SOME OBSERVATIONS ON THE CRYPTOGRAPHIC HASH FUNCTIONS

by
Lavinia Ciungu

Abstract. In this paper we will make a discussion on the conditions when a strongly collisionfree hash function is also one-way hash function and also some considerations about the security of Chaum-van Pfitzmann hash function, namely it is analyzed the possibility of birthday to this function.

1. Introduction

Hash functions are very important in cryptography where their main role is in the provision of message integrity checks and digital signatures. Because the encryption and digital signature algorithms are generally slowly for the large messages, it is faster to apply a hash function to the message and then the cryptographic algorithm to the message's hash value which is smaller compared to the message itself. Many studies are dedicated to the security of hash functions taking in consideration the properties of these functions, but there are still not definitively decided the conditions when the cryptographic hash function is considerate secure. In Introduction we remember some elements of Number Theory, the definition of cryptographic hash function and the principle of birthday attack. In Section 2 we discuss some properties of hash function and in Section 3 we make some considerations regarding the security of Chaum-van Heijt-Pfitzmann hash function.

1.1 Elements of Number Theory

We remember some elements of Number Theory, which will be used in this paper.

Definition 1. Let G be a cyclic group of order n and g a generator of G. The **discrete logarithm** (or **the index**) of y modulo n to the base g is any integer x such that $g^x = y$. The integer x is determined uniquely modulo n.

It is denoted by $x = \log_g y$ or $\mathrm{ind}_g y$.

A method to compute the discrete logarithm is given in [3].

Definition 2. If both p and 2p + 1 are prime, then p is called **Sophie Germain prime**.

The first few Sophie Germain primes are 2,3,5,11,23,29, 41,53,83,89,113,131.

Around 1825, Sophie Germain proved that the first case of Fermat's Last Theorem is true for such primes.

Definition 3. Let f, g : **N** $\rightarrow$ $\mathbf{R}_+$ and $n_o \in \mathbf{N}$, $c \in \mathbf{R}_+$ such that $f(n) \leq C$. $g(n)$, $\forall n \geq n_o$. Then we may say that $f = O(g)$.

Proposition 1. If f, g : **N** $\rightarrow$ $\mathbf{R}_+$ and $\lim_n \rightarrow \infty \frac{f(n)}{g(n)} = C > 0$ (constant) then $f = O(g)$.

From a computational point of view there are the following types of complexity:

O(1) — constant
O(n) — linear
$O(n^2)$ — quadratic
$O(n^k)$ — polynomial
$O(a^n)$ — exponential (a > 1)
O(logn) — logarithmic.

Because $\lim_{n\to\infty} \frac{C_n^k}{n^k} = \frac{1}{k!}$, then $C_n^k = O(n^k)$, so it is of a polynomial complexity.

1.2 Definitions of hash function

Through the rest of this paper we will use the following notations and definitions:

Definition 4. A hash function h : X $\rightarrow$ Y is a transformation that takes an input m and returns a fixed-size string, which is called the **hash value** or the **message digest.**

Definition 5. A hash function h is **weakly collision-free** if, given a message x, it is computationally infeasible to find a message x' $\neq$ x such that h(x') $\neq$ h(x).

Definition 6. A hash function h is **strongly collision-free** if it is computationally infeasible to find messages x and x' such that x' $\neq$ x and h(x) = h(x).

It is obviously that strong collision-free implies weakly collision-free.

Definition 7. A hash function h is **one-way** if, given a message digest y, it is computationally infeasible to find a message x such that h{x) = y.

Definition 8. A hash function is a **cryptographic hash function** if it satisfies the

following requirements:

a) h(x) is relatively easy to compute for any given x
b) h is one-way
c) h is strongly collision-free

Details and examples of cryptographic hash function can be found in [3].

1.3 The birthday attack

Theorem 1. Let h : X $\rightarrow$ Y be a cryptographic hash function and n = |Y|. If we choose k random
elements $x_1, x_2 \text{--} . x_k \in X$, then the probability of at least one collision is

$$P = 1 - e^{\frac{k^2}{2n}}$$

or, equivalent, in order to get at least one collision with the probability P, the number k of random chosen elements is estimated to be

$$k = \sqrt{2n \cdot \ln \frac{1}{1-p}}.$$

For proof see [1].

So. on attack to the hash function can be possible with the probability *P* which depends on the size of the number *n*.

In order to get P = 1/2 the number *k* is estimated to be $1.18 \cdot \sqrt{n}$.

If *X* is the set of persons and *Y* is the set of possible birthdays, then, using the above result, one can conclude that in a group of *k*=23 persons, two of them will share a birthday with a probability at least 1/2 (obviously, *n* = 365).

This surprising result is known as the **birthday paradox.**

2. On some properties of cryptographic hash functions

Let's consider $h : X \rightarrow Y$ a cryptographic hash function with *X* and *Y* finite and $|X| \geq n|Y|$, $n \in \mathbf{Z}$,where by |*A*| we denote the cardinality of the set *A*.

Let's compute the probability to have a collision, more exactly to determine the probability *P* to find $x, x' \in X$ such that $h(x') = h(x)$.

For $y \in h(X)$ put $Z(y) = \{x \in X \mid h\{x) = y\}$ and $N(y) = |Z\{y)|$. Then for any $x \in Z\{y)$, the probability to find $x' \neq x$ with $h(x') = h(x)$ is $\frac{N(y)-1}{N(y)}$.

