The MIT Center for Digital Business

A research and education initiative at the MIT Sloan School of Management

Functions that Preserve Privacy but Permit Analysis of Text

Paper 246

May 2009

Mark C. Reynolds
Marshall Van Alstyne
Sinan Aral

For more information,
please visit our website at **http://digital.mit.edu**
or contact the Center directly at **digital@mit.edu or 617-253-7054**

MITSloan

Functions that Preserve Privacy but Permit Analysis of Text

Mark C. Reynolds
Boston University
Department of Computer Science
markreyn@cs.bu.edu

Marshall Van Alstyne
Boston University
School of Management
mva@bu.edu

Sinan Aral
NYU
Information, Operations and Management Sciences Department
sinan@stern.nyu.edu

Abstract

Cryptographically strong functions can be used to preserve privacy of text content. For example, one way functions have been construed as random functions on their inputs. Given this, it is reasonable to ask if a one way function can still preserve some "property" of its inputs. Specifically, is it possible to perform some measurement on the image on a one way function that is correlated with the same measurement on the pre-image of the function? In this paper we show that this is indeed possible. If the measurement function "throws away" enough entropy it will still be possible to perform the correlated measurement. Thus it is possible to analyze properties of text while still providing security and privacy for its content.

1. Introduction

A function $h(x)$ is said to be one way [1] on its preimage if the following two properties hold: (a) There exists a polynomial time algorithm that computes $h(x)$ from x, except perhaps for a negligible subset of the preimage; (b) For every probabilistic polynomial time (PPT) algorithm A that acts on the image of $h(x)$, given y in that image, A can compute a preimage of y with at most negligible probability, namely

$$Pr[h(A(y)) = y] < \epsilon(\|A(y)\|)$$

where $\|x\|$ denotes the length of x. Condition (a) is the condition of being easy to compute, while condition (b) is the condition of being hard to invert. Note that while one way functions are widely believed to exist, there is no example of a function that has been proved to be one way.

A special category of one way functions of particular interest are hash functions. A hash function is a function that has two additional properties. First, a hash function is a compression function. If D and R denote the preimage and image of a hash function H, respectively, then $\|D\| < \|R\|$. In addition, a hash function also has the property of collision resistance. A function is collision resistant if it is hard to find x and x' (with $x \neq x'$) such that $(H(x) = H(x')$. Here "hardness" is interpreted to mean that no PPT algorithm exists.

Let us now introduce several new definitions. Suppose f is a function defined on the preimage D of a hash function H, such that $f\colon D \rightarrow R_f$. (We assume without loss of generality that the preimage of f and the preimage of H coincide.) Suppose further that g is a function defined on the image R_H of the function H, such that $g\colon R \rightarrow R_g$. A function p is said to be a *property* of the hash function H if $p(f(x))$ is approximately equal to $g(H(x))$ for a large fraction of D (these notions will be made precise momentarily). Note that it is not required that the preimage of p correspond exactly to R_f, nor that its image exactly correspond to R_g. It suffices if the preimage and image of p are supersets of R_f and R_g, respectively.

The notion of approximate equality means that the inequality

$$|(p(f(x)) - g(H(x))| < \epsilon$$

holds, except with negligible probability. This inequality can best be viewed in light of the following approximately commutative diagram:

$$\begin{array}{ccc} D & \xrightarrow{f} & R_f \\ {\scriptstyle H}\downarrow & & \downarrow{\scriptstyle p} \\ R_H & \xrightarrow{g} & R_g \end{array}$$

If E is an ensemble of values in the preimage of two functions f_1 and f_2 then we define the normalized distance between f_1 and f_2 over E as

$$nd(f_1, f_2, E) = \sum_{x \in E} |(f_1(x) - f_2(x)|/\|E\|$$

This quantity, of course, depends on the choice of the ensemble E. If we compute the maximum value of the normalized distance over all ensembles of size $S = \|E\|$ we can write $nd(f_1, f_2, S) = \max_{S=\|E\|} nd(f_1, f_2, E)$. If for any ϵ we can find a lower bound S_{min} such that for all $S > S_{min}$ the inequality $nd(f_1, f_2, S) < \epsilon$ holds, then we say that f_1 is an *approximant* of f_2 (and, symmetrically, that f_2 is an approximant of f_1).

