Exponential Bounds for Information Leakage in Unknown-Message Side-Channel Attacks

Daniel Z. Zanger, Ph.D.*

SRI International, Arlington, VA, USA

July 8, 2010

Abstract: In [1], the authors introduced an important new information theoretic numerical measure for assessing a system's resistance to unknown-message side-channel attacks and computed a formula for the limit of the numerical values defined by this measure as the number of side-channel observations tends to infinity. Here, we present corresponding quantitative (exponential) bounds that yield an actual rate-of-convergence for this limit, something not given in [1]. Such rate-of-convergence results can potentially be used to significantly strengthen the utility of the limit formula of [1] as a tool to reduce computational complexity difficulties associated with calculating the side-channel attack resistance measure presented there. In addition, our arguments here show how the arguments used in [1] to prove the limit formula can be substantially simplified.

1 Introduction

In a side-channel attack, the attacker attempts to circumvent the security of cryptographic algorithms by exploiting information inadvertently exposed within the concrete context of their actual, real-world implementations. Such possibly indirect yet nonetheless potentially quite

* Email: danielzanger@gmail.com, Mailing address: 3133 Connecticut Ave., NW #521, Washington, DC 20008, USA.

relevant information might include that revealed by running time characteristics [2], cache behavior [4], electromagnetic signals [5], or other electronic or physical properties or emanations. In an unknown-message side-channel attack scenario, the attacker cannot see or control the input that is decrypted (or encrypted) by the system. An example of such an unknown message attack is a timing attack against systems employing advanced countermeasures such as message blinding.

In [1], the authors propose an important new information theoretic measure for quantifying the resistance of systems to unknown-message side channel attacks. Their measure, denoted $\Lambda = \Lambda(n)$ (see (6) below), is a quantitative benchmark for measuring the reduction in the attacker's uncertainty (i.e., the corresponding information gain) regarding the attacked system's designated cryptographic key after n side-channel observations. In addition, in their Theorem 1 in [1], they compute an explicit formula for the limit of $\Lambda(n)$ as $n \to \infty$ (restated here as (7) below).

Our own Theorem 1 in §3 below not only recovers the qualitative result $\lim_{n\to\infty}\Lambda(n)$ appearing in [1] but very significantly strengthens and improves on it by establishing quantitative (exponential) bounds on the corresponding rate-of-convergence (see (8) below), something not addressed in [1]. Additionally, in Example 1 in §3, we show how our numerical estimate in (8) can actually be applied in a concrete, if somewhat abstract, setting. Also, we believe that an additional important contribution of our work here is that the proof of our Theorem 1 involves what we believe to be a much-simplified argument relative to that used in [1] to establish (7). In [1] (see the Full Version of that paper), a relatively involved argument is invoked using somewhat elaborate and perhaps esoteric results from information theory to compute $\lim_{n\to\infty}\Lambda(n)$. Here, our simplified argument is entirely self-contained aside from an appeal to the well-known, classical Hoeffding inequality. In fact, to simply identify the limit $\lim_{n\to\infty}\Lambda(n)$ as is done in [1] without also proving our own quantitative bound in (8), we would have actually only needed, using an argument exactly analogous to the one we use to prove our Theorem 1, just the traditional, standard Law of Large Numbers (in its weak form) from probability theory.

In [1], the authors also introduce a precise algorithm for facilitating the computation of

$\Lambda(n)$. However, as they themselves point out, the time complexity of the algorithm is exponential in n, rendering computation for large values of n infeasible. In fact, they promote their determination of the limit $\lim_{n\to\infty}\Lambda(n)$ in their Theorem 1 as a way to remedy this computational complexity problem since it allows one to compute limits for the resistance to side channel attacks without being confronted with the exponential increase in n. We believe that the quantitative bound we present here in the statement of our own Theorem 1, by giving the associated rate-of-convergence, has the potential to dramatically strengthen this computational complexity remediation strategy since it shows, of course, exactly how large n must be for $\lim_{n\to\infty}\Lambda(n)$ to be a good approximation of $\Lambda(n)$. In fact, the bound shows that the exponential increase in the time complexity of the algorithm given in [1] for computing $\Lambda(n)$ is actually offset by an exponential decrease in the numerical error resulting from using $\lim_{n\to\infty}\Lambda(n)$ as an approximation for it.

The rest of this article is organized as follows. In the next section, we give the necessary background on unknown message side channel attacks and relevant information theoretic concepts. In the final section, we state and prove our fundamental new result, Theorem 1, and present an example, Example 1, that directly applies it, as discussed above.

