Asymptotic Information Leakage under One-Try Attacks*

Michele Boreale[1], Francesca Pampaloni[2], and Michela Paolini[2]

[1] Università di Firenze, Italy
[2] IMT, Lucca, Italy

Abstract. We study the asymptotic behaviour of (a) information leakage and (b) adversary's error probability in information hiding systems modelled as noisy channels. Specifically, we assume the attacker can make a single guess after observing n independent executions of the system, throughout which the secret information is kept fixed. We show that the asymptotic behaviour of quantities (a) and (b) can be determined in a simple way from the channel matrix. Moreover, simple and tight bounds on them as functions of n show that the convergence is exponential. We also discuss feasible methods to evaluate the rate of convergence. Our results cover both the Bayesian case, where a prior probability distribution on the secrets is assumed known to the attacker, and the maximum-likelihood case, where the attacker does not know such distribution. In the Bayesian case, we identify the distributions that maximize the leakage. We consider both the min-entropy setting studied by Smith and the additive form recently proposed by Braun et al., and show the two forms do agree asymptotically. Next, we extend these results to a more sophisticated eavesdropping scenario, where the attacker can perform a (noisy) observation at each state of the computation and the systems are modelled as hidden Markov models.

Keywords: security, quantitative information leakage, information theory, Bayes risk, hidden Markov models.

1 Introduction

In recent years there has been much interest in formal models to reason about quantitative information leakage in computing systems [9,7,3,14,1,21,22]. A general situation is that of a program, protocol or device carrying out computations that depend probabilistically on a secret piece of information, such as a password, the identity of a user or a private key. We collectively designate these as *information hiding systems*, following a terminology established in [7]. During the computation, some observable information related to the secret may be disclosed. This might happen either by design, e.g. if the output of the system is directly related to the secret (think of a password checker denying access), or for reasons depending on the implementation. In the latter case, the

* Work partially supported by the EU funded project ASCENS. Corresponding author: Michele Boreale, Università di Firenze, Dipartimento di Sistemi e Informatica, Viale Morgagni 65, I-50134 Firenze, Italy. E-mail: boreale@dsi.unifi.it

M. Hofmann (Ed.): FOSSACS 2011, LNCS 6604, pp. 396–410, 2011.

observable information may take the form of physical quantities, such as the execution time or the power consumption of the device (think of timing and power attacks on smart cards [12,13]). The observable information released by the system can be exploited by an eavesdropper to reconstruct the secret, or at least to limit the search space. This is all the more true when the eavesdropper is given the ability of observing several executions of the system, thus allowing her/him to mount some kind of statistical attack.

A simple but somehow crucial remark due to Chatzikokolakis et al. [7] is that, for the purpose of quantifying the amount of secret information that is leaked, it is useful to view an information hiding system as a *channel* in the sense of Information Theory: the inputs represent the secret information, the outputs represent the observable information and the two sets are related by a conditional probability matrix. This remark suggests a natural formalization of leakage in terms of Shannon entropy based metrics, like mutual information and capacity. In fact, by a result due to Massey [18], these quantities are strongly related to the resistance of the system against *brute-force* attacks. Specifically, Shannon entropy is related to the average number of questions of the form "is the secret equal to x?" an attacker has to ask an oracle in order to identify the secret *with certainty*. In a recent paper, Smith [21] objects that, even if the number of such questions is very high, the attacker might still have a significant chance of correct guess in just one or very few attempts. Smith demonstrates that *min-entropy* quantities, based on error probability (a.k.a. *Bayes risk*), are more adequate to express leakage in this *one-try* scenario. Whatever the considered attack scenario, brute-force or one-try, the analytic computation of leakage is in general difficult or impossible. Henceforth, a major challenge is being able to give simple and tight bounds on leakage in general, or exact expressions that exploit specific properties of a system (e.g. symmetries in the channel matrix) in some special cases.

In the present paper, we tackle these issues in a scenario of one-try attacks and system re-execution. More precisely, we assume the attacker makes his guess after observing several, say n, independent executions of the system, throughout which the secret information is kept fixed. In real-world situations, re-execution may happen either forced by the attacker (think of an adversary querying several times a smart card), or by design (think of routing paths established repeatedly between a sender and a receiver in anonymity protocols like Crowds [20]). Since the computation is probabilistic, in general the larger the number n of observed executions, the more information will be gained by the attacker. Therefore, it is important to asses the resistance of a system in this scenario.

Our goal is to describe the asymptotic behaviour of the adversary's error probability and of information leakage as n goes to ∞. We show that the asymptotic values of these quantities can be determined in a simple way from the channel matrix. Moreover, we provide simple and tight bounds on error probability and on leakage as functions of n, showing that the convergence is exponential. We also discuss feasible methods for evaluating the rate of convergence. Our results cover both the Bayesian case (MAP rule), where a prior probability distribution on the secrets is assumed known to the attacker, and the maximum-likelihood case (ML rule), where the attacker does not know such distribution. In the Bayesian case, we identify the distributions that maximize leakage.

We consider both the min-entropy leakage studied by Smith [21] and the additive form recently proposed by Braun et al. [6], and show the two forms do agree asymptotically.

We next consider a more sophisticated scenario, where computations of the system may take several steps to terminate, or even not terminate at all. In any case, to each state of the computation there corresponds one (in general, noisy) observation on the part of the attacker. Hence, to each computation there corresponds a sequential *trace* of observations. The attacker may collect multiple such traces, corresponding to multiple independent executions of the system. Like in the simpler scenario, the secret is kept fixed throughout these executions. This set up is well suited to describe situations where the attacker collects information from different sources at different times, like in a coalition of different local eavesdroppers. An instance of this situation in the context of an anonymous routing application will be examined. We formalize this scenario in terms of discrete-time *Hidden Markov Models* [19] and then show that the results established for the simpler scenario carry over to the new one.

Throughout the paper, we illustrate our results with a few examples: the modular exponentiation algorithm used in public-key cryptography, the Crowds anonymity protocol, and onion routing protocols [11] in a network with a fixed topology. Additional examples are provided in [4].

Related work. The last few years have seen a flourishing of research on quantitative models of information leakage. In the context of language-based security, Clark et al. [9] first motivated the use of mutual information to quantify information leakage in a setting of imperative programs. Boreale [3] extended this study to the setting of process calculi, and introduced a notion of rate of leakage. In both cases, the considered systems do not exhibit probabilistic behaviour. Closely related to ours is the work by Chatzikokolakis, Palamidessi and their collaborators. [7] examines information leakage mainly from the point of view of Shannon entropy and capacity, but also contains results on asymptotic error probability, showing that, independently from the input distribution, the ML rule approximates the MAP rule. [8] studies error probability mainly relative to one observation ($n = 1$), but also offers a lower-bound in the case of repeated observations [8, Proposition 7.4]. This lower-bound is generalized by our results. Compositional methods based on process algebras are discussed in [5]; there, the average ML error probability is characterized in terms of MAP error probability under a uniform distribution of inputs. [6] introduces the notion of additive leakage and compares it to the min-entropy based leakage considered by Smith [21], but again in the case of a single observation.

