Healthcare Data Sharing Example using Zero-Knowledge Proofs

May 3, 2020

1 Building Blocks for Range Proof

The following are the mathematical and cryptographic techniques that are used in range proofs for zero knowledge proofs.

1.1 The Discrete Logarithm problem

In real numbers, the logarithm $\log_b a$ is a number x such that $a = b^x$, for given numbers a and b. For example, $5^3 = 125$ is logarithmic equivalent to $\log_5(125) = 3$

The discrete logarithm problem can be defined as: given a group G, a generator g of the group and an element h of G, to find the discrete logarithm to the base g of h in the group G. More formally, Find v such that

$$h = g^v \text{ or } \log_g h = v$$

Discrete logarithm problem is not always hard. The hardness of finding discrete logarithms depends on the groups. For example, if we take a cyclic group, calculating the discrete logarithm appears to be very hard and thus infeasible to compute. There are no polynomial time algorithms that solve such Discrete logarithm problem. Therefore, if we choose large enough numbers solving, the discrete logarithm becomes impractical to solve.

Note that, a cyclic group is a group of numbers generated by a single element g, called the generator or primitive root.

1.2 Integer factorization problem

In number theory, integer factorization problem that is the problem of decomposing a large positive non prime integer, referred to as composite number, into a product of smaller integers e.g., $18 = 2 * 9$. If these factors are restricted to only prime numbers then the problem is called *prime factorization problem* e.g., $10 = 2 * 5$.

1.3 Secret order principle:

Secret order principle is based on the Integer factorization problem. According to the secret order principle, if the factors of the modulus are unknown, then is hard to compute the order of an element in a group.

In cyclic group, order of an element is defined and denoted by the power which reduces the element in group to the identity element which is 1. For example, in the integer modulo 5 i.e., Z_5^*, element 2 has the order 4 as shown below:

$$2 * 2 * 2 * 2 \equiv 1 \bmod 5$$

1.4 Chinese remainder theorem

Let us consider the integer factors $n_1, \ldots, n_k$ of a composite number N such that $N = n_1 \cdot n_2 \ldots, n_k$. If these $n_1, \ldots, n_k$ are pairwise coprime meaning that they don't have common factors, except the number 1. if $a_1, \ldots, a_k$ are any integers, then the theorem states that there exists an integer x such that

$$x \equiv a_1 \bmod n_1 \tag{1}$$

$$\vdots \tag{2}$$

$$x \equiv a_k \bmod n_k \tag{3}$$

and any two such x are congruent modulo N.

Note that for prime factorization of N i.e., $N = pq$ such that p and q are prime the the system of equations

$$x \equiv a \bmod p \tag{4}$$

$$x \equiv b \bmod q \tag{5}$$

has a unique solution for $x \bmod pq$ or $x \bmod N$. For example Consider $N = 5 \cdot 7$, for $p = 5$ and $q = 7$. We want to solve following system of equation:

$$x \equiv 2 \bmod 5 \tag{6}$$

$$x \equiv 3 \bmod 7 \tag{7}$$

If we have a solution y, then $y+35$ is also a solution. So we only need to look for solutions modulo 35. By brute force, we find the only solution is $x = 17 \bmod 35$.

1.4.1 The Fujisaki-Okamoto Commitment Scheme

With the secret order principle, Fujisaki-Okamoto Commitment Scheme Damgård and Fujisaki 2002 is a special proof technique can be guaranteed: the prover proves that he knows a secret integer (which he already committed in the commitment), so that the committed integer is computationally binded and unique. Here binded means that the prover cannot change his secret value later in order to gain some kind of advantage and unique means he cannot commit to more than one secret value at a time.

In cryptography, a commitment scheme allows the prover to commit to a secret value v (i.e., integer), with the ability to reveal the committed value later. Commitments are used to bind a prover to a secret value so that he cannot change his commited secret value later.

Here we are going to define the Fujisaki-Okamoto commitment scheme that is used in our range proofs. The Fujisaki-Okamoto commitment function is designed in a large cyclic group modulo the composite modulus. Since the prime factorization of the modulus is hard, the order of the cyclic group is secret.