2 Some Preliminaries on Unknown Message Attacks and Information Theory

Our framework here is based on that of [1]. Let K be a finite set of keys, M a finite set of messages, and D an arbitrary set. We consider systems that compute functions of type $F : K \times M \to D$, and we assume the attacker can make physical observations about the implementation $\mathcal{I}_F$ of F that are associated with the computation of $F(k, m)$. We assume that the attacker can make one observation per invocation of the function F and that no measurement errors occur. A side-channel is a function $f_{\mathcal{I}_F} : K \times M \to O$, where O denotes the set of possible observations. Since K and M are finite we can assume, without loss of generality, that O too is finite, and we write $O = \{o_1, ..., o_I\}$ for some positive integer I. We assume the attacker has full knowledge regarding the implementation $\mathcal{I}_F$, i.e., $f = f_{\mathcal{I}_F}$ is known to the attacker.

In a side-channel attack, the attacker collects side-channel observations $f(k, m_1), ..., f(k, m_n)$ for ascertaining the key k or narrowing down its possible values. Such an attack is *unknown-message* if the attacker cannot observe or choose the messages $m_i \in M$. (By way of contrast, an attack is *known-message* if the attacker can observe but cannot influence the choice of $m_i \in M$, and an attack is *chosen-message* if the attacker can choose $m_i \in M$.)

Now let $p_K : K \to \boldsymbol{R}$ and $p_M : M \to \boldsymbol{R}$ be probability distributions on the sets of keys K and messages M respectively. These immediately give rise to random variables $\mathcal{K}$ and $\mathcal{M}$ modeling the respective random choices of keys and messages. We assume that $p_{\mathcal{K}} = p_K$ and $p_{\mathcal{M}} = p_M$ are known to the attacker. For positive integers n, let the random variable $\boldsymbol{\mathcal{O}}_n : K \times M^n \to O^n$ be defined by $\boldsymbol{\mathcal{O}}_n(k, m_1, ..., m_n) = (f(k, m_1), ..., f(k, m_n))$, where $p_{KM^n}(k, m_1, ..., m_n) = p_K(k)p_M(m_1)...p_M(m_n)$ is the probability distribution on $K \times M^n$. Also, we write $\mathcal{O} = \boldsymbol{\mathcal{O}}_1$.

Now naturally for each key $k \in K$, the random variable $\mathcal{O}$ gives rise to a probability distribution $p_{\mathcal{O}|\mathcal{K}=k}$ on O. For $k, k' \in K$, define $k \equiv k'$ if and only if $p_{\mathcal{O}|\mathcal{K}=k} = p_{\mathcal{O}|\mathcal{K}=k'}$. Then, $\equiv$ is an equivalence relation on K, and $V \stackrel{\text{def}}{=} K/\equiv$ denotes the (finite) set of equivalence classes and $|V|$ its cardinality. The random variable $\mathcal{V} : K \to V$ defined by $\mathcal{V}(k) = [k]$ maps every key to its equivalence class with respect to $\equiv$.

Now for any random variable $\mathcal{X}$ assuming values in the set X, we of course have the information theoretic (Shannon) entropy

$$H(\mathcal{X}) = -\sum_{x \in X} p_{\mathcal{X}}(x) \log_2(p_{\mathcal{X}}(x)). \tag{1}$$

If $\mathcal{Y}$ is another random variable defined on the same probability space taking values in some set Y, then, for $y \in Y$, we denote by $H(\mathcal{X}|\mathcal{Y} = y)$ the entropy of $\mathcal{X}$ with respect to the distribution $p_{\mathcal{X}|\mathcal{Y}=y}$. As in [1], we then define the conditional entropy as

$$H(\mathcal{X}|\mathcal{Y}) = \sum_{y \in Y} p_{\mathcal{Y}}(y) H(\mathcal{X}|\mathcal{Y} = y). \tag{2}$$

We have the identity

$$H(\mathcal{XY}) = H(\mathcal{Y}) + H(\mathcal{X}|\mathcal{Y}) \tag{3}$$

where $\mathcal{XY}$ is the random variable defined via $\mathcal{XY}(k) = (\mathcal{X}(k), \mathcal{Y}(k))$.