A model of "unknown-message" attacks is considered by Backes and Köpf in [1]. This model is basically equivalent to the information hiding systems considered in [7,8,6] and in the present paper. Backes e Köpf too consider a scenario of repeated independent observations, but from the point of view of Shannon entropy, rather than of error probability. They rely on the information-theoretic method of types to determine the asymptotic behaviour of the considered quantities, as we do in the present paper. An application of their setting to the modular exponentiation algorithm is the subject of [15], where the effect of *bucketing* on security of RSA is examined (see Section 5). This study has recently been extended to the case of one-try attacks by Köpf and Smith in [16]. Earlier, Köpf and Basin had considered a scenario of adaptive

chosen-message attacks [14]. They offer an algorithm to compute conditional Shannon entropy in this setting, but not a study of its asymptotic behaviour, which seems very difficult to characterize.

In the context of side-channel cryptanalysis, Standaert et al. propose a framework to reason on side-channel correlation attacks [22]. Both a Shannon entropy based metric and a security metric are considered. This model does not directly compare to ours, since, as we will discuss in Section 5, correlation attacks are inherently known-message – that is, they presuppose the explicit or implicit knowledge of the plaintext on the part of the attacker.

Structure of the paper. The rest of the paper is organized as follows. Section 2 establishes some notations and terminology. Section 3 introduces the model and the quantities that are the object of our study. Section 4 discusses the main results about error probability and leakage. Section 5 illustrates these results with a few examples. Section 6 presents the extension to hidden Markov models. Section 7 contains some concluding remarks. Proofs not reported in this short version for lack of space can be found in the full version [4].

2 Notations and Preliminary Notions

Let $\mathcal{A}$ be a finite nonempty set. A probability distribution on a $\mathcal{A}$ is a function $p : \mathcal{A} \to [0, 1]$ such that $\sum_{a\in\mathcal{A}} p(a) = 1$. For any $A \subseteq \mathcal{A}$ we let $p(A)$ denote $\sum_{a\in A} p(a)$. Given $n \geq 0$, we let $p^n : \mathcal{A}^n \to [0, 1]$ be the n-th extension of p, defined as $p^n((a_1, \ldots, a_n)) \stackrel{\triangle}{=} \Pi_{i=1}^n p(a_i)$; this is in turn a probability distribution on $\mathcal{A}^n$. For $n = 0$, we set $p^0(\epsilon) = 1$, where ϵ denotes here the empty string. Given two distributions p and q on $\mathcal{A}$, the *Kullback-Leibler (KL) divergence* of p and q is defined as (all the log's are taken with base 2)

$$D(p||q) \stackrel{\triangle}{=} \sum_{a\in\mathcal{A}} p(a) \cdot \log \frac{p(a)}{q(a)}$$

with the proviso that $0 \cdot \log \frac{0}{q(a)} = 0$ and that $p(a) \cdot \log \frac{p(a)}{0} = +\infty$ if $p(a) > 0$. It can be shown that $D(p||q) \geq 0$, with equality if and only if $p = q$ (*Gibbs inequality*). KL-divergence can be thought of as a sort of distance between p and q, although strictly speaking it is not – it is not symmetric, nor satisfies the triangle inequality.

$\Pr(\cdot)$ will generally denote a probability measure. Given a random variable X taking values in $\mathcal{A}$, we write $X \sim p$ if X is distributed according to p, that is for each $a \in \mathcal{A}$, $\Pr(X = a) = p(a)$.

3 Probability of Error, Leakage, Indistinguishability

An *information hiding system* is a quadruple $\mathcal{H} = (\mathcal{S}, \mathcal{O}, p(\cdot), p(\cdot|\cdot))$, composed by a finite set of *states* $\mathcal{S} = \{s_1, ..., s_m\}$ representing the secret information, a finite set of *observables* $\mathcal{O} = \{o_1, ..., o_l\}$, an a priori probability distribution on $\mathcal{S}$, p, and a *conditional probability matrix*, $p(\cdot|\cdot) \in [0, 1]^{\mathcal{S}\times\mathcal{O}}$, where each row sums up to 1. The entry of row s and column o of this matrix will be written as $p(o|s)$, and represents the probability of observing o given that s is the (secret) input of the system. For each s, the s-th row of

the matrix is identified with the probability distribution $o \mapsto p(o|s)$ on $\mathcal{O}$, denoted by p_s. The probability distribution p on $\mathcal{S}$ and the conditional probability matrix $p(o|s)$ together induce a probability distribution r on $\mathcal{S}\times\mathcal{O}$ defined as $r(s,o) \triangleq p(s)\cdot p(o|s)$, hence a pair of random variables $(S,O) \sim r$, with S taking values in $\mathcal{S}$ and O taking values in $\mathcal{O}$. Note that $S \sim p$ and, for each s and o s.t. $p(s) > 0$, $\Pr(O = o|S = s) = p(o|s)$.

Let us now discuss the attack scenario. Given any $n \geq 0$, we assume the adversary is a passive eavesdropper that gets to know the observations corresponding to n independent executions of the system, $o^n = (o_1, ..., o_n) \in \mathcal{O}^n$, throughout which the secret state s is kept fixed. Formally, the adversary knows a random vector of observations $O^n = (O_1, ..., O_n)$ such that, for each $i = 1, ..., n$, O_i is distributed like O and the individual O_i are *conditionally independent* given S, that is, the following equality holds true for each $o^n \in \mathcal{O}^n$ and $s \in \mathcal{S}$ s.t. $p(s) > 0$

$$\Pr(O^n = (o_1, \ldots, o_n) \,|\, S = s) = \Pi_{i=1}^n p(o_i|s)\,.$$

We will often abbreviate the right-hand side of the above equation as $p(o^n|s)$. For any n, the attacker strategy is modeled by a function $g : \mathcal{O}^n \to \mathcal{S}$, called *guessing function*: this represents the single guess the attacker is allowed to make about the secret state s, after observing o^n.

Definition 1 (error probability). *Let $g : \mathcal{O}^n \to \mathcal{S}$ be a guessing function. The* probability of error after n observations *(relative to g) is given by $P_e^{(g)}(n) \triangleq 1 - P_{succ}(n)$, where $P_{succ}^{(g)}(n) \triangleq \Pr(g(O^n) = S)$.*

It is well-known (see e.g. [10]) that the optimal strategy for the adversary, that is the one that minimizes the error probability, is the Maximum A Posteriori (MAP) rule, defined below.

