Exploitation as an Inference Problem

David Cock NICTA and University of New South Wales
davec@cse.unsw.edu.au

ABSTRACT

In this position paper, we suggest that an adversary seeking to exploit a side channel should be viewed as performing inference under uncertainty. We propose a set of vulnerability measures that incorporate both *observational* effort and *computational* effort. By deriving Boolean satisfiability as a special case of the marginalization problem, we justify that the measure is capable of capturing the complexity of the underlying deterministic decision problem. In the limit of unbounded computation the measure reduces to the efficiency (in number of observations) of naïve Bayesian analysis. We further hypothesize that in the limit of unbounded observations, the measure reduces to the complexity of the decision problem.

1. SIDE CHANNELS

The aim of this research is to construct a bound on the vulnerability of a secure system, X, due to *side channels* [BB03, Ber05]. We define a side channel as a list of observations $o = (o_1, \ldots, o_n) : o_i \in O$, derived stochastically from the system's current state s. The state consists of three components: $\sigma \in \Sigma$ — a hidden, sensitive state (e.g. an encryption key); $v \in V$ — a visible, or shared state (e.g. the wall clock time, or a known plaintext) and $u \in U$ — the 'uninteresting' remainder of the hidden state. We assume that the interesting hidden state σ is fixed, while v and u may vary. We further make the pessimistic assumption that v is under the control of the attacker.

First, we consider the problem as purely one of inference with unlimited resources. From the point of view of an attacker, A, the system behaves as a stochastic function $X : \Sigma \times V \times U \to O$. The attacker's task is to compute the partial inverse — $O^n \times V^n \to \Sigma$. Assume that we accept some maximum acceptable likelihood of compromise: ϵ. We will consider the system vulnerable if, with probability at least ϵ, a hypothetical attacker A can correctly deduce σ given o and v. We would expect that the likelihood of compromise should in general vary with the number of observations, and so propose our first pair of vulnerability measures: $P_c(n)$ — the probability of compromise after n observations and $N_c(\epsilon)$ — the smallest number of observations needed to compromise the system with $P > \epsilon$.

By defining our measure directly in terms of our desired security property (likelihood of compromise), we can capture some existing notions of security. In particular, the property $\forall n.\ P_c(n+1) = P_c(n)$ implies no leakage of sensitive information and $\forall n.\ P_c(n) = |\Sigma|^{-1}$ to maximal security (or *noninterference* [Rus92] between σ and v, o). The important distinction between these properties is the incorporation of *prior information*. An attacker has a greater likelihood of early success if he knows that σ is not uniformly distributed:

$$P_c(0) > |\Sigma|^{-1} \iff \exists \sigma.\ P_{prior}(\sigma) > |\Sigma|^{-1}$$

Both of the above conditions are unattainable in practice, for the same reason that simple noninterference fails as a security measure for practical cryptographic systems: the secret input to an encryption is often completely deducible from its visible inputs and outputs. A straightforward example is RSA [RSA78]: a public key, together with the exponent, uniquely determine the corresponding private key. Of course, computing the private key is (believed) hard, and modulo this assumption the system is secure.

If we want to avoid excessive pessimism, we need to consider this computational effort. We thus extend our basic measures to $P_c(n, w)$ — the probability of compromise given n observations and at no more than w effort; $N_c(\epsilon, w)$ — the minimum number of observations needed to compromise with $P \geq \epsilon$ and no more than w effort and $W_c(\epsilon, n)$ — the minimum effort needed to compromise with $P \geq \epsilon$ given n observations. These additional measures are related to our principle measure, P_c as follows:

$$N_c(\epsilon, w) = \min_n :\ P_c(n, w) \geq \epsilon$$
$$W_c(\epsilon, n) = \min_w :\ P_c(n, w) \geq \epsilon$$

2. COMPLEXITY OF INFERENCE

The complexity of an attack is thus measured in two dimensions: the number of observations, N_c, and the computational effort, W_c, required for a compromise. To give a proof of security, we need to find a lower bound for at least one of N_c or W_c. We consider N_c first.