Let δ be a security parameter and $N > 1$ be a composite number such that factorization of N are unknown to both: Prover and Verifier. Let g is an element of the modulo group Z_N^* (here $*$ means that 0 is excluded from Z_N) and h is another element of Z_N^* which is generated by g such that both the discrete logarithm of g in base h i.e., $\log_h g$ and the discrete logarithm of h in base g i.e., $\log_g h$ are unknown by Prover.

A commitment to integer secret value v in base (g, h) is defined by

$$C = \text{Commit}(v, r) = g^v h^r \mod N,$$

for r being a randomly number chosen from $-2^\delta N + 1, \ldots 2^\delta N - 1$.

1.5 Range proof Algorithm

We have to prove that the secret value v is in range $(a, a+1, \ldots, b)$. As a result of Boudot 2000 (Boudot 2000), proof of integer v in a range $(a, a+1, \ldots, b)$ can be reduced to a proof that $v - a$ and $b - v$ are non negative i.e.,

$$v \in (a, b) \text{ is true iff } (v - a + 1)(b - v + 1) > 0$$

Peng and Bao, 2010 Peng and Bao 2010 called this technique as computational bindingness through proof of knowledge in cyclic groups with secret order or CBPKCGSO in short.

Now for any random challenge γ, given by verifier, proving of statement $(v - a + 1)(b - v + 1) > 0$ is equivalent to prove $\gamma^2(v - a + 1)(b - v + 1) > 0$ that is,

$$(v - a + 1)(b - v + 1) > 0 \text{ is true iff } \gamma^2(v - a + 1)(b - v + 1) > 0$$

Notation: For simplicity we will use the notation

$$L = (v - a + 1) \tag{8}$$
$$R = (b - v + 1) \tag{9}$$

Therefore, proving that $v \in (a, b)$ is equivalent to proving the following statement:

$$\gamma^2(v - a + 1)(b - v + 1) > 0 \text{ or } \gamma^2 LR > 0 \tag{10}$$

1.6 How to prove that $\gamma^2 LR > 0$?

To prove that $\gamma^2 LR > 0$ i.e., positive, it is divided into three shares v_1, v_2, v_3, whose sum is equal to $\gamma^2 LR$. Moreover, v_3 is a square and therefore, alreday a non-negative number. If both $sv_1 + v_2 + v_3$ and $v_1 + tv_2 + v_3$ are positive where s and t are random positive integers, then $\gamma^2 LR$ is positive with an overwhelmingly large probability. So, the procedure is as follows:
Take random non-negative integer $v_1,\ v_2,\ v_3$ such that the following holds:

$$v_1 + v_2 + v_3 = \gamma^2 LR;\ v_3 = v_4^2, \tag{11}$$

for v_3 being the square of some non-negative integer v_4 hence, always positive. Let us define the two variable x and y as

$$x = sv_1 + v_2 + v_3 \tag{12}$$

$$y = v_1 + tv_2 + v_3 \tag{13}$$

Consider the fact that

$$\text{If } x > 0 \text{ and } y > 0 \implies \gamma^2 LR > 0$$

This holds with overwhelmming large probabilty. Moreover, x and y revealed negligble information about the secret value v.

1.7 How to convince the verifier?

However, publishing x and y does not give sufficient information to convince a verifier, since x and y could be random numbers. Therefore, the verifier needs to be convinced that the proover did actually use the correct commitment. This can be done by the use of secret order groups explained above .

1. First we will calculate the secret order given by N , of group $\mathbb{G}$ by using the prime factorizatoion method as follow:

   Choose two large primes P and Q such that $\frac{P-1}{2}$ is also a prime. Then, N is given by $N = PQ$.

   Note that the factorization of N remains unknown to both: Prover and Verifier.

2. Compute generators g and element h (generated by g) of the group $\mathbb{G}$ of secret order N. We will calculate this by using generators of the subgroup mod P and mod Q and chinese remainder theorem.

   Note that the Chinese Remainder Theorem (CRT) is a technique to reduce modular calculations with large moduli to similar calculations for each of the (mutually co-prime) factors of the modulus.