2. p-Security for One Way Functions

If $p(f)$ is an approximant of $g(H)$ then we call the triplet $\langle f, g, p\rangle$ a *property* of the hash function H. We will call f and g the *measurement functions* of this property, and p the *correlation function* of the property. In an informal sense, f is measuring something about the preimage of H, while g is measuring something about the image of H. The correlation function p provides a method for (approximately) computing the measurement on the image, given a corresponding measurement on the preimage.

A hash function (or, more generally a one way function) H is said to be *p-secure* with respect to the property $\langle f, g, p\rangle$ if no PPT algorithm A_p can invert H. It is implied that the algorithm A_p is allowed to use f, g, and p, so that p-security of H in the presence of a property is a stronger statement than the statement that the function H is one way. Before stating the main theorem regarding p-security, it is worthwhile to consider two informative examples. Suppose that $f(x) = \|x\|$, $g(x) = \|x\|$ and $p(x) = kx$ with $k = \|R\|/\|D\|$. If we assume (without loss of generality, see [2]) that H is length-regular, then $g(H(x)) = p(f(x))$ identically. In this case the property $\langle f, g, p\rangle$ does not provide any new information about H and the non-existence of A_p is immediate from the fact that H is one way. As a second example, suppose that p is PPT invertible by an algorithm A_1 and that f is PPT invertible by an algorithm A_2. Then for any value x which satisfies $g(H(x)) = p(f(x))$ we can recover a preimage of y by computing $x = A_2(A_1(g(y)))$. These two examples should clearly illustrate that the degree to which p-security for a property $\langle f, g, p\rangle$ of H holds depends critically on f, g and p. It would be highly desirable if we could develop a characterization for $\langle f, g, p\rangle$ such that there was a proof of security, e.g. that the existence of correlated measurements on the preimage and image did not have any adverse impact on the privacy provided by the application of the one way function H. In fact we will prove a theorem that says, informally, that if the function f loses enough entropy, privacy will be preserved even for arbitrary g and p.

Theorem 1 *If H is a one way function, $\langle f, g, p\rangle$ is a property of H, and f is one way, then H is p-secure with respect to $\langle f, g, p\rangle$ for any g and p.*

Proof. Consider the most unfavorable case, that is consider the case where p is polynomial-time invertible and g is arbitrary. Write $y = H(x)$ and $z = g(y) = g(H(x))$. Then by assumption we can compute $w = p^{-1}(z)$ in polynomial time. If there existed a probabilistic polynomial time algorithm A that inverted H, we could (except for a negligible subset) compute x in polynomial time, given $H(x)$. But then x would be a preimage of w with respect to the function f, which is a contradiction of the assumption that f is a one way function. Structurally, this results bears some similarity to those of [3].

3. Value Functions

Before proceeding further, it is necessary to formulate precisely what is meant by the "information content" of the (preimage or image) of a function. We formulate this notion in terms of a set of questions that can be asked (e.g., predicates) and the accuracy of the responses (the fidelity). Formally, suppose that E is an ensemble of values on some domain (not necessarily the preimage D of the one way function). A *value function* V is a collection of predicates $\{b_i : 0 \leq i < B\}$. An *evaluation* of the value function over an ensemble E is the set of predicate outputs $\{b_i : e \in E, 0 \leq i < B\}$ together with a set of probabilities $\{F_i : 0 \leq i < B\}$, known as the *fidelity* of each predicate. Informally, the fidelity of a particular predicate evaluation is the probability that the output is correct. Note that we assume that each predicate has access to the entire ensemble of values, so that we can define the preimage of the value function V to be the ensemble E. Given the set of fidelities calculated over the entire set of predicates, we can define the total fidelity F of V to be the product of the individual fidelities F_i. By abuse of notation we can then write $V(E) = F$ indicating that V measures the information content of some aspect of the ensemble E with aggregate fidelity F.