The probability *P* to find $x' \neq x$ such that $h(x') = h(x)$ will be:

$$P = \frac{1}{|X|} \cdot \sum_{y \in h(X)} \sum_{x \in Z(y)} \frac{N(y)-1}{N(y)} = \frac{1}{|X|} \cdot \sum_{y \in h(X)} (N(y)-1)$$

$$= \frac{|X| - |h(X)|}{|X|} \geq \frac{|X| - |Y|}{|X|} \geq \frac{|X| - \frac{|X|}{n}}{|X|}$$
$$= 1 - \frac{1}{n}$$

Now we will establish a condition sufficient to prove that the strongly collision-free implies one-way property.

Theorem 2 *If X,Y are finite sets with* $|X| \geq n\text{-}|Y|$ *with n a large integer number and* $h : X \to Y$ *is a strongly collision-free hash function, then h is one-way.*

Proof: By way of contradiction we suppose that *h* is not one-way function.

Let *x* be a random element of *X* and *y = h(x).* According to the above consideration, if *n* is a large number with a high probability we can find $x' \neq x$ such that $h(x') = h(x)$, but this is a contradiction to *h* being strongly collision-free.

Remark 1 *In case when n = 1 we cannot obtain the conclusion of Theorem* ??, *which implies that the implication of the two properties of hash function can be true or false. In case of n = 2 this property is discussed in [1].*

3. On the attacks to Chaum-van Heijst-Pfitzmann hash function

Suppose *q* a Sophie Germain large prime and *p* = 2*q* + 1. Let *a* and *b* be two primitive elements of $\mathbf{Z}_p$ such that it is computationally infeasible to compute $\log_a b$.

Definition 9. *The hash function* $h : \{0,1, ...,q\text{-}1\}X\{O,1_: ...,q\text{-}1\} \to Z_p$ *defined by* $h(x_1,x_2) = a^{x_1} b^{x_2}$ *(modulo p) is called* **Chaum-van Heijt-Pfitzmann hash function.**

In [1] it is proved that the function *h* above denned is strongly collision-free, using the fact that it is infeasible to compute the discrete logarithm $\log_a b$ when *p* is a large prime number.

Also, according to the paragraph 1.3, a birthday attack can be produced with a probability depending on the size of number *p.*

We will make some remarks regarding the security of Chaum-van Heijt-Pfitzmann hash function starting from the question: can a sufficient large Sophie Germain prime be found? The answer is affirmative: In [2] it is introduced a method to find large Sophie Germain primes *q* of the form:

$$q = c \cdot 3003 \cdot 10^6 - 1.$$

In 1994 using six computers during 110 days, it was found such a number *q*

with 4536 digits for c = 1803301 and 6 = 4526. This number is at least 10^{4526}.

On the Internet can be accessed the Sophie Germain Primes web page where the last large found Sophie Germain prime number is updated in real time.

On 4th of July 2003 the last discovered number q_2 has 33078 digits, being at least 10^{33078}.

After the following considerations, we will realize that these numbers are large enough to prevent a birthday attack.

Using the Theorem ?? in order to succeed in a birtday attack to Chaum-van Heijt-Pfitzmann hash function with the probability $P = 0.0001$ one has to choose $k = \left\lfloor \sqrt{2p \cdot \ln(\frac{1000}{999})} \right\rfloor = [0.004\sqrt{p}\,]$ random elements of $\{0,1,..., q — 1\}X\{0,1,..., q — 1\}$ in $C_{q^2}^{k}$ combinations.

For q_1 we have $k_1 = 6 \cdot 10^{2266}$ and for q_2 we have $k_2 = 6 \cdot 10^{16537}$.

The table 3.1 provides estimations of k for the smallest and largest Sophie Germain known primes. Because we meet a polynomial computational complexity, for the largest number of this form, it is infeasible to produce a birthday attack with the probability 0.001, which, in fact, is very small.

Q	P	k	q^2	$C_{q^2}^{k} = O(q^{2k})$
3	7	9	9	$O(9^{18})$
5	11	12	25	$O(25^{50})$
11	23	17	121	$O(121^{34})$
23	47	25	529	$O(529^{50})$
29	59	28	841	$O(841^{56})$
41	83	33	1681	$O(1681^{66})$
53	107	38	2809	$O(2809^{76})$
83	167	48	6889	$O(6889^{96})$
89	179	49	7921	$O(7921^{98})$
113	227	56	12769	$O(12769^{112})$
131	263	60	17161	$O(17161^{120})$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
q_1	p_1	k_1	q_1^2	$O(q_1^{2k_1})$
q_2	p_2	k_2	q_2^2	$O(q_2^{2k_2})$

Table 3.1

References

[1] D. R. Stinson, *Cryptography* , Theory and Practice, CRC Press, (1995).
[2] Harvey Dubner, *Large Sophie Germain Primes***,** Mathematics of Computation, Vol.65(No.213), (January 1996).
[3] Lavinia Ciungu, *Applications of Number Theory in Cryptography***,** Master Thesis , Faculty of Mathematics, University of Bucharest, (2003).

Author:
Lavinia Ciungu: Department of Mathematics, University of Bucharest, Academiei 14, Bucharest, Romania