3. Fujisaki-Okamoto commitment C to the secret value v is given by $C = g^v h^r \bmod N$. Usig this commitment scheme, prover publishes three

commitments C_1, C_2 and C', defined as below:

$$C_1 = \frac{C}{g^{a-1}} \bmod N \tag{14}$$

$$C_2 = \frac{g^{b+1}}{C} \bmod N \tag{15}$$

$$C' = C_1^R h^{r'} \bmod N \tag{16}$$

where r' is random integer. **It gives the proof that C_2 and C' hides the same secret.** Prover then publishes $C'' = C'^{\gamma^2} h^{r''} \bmod N$ and **this way he proofs that the hidden number is a square with base (C', h).**

4. Prover chooses v_1, v_2, v_3 such that (1.6) holds and r_1, r_2, r_3 such that $r_1 + r_2 + r_3 = \gamma^2(Rr + r') + r''$ and publishes

$$C_1' = g^{v_1} h^{r_1} \bmod N \tag{17}$$

$$C_2' = g^{v_2} h^{r_2} \bmod N \tag{18}$$

$$C_3' = g^{v_3} h^{r_3} \bmod N \tag{19}$$

By doing so the prover gives the proof that C_3' hides a square.

5. Finally x, y given by (12), $u = sr_1 + r_2 + r_3$, $w = r_1 + tr_2 + r_3$ are published.

1.8 What will verifier check?

Verifier will check if $x > 0$ and $y > 0$ that is x and y are positive and also verify the following system of equations of commitments published by prover:

$$C_1 = \frac{C}{g^{a-1}} \bmod N \tag{20}$$

$$C_2 = \frac{g^{b+1}}{C} \bmod N \tag{21}$$

$$C'' = C_1' C_2' C_3' \bmod N \tag{22}$$

$$C_1'^s C_2' C_3' = g^x h^u \bmod N \tag{23}$$

$$C_1' C_2'^t C_3' = g^y h^w \bmod N \tag{24}$$

2 Our ZKP Mechanism as Solution

Our example is built on range proof technique provide by Peng and Bao (2010). Diabetic patients shall target a blood sugar level that lies within the range from 72 to 120 milligrams per deciliter (mg/dl). The diabetic patient from which we

have data, had an average blood sugar level of 114 (mg/dl) for the period in which he donated the data. By using the zero knowlege range proof technique, this patient (prover) is able to prove to the insurence company(verifier) that his average blood sugar level lies within the expected range, without the need of disclosing each of the values from which the average emerges. Since there is a direct relationship between the average blood sugar level for a period of three months, and the HbA1c level, proving that the avergae sugar level lies within the adequate range suffices to prove that the HbA1c level lies within the range that the insurance company needs.

To prove that the average sugar level lies within the adequate range, the prover splits the secret into three non negative shares such that one share is sqaured terms. These three shares are still hidden from the verifirer. In order to be sure that the prover will not change the secret value later, he commit to the original secret (114) using a commitment scheme. He also commit to the three shares by using the same commitment scheme. Now, the verifier can verify that whether the commitment to the original secret value is equal to the sum of commitments to three shares. This way verifier will be convinced with overwhelming propability that the average value lies within the given range and infact the prover is telling truth. The whole procedure of zero knowledge range proof works as follows:

We denote the values 114 by v, 72 by a and 120 by b. Here, v is the secret number, $a = 72$ is lowest possible value of blood sugar level considered as normal while $b = 120$ is highest possible value of blood sugar level considered as normal. Since, proving that an integer v lies in range $(a, a+1, \ldots, b)$ is equivalent to providing the two proofs that $v - a + 1$ and $b - v + 1$ are non negative. Form the fact that multiplication of two non-negative integer wil always result in a non-negative integer, we can combine these two proofs into one by simply multiplying them. Therefore, v belongs to $(a, a+1, \ldots b)$ is true if and only if $(v - a + 1) * (b - v + 1)$ is non-negative. Square of any integer is always positive and multiplying it to any non-negative integer will not change the sign, this will convince the verifier. Thus if prover can provide the proof that $\gamma^2(v - a + 1) * (b - v + 1)$ is non-negative for any random challenge γ^2 his job is done. For example, if random challenge is $\gamma = 2$, value of $\gamma^2(v-a+1)*(b-v+1)$ is equal to 1204. So, now the new secret becomes $\gamma^2(v-a+1)*(b-v+1) = 1204$.

To do so, prover randomly split the new secret 1204 into three non negative integers say v_1, v_2, v_3, where v_3 is a square of another non-negative integer say v_4 . For example if he chooses $v_1 = 300, v_2 = 279$ and $v_3 = 625$ then $v_1 + v_2 + v_3 = 1204$, also $v_3 = v_4^2$ for $v_4 = 25$. Prover also chooses three random integers as $r_1 = 26, r_2 = 30, r_3 = 100$ to hide the shares v_1, v_2 and v_3. Note that, sum of r_1, r_2, and r_3 should equal to $\gamma^2((b - v + 1)r + r') + r''$.