Similar to the definition in [1], we also define the *type* $\boldsymbol{t}_{\boldsymbol{O}_n}$ of a sequence $\boldsymbol{o}_n \in O^n$ to be the vector of respective numbers of occurrences of each element $o_i \in O$ in the sequence $\boldsymbol{o}_n$. Thus we have $\boldsymbol{t}_{\boldsymbol{O}_n} = (t_1, ..., t_I)$ where, for all $i = 1, ..., I$,

$$t_i = |\{o_{nj} | o_{nj} = o_i, j = 1, ..., n\}|, \tag{4}$$

with $\boldsymbol{o}_n = (o_{n1}, ..., o_{nn})$ and $|\cdot|$ denoting the cardinality of the enclosed set. Finally, in the next section, we will also require a measure of how far apart, in a sense, the distributions of the form $p_{\mathcal{O}|\mathcal{V}=B}, B \in V$ are. So, for this, set

$$\epsilon_V = \min_{B_1, B_2 \in V} \max_{i=1,...,I} \frac{|p_{\mathcal{O}|\mathcal{V}=B_1}(o_i) - p_{\mathcal{O}|\mathcal{V}=B_2}(o_i)|}{3}. \tag{5}$$

From the definition of $V = K/\equiv$, it is clear that $\epsilon_V > 0$.

3 An Exponential Bound on Side-Channel Observation Information Gain

In [1], the authors introduce, employing the notation in (2) above,

$$\Lambda(n) \stackrel{\text{def}}{=} H(\mathcal{K}|\boldsymbol{\mathcal{O}}_n), \; n \text{ a positive integer}, \tag{6}$$

as a practical measure for potentially quantifying the resistance of systems against unknown message side channel attacks. In addition, they prove that

$$\lim_{n\to\infty} H(\mathcal{K}|\boldsymbol{\mathcal{O}}_n) = H(\mathcal{K}|\mathcal{V}). \tag{7}$$

Our own Theorem 1 below recovers the qualitative result (7) above that appears in [1] but very significantly strengthens and improves on it by establishing a quantitative (exponential) estimate for the difference $H(\mathcal{K}|\boldsymbol{\mathcal{O}}_n) - H(\mathcal{K}|\mathcal{V})$ (see (8) below). This yields a bound on the corresponding convergence rate, something not given in [1]. Moreover, we believe that an additional important contribution of our work here is that the proof of our Theorem 1 involves a much-simplified argument relative to that used in [1] to establish (7), as discussed in more detail in the Introduction.

So, without further ado, our numerical estimate on information gain after repeated unknown message side-channel measurements is given in the following theorem.

Theorem 1. *We have*

$$0 \leq H(\mathcal{K}|\boldsymbol{\mathcal{O}}_n) - H(\mathcal{K}|\mathcal{V}) \leq 2|V|^2 I \exp(-2n\epsilon_V^2). \tag{8}$$

We illustrate the result of Theorem 1 with the following example.

Example 1. The framework for this example is based on that of Example 3, again from [1]. Let $K = \{0,1\}^j$ and $M = \{1, ..., j\}$, j being any positive integer, and set $O = \{0,1\}$. We assume the probability distributions $p_K : K \to \boldsymbol{R}$ and $p_M : M \to \boldsymbol{R}$ to be uniform distributions. Consider the function $f : K \times M \to O$ defined via $f(k,m) = k_m$, where $k = (k_1, ..., k_j)$. For computing $H(\mathcal{K}|\mathcal{V})$, notice that for $k, k' \in K$, $p_{\mathcal{O}|\mathcal{K}=k} = p_{\mathcal{O}|\mathcal{K}=k'}$ if and only if the numbers of 1-bits in k and k' are equal, i.e., if k and k' have the same Hamming weight. The number of j-bit values with Hamming weight h is given by $\binom{j}{h} = \frac{j!}{(j-h)!h!}$. Since, in general, $H(\mathcal{K}|\mathcal{V}) = \frac{1}{|K|}\sum_{B \in V} |B| \log_2 |B|$ for uniformly distributed keys k, the statement of our Theorem 1 implies that

$$\frac{1}{2^j}\sum_{h=0}^{j}\binom{j}{h}\log_2\binom{j}{h} \leq H(\mathcal{K}|\boldsymbol{\mathcal{O}}_n) \leq \frac{1}{2^j}\sum_{h=0}^{j}\binom{j}{h}\log_2\binom{j}{h} + 4j^2\exp\left(-\frac{2n}{9j^2}\right). \tag{9}$$

Lemma 1 and its proof which follow just below essentially appear within the proof of Lemma 1 of [1]. The argument establishing (10) in essence involves only careful manipulation of relevant definitions.

