40, No. 5, 1994, pp. 1573–1585.
[7] J.L. Massey, Contemporary cryptology – an Introduction, in *Contemporary cryptology – the science of information integrity*, G.J. Simmons (Ed.), IEEE Press, 1992.
[8] U.M. Maurer, Secret key agreement by public discussion from common information, *IEEE Transactions on Information Theory*, Vol. 39, No. 3, pp. 733-742, May 1993.
[9] R. L. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, *Communications of the ACM*, Vol. 21, No. 2, 1978, pp. 120–126.
[10] U. Rosenbaum, A lower bound on authentication after having observed a sequence of messages, *J. of Cryptology*, Vol. 6, No. 3, 1993, pp. 135–156.
[11] A. Sgarro, Information divergence bounds for authentication codes, *Advances in Cryptology – Eurocrypt '89*, J.-J. Quisquater and J. Vandewalle (Eds.), Lecture Notes in Computer Science, No. 434. Berlin: Springer Verlag, 1985, pp. 93–101.
[12] A. Sgarro, *Advances in Cryptology – Eurocrypt '92*, R.A. Rueppel (Ed.), Lecture Notes in Computer Science, No. 658. Berlin: Springer Verlag, 1992, pp. 467–471.
[13] A. Sgarro, Information-theoretic bounds for authentication frauds, *Journal of Computer Security*, Vol. 2, No. 1, IOS Press, 1993, pp. 53–63.
[14] A. Sgarro, Blind coding: authentication frauds from the point of view of rate-distortion theory, preprint, presented at the Workshop on Coding and Cryptography, Paris, Jan. 1999.
[15] C. E. Shannon, Communication theory of secrecy systems, *Bell System Technical Journal*, Vol. 28, Oct. 1949, pp. 656–715.
[16] G. J. Simmons, Authentication theory/coding theory, in *Advances in Cryptology – CRYPTO 84*, G.R. Blakley and D. Chaum (Eds.), Lecture Notes in Computer Science, No. 196, Berlin: Springer Verlag, 1985, pp. 411–431.

[17] G.J. Simmons and B. Smeets, A paradoxical result in unconditionally secure authentication codes – and an explanation, in *Cryptography and Coding II*, C. Mitchell, Ed., Oxford: Clarendon, 1992, pp. 231–258.
[18] B. Smeets, Bounds on the Probability of Deception in Multiple Authentication, *IEEE Transactions of Information Theory*, Vol. 40, No. 5, 1994, pp. 1586–1591.
[19] B. Smeets, P. Vanroose, and Zhe-Xian Wan, On the construction of authentication codes with secrecy and codes which stand against spoofing attacks of order $L \geq 2$, *Advances in Cryptology – Eurocrypt '90*, I.B. Damgård, Ed., Lecture Notes in Computer Science, No. 473, Berlin: Springer Verlag, 1991, pp. 306–312.
[20] D. R. Stinson, Some constructions and bounds for authentication codes, *Journal of Cryptology*, Vol. 1, No. 1, 1988, pp. 37–51.
[21] M. Walker, Information-theoretic bounds for authentication schemes, *J. of Cryptology*, Vol 2, No. 3, 1990, pp. 131–143.