Verifier knows that one share is in square form so it is automatic non negative. However, to be sure that v_1 and v_2 are really non-negative, verifier sends two positive random integer $s = 11$ and $t = 7$ to the prover. From the fact that multiplication of two non-negative integer is also a non-negative integer. If values of $x = sv_1 + v_2 + v_3$ and $y = v_1 + tv_2 + v_3$ calculated by the prover are positive then the verifier will be sure that v_1 and v_2 are also

positive. So, the prover calculates $x = 4204$ and $y = 2878$. For the same reason, he also calculates u and w as $u = sr_1 + r_2 + r_3 = 416$ and $w = r_1 + tr_2 + r_3 = 336$ are published.

He choose the value of $N = 21(P = 7\&Q = 3)$, $g = 5$ $h = 4$ and $r = 5$ in the Fujisaki-Okamoto commitment Scheme to calculate the following commitment values:

1. Commitment C to the secret value $v = 114$ as $C = g^v h^r \bmod N$ which equals to 16.
2. Commitment C_1 to the value $(v - a + 1) = 43$ as $C_1 = \frac{C}{g^{a-1}} \bmod N$ which equals to 17.
3. Commitment C_2 to the value $(b - v + 1) = 7$ as $C_2 = \frac{g^{b+1}}{C} \bmod N$ which equals to 20.
4. Commitment C' to the value $(v - a + 1) * (b - v + 1) = 301$ as $C' = C_1^{(b-v+1)} h^{r'} \bmod N$ which equals to 20, for $r' = 2$ being another random integer. This gives the proof to the verifier that C_2 and C' hides the same secret $(b - v + 1)$.
5. Commitment C'' to the new secret value $\gamma^2(v - a + 1) * (b - v + 1) = 1204$, $C'' = C'^{\gamma^2} h^{r''} \bmod N$ which equals to 16, for $r'' = 8$ being another random integer. This gives the proof to the verifier that the secret number is a square (γ^2) .
6. Commitment C_1' to the first share v_1 as $C_1' = g^{v_1} h^{r_1} \bmod N$ which equals to 16.
7. Commitment C_2' to the second share v_2 as $C_2' = g^{v_2} h^{r_2} \bmod N$ which equals to 20.
8. Commitment C_3' to the third share v_3 as $C_3' = g^{v_3} h^{r_3} \bmod N$ which equals to 20. This also gives the proof to the verifier that C_3' hides a square v_4.

After calculating all the values and commitments needed to provide the proof, prover publishes $x, y, u, w, C, C_1, C_2, C', C'', C_1', C_2'$ and C_3'.

Now the verifier knows all the values published by prover along with g, h and N. Therefore, the verifier will check the following:

1. If values of x and y are really non-negative?
2. Whether the value of C_1 and C_2 are calculated correctly? For this he checks if C_1 equals to $\frac{C}{g^{a-1}} \bmod N$ and C_2 equals to $\frac{g^{b+1}}{C} \bmod N$ or not.
3. Whether all the split secret together creates the original secret or not? For this he checks if C'' equals to $C_1' C_2' C_3' \bmod N$ or not?
4. Whether $C_1'^s C_2' C_3'$ equals to $g^x h^u \bmod N$ and $C_1' C_2''^t C_3'$ equals to $g^y h^w \bmod N$ or not?

We can see that all of the above statements are true. Therefore verifier will be convinced that the prover is telling the truth.

References

Boudot, F. (2000). *Efficient Proofs that a Committed Number Lies in an Interval*. Ed. by B. Preneel. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 431–444.

Damgård, I. and E. Fujisaki (2002). "A statistically-hiding integer commitment scheme based on groups with hidden order". In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, pp. 125–142.

Peng, K. and F. Bao (2010). "An Efficient Range Proof Scheme". In: *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SocialCom / IEEE International Conference on Privacy, Security, Risk and Trust, PASSAT 2010, Minneapolis, Minnesota, USA, August 20-22, 2010*. Ed. by A. K. Elmagarmid and D. Agrawal, pp. 826–833.