Definition 2 (Maximum A Posteriori rule, MAP). *A function $g : \mathcal{O}^n \to \mathcal{S}$ satisfies the* Maximum A Posteriori (MAP) criterion *if for each o^n and s, $g(o^n) = s$ implies $p(o^n|s)\,p(s) \geq p(o^n|s')p(s')$ for each s'.*

In the above definition, for $n = 0$ one has $o^n = \epsilon$, and it is convenient to stipulate that $p(\epsilon|s) = 1$: that is, with no observations at all, g selects some s maximizing the prior distribution. With this choice, $P_e^{(g)}(0)$ denotes $1 - \max_s p(s)$. It worthwhile to note that, once n and $p(s)$ are fixed, the MAP guessing function is not in general unique. It is readily checked, though, that $P_e(n)$ does *not* depend on the specific MAP function g that is chosen. Hence, throughout the paper we assume w.l.o.g. a fixed guessing function g for each given n and probability distribution $p(s)$. We shall omit the superscript $^{(g)}$, except where this might cause confusion.

Another widely used criterion is *Maximum Likelihood (ML)*, which given o^n selects a state s maximizing the likelihood $p(o^n|s)$ among all the states. ML coincides with MAP if the uniform distribution on the states is assumed. ML is practically important because it requires no knowledge of the prior distribution, which is often unknown in security applications. Our main results will also apply to the ML rule (see Remark 2 in the next section).

We now come to information leakage: this is a measure of the information leaked by the system, obtained by comparing the prior and the posterior (to the observations)

success probabilities. Indeed, two flavours of this concept naturally arise, depending on how the comparison between the two probabilities is expressed. If one uses subtraction, one gets the additive form of [6], while if one uses the ratio between them, one gets a multiplicative form. In the latter case, one could equivalently consider the difference of the log's, obtaining the *min-entropy* based definition considered by Smith [21].

Definition 3 (Additive and Multiplicative Leakage [6,21]). *The* additive *and* multiplicative leakage *after n observations are defined respectively as* $L_+(n) \stackrel{\triangle}{=} P_{succ}(n) - \max_s p(s)$ *and* $L_\times(n) \stackrel{\triangle}{=} \frac{P_{succ}(n)}{\max_s p(s)}$.

In an information hiding system, it may happen that two secret states induce the same distribution on the observables. A common example is that of a degenerate channel matrix modelling a deterministic function $\mathcal{S} \to \mathcal{O}$ with $|\mathcal{O}| < |\mathcal{S}|$. An important role in determining the fundamental security parameters of the system will be played by an indistinguishability equivalence relation over states, which is defined in the following. Recall that, for each $s \in \mathcal{S}$, we let p_s denote the probability distribution $p(\cdot|s)$ on $\mathcal{O}$.

Definition 4 (Indistinguishability). *Given* $s, s' \in \mathcal{S}$, *we let* $s \equiv s'$ *iff* $p_s = p_{s'}$.

Concretely, two states are indistinguishable iff the corresponding rows in the conditional probability matrix are the same. This intuitively says that there is no way for the adversary to tell them apart, no matter how many observations he performs. We stress that this definition does not depend on the prior distribution on states, nor on the number n of observations.

4 Bounds and Asymptotic Behaviour

Let $\mathcal{S}/\equiv$ be $\{C_1, ..., C_K\}$, the set of equivalence classes of $\equiv$. For each $i = 1, ..., K$, let

$$s_i^* \stackrel{\triangle}{=} \text{argmax}_{s \in C_i} p(s) \quad \text{and} \quad p_i^* \stackrel{\triangle}{=} p(s_i^*)\,. \tag{1}$$

We assume wlog that $p_i^* > 0$ for each $i = 1, ..., K$ (otherwise all the states in class C_i can be just discarded from the system).

Main results. We shall prove the following bounds and asymptotic behaviour for $P_e(n)$.

Theorem 1. $P_e(n)$ *converges exponentially fast to* $1 - \sum_{i=1}^K p_i^*$. *More precisely, there is* $\epsilon > 0$ *s.t.*

$$1 - \textstyle\sum_{i=1}^K p_i^* \;\le\; P_e(n) \;\le\; 1 - (\sum_{i=1}^K p_i^*) \cdot r(n)$$

where $r(n) = 1 - (n+1)^{|\mathcal{O}|} \cdot 2^{-n\epsilon}$. *Here, the lower-bound holds true for any n, while the upper-bound holds true for any* $n \ge n_0 \stackrel{\triangle}{=} \epsilon^{-1} \cdot \max_{i,j} \log(\frac{p_i^*}{p_j^*})$. *Moreover,* ϵ *only depends on the rows* $p_{s_i^*}$ $(i = 1, ..., K)$ *of the conditional probability matrix* $p(\cdot|\cdot)$.

Note that in the practically important case of the uniform distribution on states, we have $n_0 = 0$, that is the upper-bound as well holds true for any n. The theorem has a simple interpretation in terms of the attacker's strategy: after infinitely many observations, he can determine the indistinguishability class of the secret, say C_i, and then guess the most

likely state in that class, s_i^*. In order to discuss this result, we recall some terminology and a couple of preliminary results from the information-theoretic method of types [10, Ch.11]. Given $n > 0$, a sequence $o^n \in \mathcal{O}^n$ and a symbol $o \in \mathcal{O}$, let us denote by $\mathrm{n}(o, o^n)$ the number of occurrences of o inside o^n. The *type* (or empirical distribution) of o^n is the probability distribution t_{o^n} on $\mathcal{O}$ defined as: $t_{o^n}(o) \stackrel{\triangle}{=} \frac{\mathrm{n}(o,o^n)}{n}$. Let q any probability distribution on $\mathcal{O}$. A *neighborhood* of q is a subset of n-sequences of $\mathcal{O}^n$ whose empirical distribution is close to q. Formally, for each $n \geq 1$ and $\epsilon > 0$

$$U_q^{(n)}(\epsilon) \stackrel{\triangle}{=} \{o^n \in \mathcal{O}^n \mid D(t_{o^n} || q) \leq \epsilon\}\,.$$

The essence of the method of types is that (i) there is only a polynomial number of types in n, and that (ii) the probability under q of the set of n-sequences of a given type decreases exponentially with n, at a rate determined by the KL-divergence between q and that type. These considerations are made precise and exploited in the proof of the following lemma, which can be found in [10, Ch.11]. The lemma basically says that the probability that a sequence falls in a neighborhood of q of radius ϵ approaches 1 exponentially fast with n.