Lemma 1. *We have*

$$H(\mathcal{K}|\boldsymbol{\mathcal{O}}_n) - H(\mathcal{K}|\mathcal{V}) = H(\mathcal{V}|\boldsymbol{\mathcal{O}}_n). \tag{10}$$

Proof. Note that $p_{\mathcal{O}|\mathcal{K}=k} = p_{\mathcal{O}|\mathcal{V}=[k]}$, so that $H(\boldsymbol{\mathcal{O}}_n|\mathcal{K}=k) = H(\boldsymbol{\mathcal{O}}_n|\mathcal{V}=[k])$ and

$$\begin{aligned} H(\boldsymbol{\mathcal{O}}_n|\mathcal{K}) &= \sum_{k\in K} p_{\mathcal{K}}(k) H(\boldsymbol{\mathcal{O}}_n|\mathcal{K}=k) = \sum_{k\in V}\sum_{k\in B} p_{\mathcal{K}}(k) H(\boldsymbol{\mathcal{O}}_n|\mathcal{K}=k) \\ &= \sum_{B\in V} p_{\mathcal{V}}(B) H(\boldsymbol{\mathcal{O}}_n|\mathcal{V}=B) = H(\boldsymbol{\mathcal{O}}_n|\mathcal{V}). \end{aligned} \tag{11}$$

Now observe that $H(\mathcal{K}\mathcal{V}) = H(\mathcal{K})$, since $\mathcal{V}$ is determined by $\mathcal{K}$. Hence, invoking (3), $H(\mathcal{K}|\boldsymbol{\mathcal{O}}_n) - H(\mathcal{K}|\mathcal{V}) = (H(\boldsymbol{\mathcal{O}}_n|\mathcal{K}) + H(\mathcal{K}) - H(\boldsymbol{\mathcal{O}}_n)) - (H(\mathcal{K}\mathcal{V}) - H(\mathcal{V})) = H(\boldsymbol{\mathcal{O}}_n|\mathcal{K}) + H(\mathcal{V}) - H(\boldsymbol{\mathcal{O}}_n)$. Using (11), it now follows that $H(\boldsymbol{\mathcal{O}}_n|\mathcal{K}) + H(\mathcal{V}) = H(\boldsymbol{\mathcal{O}}_n|\mathcal{V}) + H(\mathcal{V}) = H(\boldsymbol{\mathcal{O}}_n\mathcal{V})$. As $H(\boldsymbol{\mathcal{O}}_n\mathcal{V}) - H(\boldsymbol{\mathcal{O}}_n) = H(\mathcal{V}|\boldsymbol{\mathcal{O}}_n)$, (10) immediately follows.

Proof of Theorem 1. By Lemma 1 above, it is sufficient to show that $H(\mathcal{V}|\boldsymbol{\mathcal{O}}_n) \leq 2|V|^2 I\exp(-2n\epsilon_V^2)$. Note first that from the definitions (1) and (2), it follows that

$$H(\mathcal{V}|\boldsymbol{\mathcal{O}}_n) = \sum_{\boldsymbol{o}_n \in O^n} p_{\boldsymbol{\mathcal{O}}_n}(\boldsymbol{o}_n)(-\sum_{B\in V} p_{\mathcal{V}|\boldsymbol{\mathcal{O}}_n=\boldsymbol{o}_n}(B)\log_2(p_{\mathcal{V}|\boldsymbol{\mathcal{O}}_n=\boldsymbol{o}_n}(B))). \tag{12}$$

Now, for each $i = 1, ..., I$, let $p_{ni} = p_{ni}(o_i, \boldsymbol{o}_n) = \frac{t_i}{n}$, where t_i is as in (4), and, for each $B \in V = K/\equiv$ and $\boldsymbol{o}_n \in O^n$ where n is any positive integer, write

$$m(\boldsymbol{o}_n, B) \stackrel{\text{def}}{=} \max_{i=1,...,I}|p_{ni}(o_i, \boldsymbol{o}_n) - p_{\mathcal{O}|\mathcal{V}=B}(o_i)|. \tag{13}$$

Note that, by iterated application of the classical Hoeffding's inequality (see, for example, Theorem 2.3(a) in [3]), it follows that, for any $B \in V$ and $\epsilon > 0$,

$$p(m(\boldsymbol{\mathcal{O}}_n, B) > \epsilon) \leq 2I\exp(-2n\epsilon^2). \tag{14}$$