Given these definitions, we can next ask for the dependency of F on E. Informally, we would like it to be the case that if we increase the population of E this somehow leads to more information, and that as a result the aggregate fidelity (at least) does not decrease. Since V may not be a deterministic function, it may be too strong a requirement that V be strictly ordered on the cardinality of E, however. Therefore, we define a less restrictive notion, that of a weak

ordering. We say that a value function is *weakly ordered* if two conditions are satisfied. The first condition is the *threshold condition*, which states that

$$\|E\| \geq \|E_{min}\|$$

namely that we have at least some minimum population. The second condition is the *weak ordering* condition, which states that if $F_1 \leq 1$ and $F_2 \leq 1$, and $r = F_2/F_1 > 1$, then there exists some r' such that

$$\|E'\| > r'\|E\| \Rightarrow V(E') > rV(E)$$

for all ensembles E and E'. Roughly speaking, the more accurate we want our answer to be, the larger a population we must have. Henceforth in this paper we will always assume that any value function V is weakly increasing.

We observe immediately that if V is weakly increasing, and therefore weakly ordered, we can take the contrapositive of the statement above; namely, if $r < 1$ (a decrease in information content) then we can find some $r' < 1$ such that

$$V(E') < rV(E) \Rightarrow \|E'\| < r'\|E\|$$

so long as the threshold condition is also satisfied.

We can now use these definitions to define what we mean by the "loss of information" due to the application of an approximation function as described in the first section. Let f by a function acting on the preimage of a one way function H, let g by a function on the image of that one way function, and let p be a correlation function of f and g. Let E_f be an ensemble of values in the image of f and E_g be an ensemble of values in the image of g. Let V be a value function that acts on both E_f and E_g. Then we define the *differential value* of the two ensembles as:

$$DV(E_f, E_g) = V(E_g)/V(E_f)$$

We assume throughout that both E_f and E_g satisfy the threshold condition and also that the pathological case $V(E) = 0$ is not realized. In a heuristic sense, DV is a measure of the amount of information lost due to applying the one way function; this is the baseline against which we will now measure further losses due to the approximation process. Given a correlation function p, rather than looking at the output obtained directly from g, consider instead the output obtained from the functional composition of p and f. In this case we define the *p-differential value* as:

$$DV(p, E_f, E_g) = V(E_p)/V(E_f)$$

where we have used E_p to denote the ensemble of values $\{p(f(z)) : f(z) \in E_f\}$. It may be the case that $DV(p, E_f, E_g)$ approximates $DV(E_f, E_g)$ at least as well as $p(f(x))$ approximates $g(H(x))$. We would therefore like to ask how much does one need to increase the source population $\|E_f\|$ in order to insure that the p-differential value is at least as large as the differential value. We define the *information loss* due to p as:

$$IL(p) = DV(p, E_f, E_g)/DV(E_f, E_g)$$

This is an instantaneous measurement in that it depends on the populations E_f and E_g. Therefore, we also define the *population multiplier* to be the value k such that for all E_f and E_g and for any ϵ we have that

$$\|E_p\| > k\|E_f\| \Rightarrow IL(p) \geq (1 - \epsilon)$$

Thus, if we insure that the target population E_p is at least k times as large as the source population E_f then almost no information is lost. Using the theorem above we have immediately

Corollary 1 *If f is a one way function then $k = 1$.*

This corollary expresses the fact that if f is itself a one way function then sufficient entropy is lost in making measurements $f(x)$ that it is possible to obtain (almost) perfect correlation with the measurements $g(H(x))$, e.g. that $p(x)$ can be any deterministic polynomially invertible function, including the identity function.

4. Conclusion

Why does this matter? The theorem and is corollary show that we can offer privacy and security for stored and transmitted text while still permitting statistical analysis of content. If x is an element of the ensemble E, we can prevent identification of x via a one way function, such as a hash function, while preserving the properties associated with x and E. Although x will be unrecoverable via any polynomial time algorithm, its properties are still measurable.

References

[1] J. Katz, Y. Lindell, *Introduction to Modern Cryptography*, Chapman & Hall/CRC, 2007.

[2] O. Goldreich, *Foundations of Cryptography: Volume 1, Basic Tools*, Cambridge Univeristy Press, 2007.

[3] M. Naor, M. Yung, *Universal One-Way Hash Functions and Their Cryptographic Applications*, Proc. Twenty First ACM Symposium on the Theory of Computation, 1989.

References