Lemma 1. *Let q be a probability distribution on $\mathcal{O}$. Then $q^n(U_q^{(n)}(\epsilon)) \geq 1-(n+1)^{|\mathcal{O}|}\cdot 2^{-n\epsilon}$.*

For any $s \in \mathcal{S}$, we let $A_s^{(n)} \stackrel{\triangle}{=} g^{-1}(s) \subseteq \mathcal{O}^n$ be the *acceptance region* for state s. We note that it is not restrictive to assume that g maps each o^n in one of the K representative elements $s_1^*, ..., s_K^*$ that maximize the prior: indeed, if this were not the case, it would be immediate to build out of g a new MAP function that fulfills this requirement. Thus, from now on we will assume w.l.o.g. that $A_s^{(n)} = \emptyset$ for $s \neq s_1^*, ..., s_K^*$. For the sake of notation, from now on we will denote $U_{p_{s_i^*}}^{(n)}$ as $U_i^{(n)}$ and $A_{s_i^*}^{(n)}$ as $A_i^{(n)}$, for $i = 1, ..., K$. The sets $U_i^{(n)}$ and $A_i^{(n)}$ are related by the following lemma.

Lemma 2. *There is $\epsilon > 0$, not depending on the prior probability on states, such that for each $n \geq n_0$ as defined in Theorem 1 and for each $i = 1, ..., K$, it holds that $U_i^{(n)}(\epsilon) \subseteq A_i^{(n)}$.*

We now come to the proof of the main theorem above.

Proof. (of Theorem 1). We focus equivalently on the probability of success, $P_{succ}(n)$. Under the assumptions on g explained above, we compute as follows

$$\begin{aligned} P_{succ}(n) &= \textstyle\sum_{s\in\mathcal{S}} \Pr(g(O^n) = S \,|\, S = s)p(s) = \sum_{s\in\mathcal{S}} p_s^n(A_s^{(n)})p(s) \\ &= \textstyle\sum_{i=1}^K \underbrace{p_{s_i^*}^n(A_i^{(n)})}_{\leq 1} p_i^* \qquad\qquad \leq \qquad \sum_{i=1}^K p_i^* \end{aligned}$$

which implies the lower-bound in the statement. Choose now ϵ as given by Lemma 2. Let $n \geq n_0$. Note that for $n = 0$ the upper-bound holds trivially, as $P_e(0) = 1 - \max_s p(s)$, so assume $n \geq 1$. For each $i = 1, ..., K$ we have

$$p_{s_i^*}^n(A_i^{(n)}) \;\geq\; p_{s_i^*}^n(U_i^{(n)}(\epsilon)) \;\geq\; 1 - (n+1)^{|\mathcal{O}|} \cdot 2^{-n\epsilon}$$

where the first inequality comes from Lemma 2 and second one from Lemma 1. In the end, from $P_{succ}(n) = \sum_{i=1}^K p_{s_i^*}^n(A_i^{(n)})p_i^*$, we obtain that for $n \geq n_0$

$$P_{succ}(n) \;\geq\; (\sum_{i=1}^K p_i^*) \cdot (1 - (n+1)^{|\mathcal{O}|} \cdot 2^{-n\epsilon})$$

which implies the upper-bound in the statement.

Remark 1. In the expression for $r(n)$, the term $(n+1)^{|O|}$ is a rather crude upper bound on the number of types of n-sequences. It is possible to replace this term with the expression $\binom{n+|O|-1}{|O|-1}$, which is less easy to manipulate analytically, but gives the exact number of types, hence a more accurate upper bound on $P_e(n)$.

The following results show that, asymptotically, the security of the systems is tightly connected to the number of its indistinguishability classes – and in the case of uniform prior distribution *only* depends on this number.

Corollary 1. *If the a priori distribution on $\mathcal{S}$ is uniform, then $P_e(n)$ converges exponentially fast to* $1 - \frac{K}{|\mathcal{S}|}$.

Remark 2 (on the ML *rule).* [5] shows that the probability of error under the ML rule, *averaged* on all distributions, coincides with the probability of error under the MAP rule and the uniform distribution. From Corollary 1 we therefore deduce that the average ML error converges exponentially fast to the value $1 - \frac{K}{|\mathcal{S}|}$ as $n \to \infty$.

We discuss now some consequences of the above results on information leakage. Recall that for $i = 1, ..., K$, we call s_i^* a representative of the indistinguishability class C_i that maximizes the prior distribution $p(s)$ in the class C_i, and let $p_i^* = p(s_i^*)$. Assume w.l.o.g. that $p_1^* = \max_s p(s)$. In what follows, we denote by p_{max} the distribution on $\mathcal{S}$ defined by: $p_{max}(s) = \frac{1}{K}$ if $s \in \{s_1^*, ..., s_K^*\}$ and $p_{max}(s) = 0$ otherwise.

Corollary 2. *1. $L_+(n)$ converges exponentially fast to $\sum_{i=2}^K p_i^*$. This value is maximized by the prior distribution p_{max}, which yields the limit value* $1 - \frac{1}{K}$.

2. $L_\times(n)$ converges exponentially fast to $\frac{\sum_{i=1}^K p_i^*}{p_1^*}$. This value is maximized by the prior distribution p_{max}, which yields the limit value K.

Remark 3. A consequence of Corollary 2(2) is that, in the case of uniform distribution on states, the multiplicative leakage coincides with the number of equivalence classes K. This generalizes a result of [21] for deterministic systems.

In [6] additive and multiplicative leakages are compared in the case of a single observation ($n = 1$). It turns out that, when comparing two systems, the two forms of leakage are in agreement, in the sense that they individuate the same maximum-leaking system w.r.t. a fixed prior distribution on inputs. However, [6] also shows that the two forms disagree as to the distribution on inputs that maximizes leakage w.r.t. a fixed system. This is shown to be the uniform distribution in the case of multiplicative leakage, and a function that uniformly distributes the probability on the set of "corner points" in the case of additive leakage (see [6] for details). Here, we have shown that, despite this difference, additive and multiplicative leakage do agree on the maximizing distribution asymptotically.

Rate of convergence. The quantity ϵ in the statement of Theorem 1 determines how fast the error probability approaches its limit value. Let us call *achievable* any $\epsilon > 0$ for which the upper bound in Theorem 1 holds true for any $n \geq n_0$. The following result gives sufficient and practical conditions for achievability. Let us stress that the achievable rates given by this proposition do not depend on the prior distribution, but only on

the relation $\equiv$, and specifically on the minimum norm 1 distance between equivalence classes: the larger this distance, the higher the achievable rates. This result is essentially a re-elaboration on [10, Lemma 11.6.1].