In fact, (14) directly follows from the Hoeffding inequality because, for each $i \in I$, we can define indicator random variables that are set to 1 if $o_{nj} = o_i$ where $\boldsymbol{o}_n = (o_{n1}, ..., o_{nn})$ but 0 otherwise, so that the expectation in the statement of Hoeffding's inequality is simply the probability $p_{\mathcal{O}|\mathcal{V}=B}(o_i)$. In any case, notice as well that

$$\begin{aligned} p(m(\boldsymbol{\mathcal{O}}_n, \mathcal{V}) \leq \epsilon) &= 1 - p(m(\boldsymbol{\mathcal{O}}_n, \mathcal{V}) > \epsilon) \\ &= 1 - \sum_{B\in V} p(\{\mathcal{V} = B\} \cap \{m(\boldsymbol{\mathcal{O}}_n, B) > \epsilon\}) \\ &\geq 1 - 2I|V|\exp(-2n\epsilon^2). \end{aligned} \tag{15}$$

Now consider that the definition of $V = K/\equiv$ clearly implies that, if ϵ is small enough, in particular if $\epsilon = \epsilon_V$ with ϵ_V defined as in (5) above, then, for any $\boldsymbol{o}_n \in O^n$, any $B \in V$ satisfying $m(\boldsymbol{o}_n, B) \leq \epsilon$ (if such $B \in V$ does exist) must be unique, and we denote this corresponding unique equivalence class B via $B_{\boldsymbol{o}_n}$. Hence, for such small ϵ, the random variable $\mathcal{V}$ is fully determined by the random variable $\boldsymbol{\mathcal{O}}_n$ under $m(\cdot,\cdot)$ and in fact we have $\mathcal{V} = B_{\boldsymbol{\mathcal{O}}_n}$. So, for $\epsilon = \epsilon_V$, we can rewrite the conclusion of (15) as

$$p(m(\boldsymbol{\mathcal{O}}_n, B_{\boldsymbol{\mathcal{O}}_n}) \leq \epsilon_V) \geq 1 - 2I|V|\exp(-2n\epsilon_V^2). \tag{16}$$

It therefore follows that, with probability $1 - 2I|V|\exp(-2n\epsilon_V^2)$ over all $\boldsymbol{o}_n \in O^n$,

$$p_{\mathcal{V}|\boldsymbol{\mathcal{O}}_n=\boldsymbol{o}_n}(B_{\boldsymbol{o}_n}) = 1 \text{ and } p_{\mathcal{V}|\boldsymbol{\mathcal{O}}_n=\boldsymbol{o}_n}(B) = 0 \text{ for } B \in V, B \neq B_{\boldsymbol{o}_n}. \tag{17}$$

Thus, with probability $1 - 2I|V|\exp(-2n\epsilon_V^2)$ over all $\boldsymbol{o}_n \in O^n$,

$$-\sum_{B \in V} p_{\mathcal{V}|\boldsymbol{\mathcal{O}}_n=\boldsymbol{o}_n}(B)\log_2(p_{\mathcal{V}|\boldsymbol{\mathcal{O}}_n=\boldsymbol{o}_n}(B)) = 0. \tag{18}$$

Hence, since for $x \in (0,1]$, $|x\log_2(x)| \leq 1$, it follows directly from (12) along with (18) that

$$0 \leq H(\mathcal{V}|\boldsymbol{\mathcal{O}}_n) \leq 2|V|^2 I\exp(-2n\epsilon_V^2). \tag{19}$$

References

[1] Backes, M. and B. Kopf(2008):Formally Bounding the Side-Channel Leakage in Unknown-Message Attacks. 13th European Symposium On Research In Computer Security (ESORICS, 2008). (Full version of the paper available at the Cryptology ePrint Archive, Report 2008/162,2008.)

[2] Boneh, D. and D. Brumley(2003):Remote Timing Attacks are Practical. In *Proc. USENIX Security Symposium '03.*

[3] McDiarmid, C.(1998):"Concentration" in *Probabilistic Methods for Algorithmic Discrete Mathematics,* New York: Springer, 195-228.

[4] Osvik, D. A.,A. Shamir, and E. Tromer(2006):Cache Attacks and Countermeasures: the Case of AES. In*Proc. CT-RSA '06,* LNCS 3860, pages 1-20, Springer.

[5] Quisquater J.-J. and D. Samyde(2001): Electromagnetic Analysis(EMA): Measures and Countermeasures for Smart Cards. In *Proc. E-smart '01,* LNCS 2140, pages 200-210, Springer.