The attacker is making a decision under uncertainty, and after n observations must select a $\sigma' \in \Sigma$ as his best guess at the secret σ. The optimal attack strategy (with respect to our measure P_c) must select $\sigma' = \sigma$ with greatest expected probability (with expectation taken over $\Sigma \times O$). One algorithm satisfying this is Bayesian inference with MAP (Maximum A Posteriori) assignment [Mac04]. In brief, this relies on having access to the distribution $P(o_i|\sigma)$ and calculating $P(\sigma|o)$ by repeated application of Bayes' rule:

$$P(\sigma|o_{1\ldots i}) = \frac{P(o_i|\sigma)P(\sigma|o_{1\ldots i-1})}{P(o_i|o_{1\ldots i-1})}$$

The guess, σ', is then chosen to maximize $P(\sigma'|o)$. We can thus in principle bound $N_c(\epsilon, w)$ by $N_c(\epsilon, \infty)$:

$$N_c(\epsilon, \infty) = \min_n : P(\sigma'_n = \sigma) \geq \epsilon$$
$$\sigma'_n = \operatorname*{argmax}_\sigma P(\sigma|o_{1\ldots n})$$

AISec'11, October 21, 2011, Chicago, Illinois, USA.
ACM 978-1-4503-1003-1/11/10.

Bounding W_c requires us to reason about the *minimum* complexity of any algorithm solving the decision problem with sufficient accuracy. We first note that the stochastic function:

$$X : \Sigma \times V \times U \to O$$

is equivalent to the distribution:

$$P(o_i|\sigma, v_i, u_i)$$

and that solving for the partial inverse:

$$O^n \times V^n \to \Sigma$$

is simply the marginalization problem:

$$P(o, \sigma, v) = \sum_u P(o, \sigma, v, u)$$

i.e. we must marginalize over the hidden state. To solve for σ' given o and v, we simply maximize the resulting distribution:

$$\sigma' = \operatorname*{argmax}_\sigma P(o, \sigma, v)$$

Unfortunately for our attacker (but fortunately for us), both marginalization and maximization are hard problems in general. The question is: how hard, and in what circumstances? We note that any problem expressible as the satisfying assignment of a finite Boolean propositional formula is encodable using the following construction (for each translation, α is a fresh name representing a new Boolean variable):

$$\begin{aligned} T(Q) &= \Delta(Q \text{ is true}) \\ T(\neg a) &= \Delta(\alpha = \neg a)T(a) \\ T(a \wedge b) &= \Delta(\alpha = a \wedge b)T(a)T(b) \\ T(a \vee b) &= \Delta(\alpha = a \vee b)T(a)T(b) \\ \Delta Q &= \begin{cases} 1 & : Q \\ 0 & : \neg Q \end{cases} \end{aligned}$$

'Marginalizing' over some subset of $\{Q\}$ amounts to counting the satisfying assignments to that subset of propositions (strictly, the *proportion* of satisfying assignments), and maximization to finding one (if the maximum is 0, there are no such assignments). These problems are respectively #P- and NP-hard [Val79, Coo71]. So there certainly exist difficult marginalization problems, but how can we establish that for a given side channel, *any* associated marginalization problem is hard?

One approach is to show that the associated marginalization problem must encode a known (or presumed) hard problem. For example, if the side-channel observations are known to depend only on the result of an RSA encryption, then marginalization over the secret key must be *at least* as hard as solving the RSA problem.

It would be convenient to leverage existing hardness results, by reducing the stochastic inference problem to a known deterministic problem. To that end, we propose the following hypothesis:

HYPOTHESIS 1. *As n approaches infinity, then all stochastic elements of observations disappear: we are left with the underlying discrete inference problem. Specifically, $W_c(\epsilon, \infty)$ is a safe lower bound for $W_c(\epsilon, n)$, and corresponds to the difficulty of the underlying decision problem.*

If true, this would allow us to derive a lower bound on complexity by appealing to existing results.

3. RESEARCH DIRECTIONS

This approach to assessing vulnerability suggests a number of interesting questions, which we intend to address in future work:

- Is Hypothesis 1 true?
- Is it possible to automatically identify a control-flow-based side channel, and build a minimal corresponding inference model? Doing so would provide a powerful new tool for vulnerability analysis.
- Are countermeasures based on increasing the complexity of inference practical, and if so how should they be implemented?
- Can this style of probabilistic reasoning be rigorously integrated with cryptographic techniques, to optimally combine information gleaned from cryptanalysis (algorithmic weaknesses) and side channels (implementation weaknesses)?

4. REFERENCES

[BB03] David Brumley and Dan Boneh. Remote timing attacks are practical. In *Proceedings of the 12th USENIX Security Symposium*, Berkeley, CA, USA, 2003. USENIX.

[Ber05] Daniel J. Bernstein. Cache-timing attacks on AES. 2005.

[Coo71] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, STOC '71, pages 151–158, New York, NY, USA, 1971. ACM.

[Mac04] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge university press, 2004.

[RSA78] R.L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21:120–126, 1978.

[Rus92] John Rushby. Noninterference, transitivity, and channel-control security policies. Technical report, SRI International, December 1992.

[Val79] Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, 1979.