Proposition 1. *Let* $\delta \stackrel{\triangle}{=} \min_{s_i \not\equiv s_j} ||p_{s_i} - p_{s_j}||_1$. *Then any rate* ϵ *satisfying* $0 < \epsilon < \frac{\delta^2}{16 \ln 2}$ *is achievable. Moreover, if* $p_1^* = p_2^* = \cdots = p_K^*$, *the second inequality can be weakened to* $\epsilon < \frac{\delta^2}{8 \ln 2}$.

The above result prompts the following question. Suppose one somehow ignores the rows of $p(\cdot|\cdot)$ that are close together with each other, and only consider rows that are far from each other: is it then possible to achieve a higher rate of convergence ϵ? The answer is expected to be *yes*, although ignoring some rows might lead to a possibly higher asymptotic error probability. In other word, it should be possible to trade off accuracy in guessing with rate of convergence. This is the content of the next proposition.

Proposition 2. *Let* $\emptyset \neq \mathcal{S}_0 \subseteq \{s_1^*, ..., s_K^*\}$. *Then there is* $\epsilon > 0$ *only depending on the rows* p_s, $s \in \mathcal{S}_0$, *of* $p(\cdot|\cdot)$, *such that for each* $n \geq n_0 \stackrel{\triangle}{=} \epsilon^{-1} \max_{s_i^*, s_j^* \in \mathcal{S}_0} \log(\frac{p_i^*}{p_j^*})$, *it holds true that*

$$P_e(n) \leq 1 - (\sum_{s_j^* \in \mathcal{S}_0} p_j^*) \cdot r(n) \quad \textit{with} \quad r(n) = 1 - (n+1)^{|O|} \cdot 2^{-n\epsilon}\,.$$

These concepts are demonstrated in the following example.

An example. Let $\mathcal{S} = \{s_1, s_2, s_3, s_4\}$ and $O = \{o_1, o_2, o_3\}$. The prior probability distribution on $\mathcal{S}$ is defined by: $p(s_1) = p(s_3) = \frac{1}{2} - 10^{-9}$ and $p(s_2) = p(s_4) = 10^{-9}$. The conditional probability matrix is defined in the table on the right.

	o_1	o_2	o_3
s_1	$\frac{1}{2}$	0	$\frac{1}{2}$
s_2	$\frac{1}{2}$	0	$\frac{1}{2}$
s_3	0	$\frac{1}{2}$	$\frac{1}{2}$
s_4	0	$\frac{1}{2} - 10^{-5}$	$\frac{1}{2} + 10^{-5}$

Note that $s_1 \equiv s_2$. Applying Theorem 1, we find that, for n sufficiently large, $1 - E \leq P_e(n) \leq 1 - E \cdot r(n)$, where $E = 1 - 10^{-9}$ and $r(n) = 1 - (n+1)^3 \cdot 2^{-n\epsilon}$. Applying Proposition 1, we find that any rate $\epsilon < 3.6067 \times 10^{-11}$ is achievable. Thus the convergence to the value $1 - E = 10^{-9}$ is very slow. One wonders if there is some value $1 - E'$ that is only slightly higher than $1 - E$, but that can be reached much faster. This is indeed the case. Observe that the rows s_3 and s_4 are very close with each other in norm-1 distance: $||p_{s_3} - p_{s_4}||_1 = 2 \times 10^{-5}$. We can discard s_4, which has a very small probability, and apply Proposition 2 with $\mathcal{S}_0 = \{s_1, s_3\}$ to get $P_e(n) \leq 1 - E' \cdot r'(n)$, where $E' = \frac{1}{2} - 10^{-9} + \frac{1}{2} - 10^{-9} = 1 - 2 \times 10^{-9}$ and $r'(n) = 1 - (n+1)^3 \cdot 2^{-n\epsilon'}$. The rate ϵ' can be computed by applying the second part of Proposition 1, as $p(s_1) = p(s_3)$. By doing so, we get that any $\epsilon' < 0.18034$ is achievable. This implies that the value $1 - E'$ is approached much faster as n grows. For instance, already after $n = 350$ observations we get that $(1 - E')/P_e(n) > 0.99$.

5 Examples

Timing leaks and blinding in modular exponentiation. In the '90's, P. Kocher showed that RSA and other public-key crypto-systems are subject to side-channel attacks exploiting information leaked by implementations of the modular exponentiation algorithm,

such as execution time [12] and/or power consumption [13]. Many of these attacks are based on the assumption that the attacker can observe repeated independent execution of the system, throughout which the exponent – the secret key – is kept fixed. Here, we concentrate on timing attacks. *Blinding* [12] was early proposed as a countermeasure to thwart such attacks. The essence of blinding is that exponentiation is performed on a random message unknown to the attacker, rather than on the original message (to be decrypted or digitally signed) known to the attacker. This appears to be sufficient to thwart Kocher's attack, which is of chosen-ciphertext type.

Köpf and Dürmuth [15] have recently quantified the degree of protection provided by blinding when it is enhanced by *bucketing*, a technique by which the algorithm's execution times are adjusted so as to always fall in one of few predefined values. Köpf and Smith have extended this result to the case of one-try attacks and min-entropy [16]. Below, we refine Köpf and Smith's analysis, under a reasonable assumption on the functioning of the algorithm that will be described shortly. We consider an implementation of the modular exponentiation algorithm with blinding, but *no* bucketing. To such an implementation, there corresponds an information hiding system where: $\mathcal{S} = \mathcal{K} = \{0,1\}^N$ is the set of private keys, i.e. the possible exponents of the algorithm, over which we assume a uniform distribution[1]; $\mathcal{O} = \{t_1, t_2, ...\}$ is the set of possible execution times; $p(t|k)$ is the probability that, depending on the deciphered message, the execution of the algorithm takes times t given that the private key is k. To be more specific about the last point, we assume an underlying set of messages $\mathcal{M}$, with a known prior distribution $p_M(m)$, and a function time : $\mathcal{M}\times\mathcal{S} \to \mathcal{O}$ that yields the duration of the execution of the algorithm when its argument is a given pair (m,k). Then the entries of the probability matrix $p(t|k)$ can be defined thus

$$p(t|k) \stackrel{\triangle}{=} \sum_{m\in\mathcal{M}:\mathrm{time}(m,k)=t} p_M(m)\,.$$

Now, modular exponentiation functions in such a way that at the i-th iteration ($0 \le i < N$), either a squaring or *both* a squaring and a multiply are performed, depending on whether the i-th bit of the exponent is 0 or 1. Given this functioning, it seems reasonable to assume that, for each m, *the execution time only depends on the number of '1' digits in* k. In other words, we assume that whenever k and k' have the same Hamming weight, $\mathrm{time}(m,k) = \mathrm{time}(m,k')$, for any m. From this assumption and the definition of $p(t|k)$, it follows that whenever k and k' have the same Hamming weight then $p(\cdot|k) = p(\cdot|k')$. So, in the system there are *at most* as many $\equiv$-classes as Hamming weights, that is $N+1$. The results in Section 4 then allow us to conclude that for any n

$$P_e(n) \ge 1 - \tfrac{N+1}{2^N}\,.$$

For any practical size of the key, say $N = 1024$, this value is ≈ 1. Accordingly, additive and multiplicative leakage satisfy, asymptotically,

$$L_+ \le \tfrac{N}{2^N} \quad \text{and} \quad L_\times \le N+1\,.$$

For any practical size of the key, say $N = 1024$, these upper bounds yield negligible values: $L_+ \approx 0$ and $L_\times \le 1025$. In the latter case, taking the log we obtain that no more than $\log(1025) = 10.001$ bits of min-entropy are leaked, out of 1024. In conclusion,

[1] In the case of RSA, a negligible fraction of the exponents is ruled out by virtue of number theoretic requirements, so the resulting distribution is not exactly uniform on $\mathcal{S}$. This fact does not substantially affect the significance of our analysis.

under the further assumption on the behaviour of modular exponentiation we made above, blinding alone appears to provide satisfactory guarantees of security against one-try attacks.

Protocol re-execution in Crowds. The Crowds protocol [20] is designed for protecting the identity of the senders of messages in a network where some of the nodes may be corrupted, that is, under the control of an attacker. Omitting a few details, the functioning of the protocol can be described quite simply: the sender first forwards the message to a node of the network chosen at random; at any time, any node holding the message can decide whether to (a) forward in turn the message to another node chosen at random, or (b) submit it to the final destination. The choice between (a) and (b) is made randomly, with alternative (a) being assigned probability p_f (forwarding probability) and alternative (b) probability $1 - p_f$. The rationale here is that, even if a corrupted node C receives the message from a node N (in the Crowds terminology, C *detects* N), C, hence the attacker, cannot decide whether N is the original sender or just a forwarder. In fact, given that N is detected, the probability of N being the true sender is only slightly higher than that of any other node being the true sender. So the attacker is left with a good deal of uncertainty as to the sender's identity. Reiter and Rubin have showed that, depending on p_f, on the fraction of corrupted nodes in the network and on a few other conditions, Crowds offers very good guarantees of anonymity (see [20]).

Chatzikokolakis et al. have recently analyzed Crowds from the point of view of information hiding systems and one-try attacks [7,8]. In their modelling, $\mathcal{S} = \{s_1, ..., s_m\}$ is the set of possible senders (honest nodes), while $\mathcal{O} = \{d_1, ..., d_m\}$ is the set of observables. Here each d_i has the meaning that node s_i has been detected by some corrupted node. The conditional probability matrix is given by

$$p(d_j|s_i) \triangleq \Pr(s_j \text{is detected} \mid s_i \text{is the true sender and some honest node has been detected})$$

(see [20] for details of the actual computation of these quantities). An example of such a system with $m = 20$ users, borrowed from [8], is given in the table below.

	d_1	d_2	$\cdots$	d_{20}
s_1	0.468	0.028	$\cdots$	0.028
s_2	0.028	0.468	$\cdots$	0.028
$\vdots$				$\vdots$
s_{20}	0.028	0.028	$\cdots$	0.468

The interesting case for us is that of re-execution, in which the protocol is executed several times, either forced by the attacker himself (e.g. by having corrupted nodes suppress messages) or by some external factor, and the sender is kept fixed through the various executions. This implies the attacker collects a sequence of observations $o^n = (o_1, ..., o_n) \in \mathcal{O}^n$, for some n. The repeated executions are assumed to be independent, hence we are precisely in the setting considered in this paper. This case is also considered in [8], which gives lower bounds for the error probability holding for any n. Our results in Section 4 generalize those in [8] by providing both lower- and upper- bounds converging exponentially fast to the asymptotic error probability. As an example, for the system in the table above, we have $P_e(n) \to 0$, independently of the prior distribution on the senders. An achievable convergence rate, estimated with the method of Proposition 1, is $\epsilon \approx 0.13965$. This implies that already after observing $n = 1000$ re-executions the probability of error is, using the refined bound given in Remark 1, < 0.01.

It is worth to stress that protocol re-execution is normally prevented in Crowds for the very reason that it decreases anonymity, although it may be necessary in some cases. See the discussion on static vs. dynamic paths in [20].

6 Sequential Observations and Hidden Markov Models

We consider in this section an attack scenario where to each state of the computation there corresponds one observation on the part of the attacker. Hence, to each computation of the system there corresponds a sequential *trace* of observations. Discrete-time Hidden Markov Models [19] provide a convenient setting to formally model such systems, which we may designate as *sequential* information hiding system.

Definitions. Let $\mathcal{S}$ and $\mathcal{O}$ be finite sets of states and observations, respectively. A (discrete-time, homogeneous) *Hidden Markov Model (*HMM*)* with states in $\mathcal{S}$ and observations in $\mathcal{O}$ is a a pair of random processes $\langle (S_i)_{i\geq 1}\,,\, (O_i)_{i\geq 1}\rangle$, such that, for each $t \geq 1$

- S_t and O_t are random variables taking values in $\mathcal{S}$ and $\mathcal{O}$, respectively; and,
- the following equalities hold true (whenever the involved conditional probabilities are defined):

$$\Pr(S_{t+1} = s_{t+1} | S_t = s_t, O_t = o_t, ..., S_1 = s_1, O_1 = o_1) = \Pr(S_{t+1} = s_{t+1} | S_t = s_t) \quad (2)$$

$$\Pr(O_t = o_t | S_t = s_t, S_{t-1} = s_{t-1}, O_{t-1} = o_{t-1}, ..., S_1 = s_1, O_1 = o_1) = \Pr(O_t = o_t | S_t = s_t) \quad (3)$$

  Moreover, the value of the above probabilities does not depend on the index t, but only on s_t, s_{t+1} and o_t.

Equation (2) says that the state at time $t + 1$ only depends on the state at time t, that is $(S_i)_{i\geq 1}$ forms a Markov chain. Equation (3) says that the observation at time t only depends on the state at time t. A consequence of this equation is that the state at time $t + 1$ is independent from the observation at time t, given the state at time t, that is

$$\Pr(O_t = o_t, S_{t+1} = s_{t+1} | S_t = s_t) \;=\; \Pr(O_t = o_t | S_t = s_t) \cdot \Pr(S_{t+1} = s_{t+1} | S_t = s_t)\,. \quad (4)$$

Assume now $\mathcal{S} = \{s_1, ..., s_m\}$ and $\mathcal{O} = \{o_1, ..., o_l\}$. A finite-state HMM on $\mathcal{S}$ and $\mathcal{O}$ is completely specified by, hence can be identified with, a triple (π, F, G) such that:

- $\pi \in \mathbb{R}^{1\times m}$ is a row-vector representing the prior distribution on $\mathcal{S}$, that is $\pi(i) = p(S_1 = s_i)$ for each $1 \leq i \leq m$;
- $F \in \mathbb{R}^{m\times m}$ is a matrix such that $F(i, j)$ is the probability of transition from s_i to s_j, for $1 \leq i, j \leq m$;
- $G \in \mathbb{R}^{m\times l}$ is a matrix such that $G(i, j)$ is the probability of observing o_j at state s_i, for $1 \leq i \leq m$ and $1 \leq j \leq l$.

In our scenario, a Bayesian attacker targets the first state of the computation, that is the value of S_1. We are interested in analyzing the attacker's probability of error after observing n traces of length t, corresponding to n conditionally independent executions of the system up to and including time t, as both n and t go to $+\infty$. This we define in the following. Let σ range over the set of observation traces, that is $\mathcal{O}^*$. For any $\sigma = o_1 \cdots o_t$ $(t \geq 0)$ and $s \in S$, define

$$p(\sigma \mid s) \stackrel{\triangle}{=} \Pr(O_1 = o_1, O_2 = o_2, ..., O_t = o_t \mid S_1 = s)$$

with the proviso that $p(\epsilon \mid s) \triangleq 1$. We note that for any fixed $t \geq 0$ and $s \in \mathcal{S}$, $p(\sigma|s)$ defines a probability distribution as σ ranges over O^t, the set of traces of length t, or *t-traces*. In other words, for *any fixed t*, we have an information hiding system in the sense of Section 3, with $\mathcal{S}$ as a set of states, O^t as a set of observables, a conditional probability matrix $p(\sigma|s)$ ($s \in \mathcal{S}, \sigma \in O^t$) and π as a prior distribution on states. Call $\mathcal{H}^{(t)}$ this system. We have the following error probabilities of interest ($t \geq 0$):

$$P_e^{(t)}(n) \triangleq \text{ probability of error after } n \text{ observations (of } t\text{-traces) in } \mathcal{H}^{(t)} \tag{5}$$

$$P_e^{(t)} \triangleq \lim_{n\to\infty} P_e^{(t)}(n) \tag{6}$$

$$P_e \triangleq \lim_{t\to\infty} P_e^{(t)} . \tag{7}$$

We will show in the next paragraph that the above two limits exist and are easy to compute. Correspondingly, we have the information leakage quantities of interest (here $P_{succ} = 1 - P_e$): $L_+^{(t)}(n) \triangleq P_{succ}^{(t)}(n) - \max_s \pi(s)$ $L_+^{(t)} \triangleq P_{succ}^{(t)} - \max_s \pi(s)$ $L_+ \triangleq P_{succ} - \max_s \pi(s)$. Multiplicative leakages are defined similarly.

Results. That the limit (6) exists is an immediate consequence of Theorem 1 applied to $\mathcal{H}^{(t)}$. Indeed, let us denote by $\equiv^{(t)}$ the indistinguishability relation on states for $\mathcal{H}^{(t)}$, that is, explicitly $s \equiv^{(t)} s'$ iff for each $\sigma \in O^t : p(\sigma|s) = p(\sigma|s')$ .. Let $C_1^{(t)}, ..., C_{K_t}^{(t)}$ be the equivalence classes of $\equiv^{(t)}$ and let $p_i^{*(t)} \triangleq \max_{s\in C_i^{(t)}} \pi(s)$. Then we have by Theorem 1 that $P_e^{(t)} = 1 - \sum_{i=1}^{K_t} p_i^{*(t)}$. Note that, for any fixed t, Corollary 2 carries over to $\mathcal{H}^{(t)}$. We now consider the case $t \to \infty$. We introduce the following fundamental relation.

Definition 5 (Indistinguishability for HMM). *The* indistinguishability relation *on a* HMM *is defined as* $\equiv \triangleq \bigcap_{t\geq 0} \equiv^{(t)}$. *Equivalently,* $s \equiv s'$ *iff for every* $\sigma \in O^*$, $p(\sigma|s) = p(\sigma|s')$.

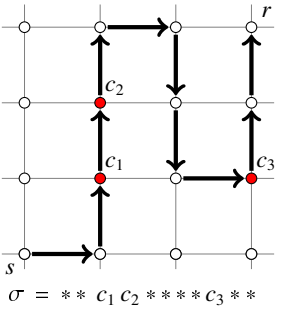
It is immediate to check that $\equiv$ is an equivalence relation. Let $C_1, ..., C_K$ be its equivalence classes and let $p_i^* \triangleq \max_{s\in C_i} \pi(s)$, for $i = 1, ..., K$.

Proposition 3. *The limit (7) is given by* $P_e = 1 - \sum_{i=1}^{K} p_i^*$.

The actual computation of P_e, and of the corresponding information leakage quantities, is therefore reduced to the computation of $\equiv$. Below, we show that this computation can indeed be performed quite efficiently. We do so by using some elementary linear algebra. Let us introduce some additional notation. We define the transition matrices $M_{o_k} \in \mathbb{R}^{m\times m}$, for any $o_k \in O$, as follows: $M_{o_k}(i, j) \triangleq \Pr(S_{t+1} = s_j, O_t = o_k|S_t = s_i) = F(i, j) \cdot G(i, k)$, where the last equality is justified by equation (4). For any $\sigma = o_1 \cdots o_t$, we let M_σ denote $M_{o_1} \times \cdots \times M_{o_t}$. Finally, we let $e_i \in \mathbb{R}^{1\times m}$ denote the row vector with 1 in the i-th position and 0 elsewhere and let $e \triangleq \sum_{i=1}^{m} e_i$ denote the everywhere 1 vector. We say a row vector v is *orthogonal* to a set of column vectors U, written $v \perp U$, if $vu = 0$ for each $u \in U$.

Theorem 2. *Let B be a basis of the (finite-dimensional) sub-space of* $\mathbb{R}^{m\times 1}$ *spanned by* $\bigcup_{\sigma\in O^*}\{M_\sigma e^T\}$. *For* $s_i, s_j \in \mathcal{S}$, $s_i \equiv s_j$ *iff* $(e_i - e_j) \perp B$.

A basis B of span($\bigcup_\sigma \{M_\sigma e^T\}$) can be expressed as $B = \{M_\sigma e^T \mid \sigma \in \mathcal{F}\}$ for a suitable finite, prefix-closed $\mathcal{F} \subseteq O^*$. More precisely, B can be computed by a procedure that starts with the set $B := \{e^T\}$ and iteratively updates B by joining in the vectors $M_{o\cdot\sigma} e^T = M_o \cdot (M_\sigma e^T)$, with $M_\sigma e^T \in B$ and $o \in O$, that are linearly independent from the vectors already present in B, until no other vector can be joined in. This procedure must terminate in a number of steps $\leq m$. The set of strings $\mathcal{F}$ can be computed alongside with B. In the full version of the paper we also discuss a method to compute the rate of convergence to P_e.



An example: hiding routing information. We outline a simple anonymity protocol in the vein of onion routing [11]. The protocol aims at protecting the identity of the sender *and* of the receiver of a transaction in a network where some of the nodes are compromised by local eavesdroppers. The routing paths are established randomly. The local eavesdroppers have limited observation capabilities and, perhaps because of encryption, can only tell whether, at any discrete time step, the compromised node is holding a message in the target transaction, or not. Assume the topology of the network is specified by a nonempty graph $\mathcal{G} = (V, E)$. For each node $v \in V$, we let $N(v)$ denote the set of neighbours of v, that is the set of nodes u for which an arc $\{v, u\}$ in E exists; $N(v)$ is always assumed nonempty. Let $C \subseteq V$ represent a subset of corrupted nodes. We let $\mathcal{S} \triangleq V \times V$ be the set of states of the system and $O \triangleq C \cup \{*\}$ be the set of observables. State $(s, r) \in \mathcal{S}$ means the message is hold by s and that r is the final receiver. Observation $c \in C$ means that the message is presently hold by the node c, while $*$ means no observation other than the elapse of a discrete time unit. What the attacker can observe are therefore traces σ like in the picture above. The exact definition of the transition and observation matrices F and G of the HMM, as well as the outcome of several experiments conducted with this simple model, are reported in the full version of the paper [4].

7 Conclusion and Further Work

We have characterized the asymptotic behaviour of error probability, and information leakage in terms of indistinguishability in a scenario of one-try attacks after repeated independent, noisy observations. We have first examined the case in which each execution gives rise to a single observation, then extended our results to the case where each state traversed during an execution induces one observation.

In the future, we would like to systematically characterize achievable rates of convergence given an error probability threshold, thus generalizing Proposition 1. It would also be natural to generalize the present one-try scenario to the case of k-tries attack, for $k \geq 2$. Experiments and simulations with realistic anonymity protocols may be useful to asses at a practical level the theoretical results of our study. For example, we believe that HMM's are relevant to security in peer-to-peer overlays. We would also like to clarify the relationship of our model with the notion of probabilistic *opacity* [2], and with the huge amount of work existing on *covert channels* (see e.g. [17] and references therein).

References

1. Backes, M., Köpf, B.: Formally Bounding the Side-Channel Leakage in Unknown-Message Attacks. In: Jajodia, S., Lopez, J. (eds.) ESORICS 2008. LNCS, vol. 5283, pp. 517–532. Springer, Heidelberg (2008)
2. Bérard, B., Mullins, J., Sassolas, M.: Quantifying Opacity. In: Proc. of QEST 2010, pp. 263–272. IEEE Society, Los Alamitos (2010)
3. Boreale, M.: Quantifying information leakage in process calculi. Information and Computation 207(6), 699–725 (2009)
4. Boreale, M., Pampaloni, F., Paolini, M.: Asymptotic information leakage under one-try attacks, Full version of the present paper `http://rap.dsi.unifi.it/~boreale/Asympt.pdf`
5. Braun, C., Chatzikokolakis, K., Palamidessi, C.: Compositional Methods for Information-Hiding. In: Amadio, R.M. (ed.) FOSSACS 2008. LNCS, vol. 4962, pp. 443–457. Springer, Heidelberg (2008)
6. Braun, C., Chatzikokolakis, K., Palamidessi, C.: Quantitative Notions of Leakage for One-try Attacks. In: Proc. of MFPS 2009. Electr. Notes Theor. Comput. Sci, vol. 249, pp. 75–91 (2009)
7. Chatzikokolakis, K., Palamidessi, C., Panangaden, P.: Anonymity protocols as noisy channels. Information and Computation 206(2-4), 378–401 (2008)
8. Chatzikokolakis, K., Palamidessi, C., Panangaden, P.: On the Bayes risk in information-hiding protocols. Journal of Computer Security 16(5), 531–571 (2008)
9. Clark, D., Hunt, S., Malacaria, P.: Quantitative Analysis of the Leakage of Confidential Data. Electr. Notes Theor. Comput. Sci. 59(3) (2001)
10. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2/e. John Wiley & Sons, Chichester (2006)
11. Goldschlag, D.M., Reed, M.G., Syverson, P.F.: Anonymous Connections and Onion Routing. IEEE Journal on Selected Areas in Communication, Special Issue on Copyright and Privacy Protection (1998)
12. Kocher, P.C.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996)
13. Kocher, P.C., Jaffe, J., Jun, B.: Differential Power Analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
14. Köpf, B., Basin, D.A.: An information-theoretic model for adaptive side-channel attacks. In: ACM Conference on Computer and Communications Security 2007, pp. 286–296 (2007)
15. Köpf, B., Dürmuth, M.: A Provably Secure and Efficient Countermeasure against Timing Attacks. In: CSF 2009, pp. 324–335 (2009)
16. Köpf, B., Smith, G.: Vulnerability Bounds and Leakage Resilience of Blinded Cryptography under Timing Attacks. In: CSF 2010, pp. 44–56 (2010)
17. Mantel, H., Sudbrock, H.: Information-Theoretic Modeling and Analysis of Interrupt-Related Covert Channels. In: Degano, P., Guttman, J., Martinelli, F. (eds.) FAST 2008. LNCS, vol. 5491, pp. 67–81. Springer, Heidelberg (2009)
18. Massey, J.L.: Guessing and Entropy. In: Proc. 1994 IEEE Symposium on Information Theory (ISIT 1994), vol. 204 (1994)
19. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proc. of the IEEE 77(2), 257–286 (1989)
20. Reiter, M.K., Rubin, A.D.: Crowds: Anonymity for Web Transactions. ACM Trans. Inf. Syst. Secur. 1(1), 66–92 (1998)
21. Smith, G.: On the Foundations of Quantitative Information Flow. In: de Alfaro, L. (ed.) FOSSACS 2009. LNCS, vol. 5504, pp. 288–302. Springer, Heidelberg (2009)
22. Standaert, F.-X., Malkin, T.G., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 443–461. Springer, Heidelberg (2009)