Setting 128-bit security: Encryption Scheme Parameters Selection

Georgios Sakellariou   Anastasios Gounaris   Konstantinos A. Draziotis

July 31, 2018

We follow the analysis of [WH12] and [GHS12] to select encryption scheme parameters. The symbols that appeared in the following analysis are explained on Table 1.

Table 1: Encryption Scheme Parameters Selection Symbols Explanation

Symbol	Explanation
q	The value of modulo used by the encryption scheme in the ciphertexts.
σ	The standard deviation of the discrete Gaussian distribution.
$\phi(m)$, N	The degree of polynomials used by the encryption scheme.
p	The product of t prime numbers, $p = \prod_{i=1}^{t} prime_i$
h	The exact number of nonzero coefficients in Hamming weighted distribution.
b	Base decomposition parameter
Θ_{res}	The amount of noise that enclosed in the result of a formula.
Θ_i	The amount of noise of a ciphertext c_i, which is equal to E_{clean}, except as otherwise specified.
$\lVert E_{clean}\rVert_{\infty}^{can}$	The minimum noise of a ciphertext
$cipher_{[p_i \in clusterj]}$	The encrypted value of 1 if the point p_i belongs to $cluster_j$, otherwise the encrypted value of 0.

To begin our analysis, we assume that the minimum noise of a ciphertext is that produced during the encryption. Namely, this means that no homomorphic operation can reduce the noise level of a ciphertext lower than the one reached during the encryption. Furthermore, the upper acceptable bound of the noise produced during the encryption is given by [WH12]:

$$\lVert E_{clean}\rVert_{\infty}^{can} \leq 8\sqrt{2}\sigma\phi(m) + 6\sigma p\sqrt{\phi(m)} + 16\sigma p\sqrt{h\phi(m)}$$

While, the acceptable upper bounds of noise added by each homomorphic operation are:

- *Switch Key*: $\lVert E_{KS}\rVert_{\infty}^{can} \leq 6\sigma b(n_s+1)\lceil \log_b q\rceil \phi(m)^{\frac{3}{2}}$, where $n_s = 2$ in case of switch key between two different keys and $n_s = 4$ when switch key is conducted after a homomorphic multiplication.

- *Homomorphic Addition*: the noise of a ciphertext which is the product of homomorphic addition between two ciphertexts is equal with the sum of their noises.

- *Homomorphic Multiplication*[1]: $E_{mult} \leq p[2\phi(m)(6\sqrt{h}+4)+1]E + \left[\frac{1}{2}\phi(m)(1+12\sqrt{h}+16h)\right] + \left[p^2\phi^2(m)[2(6\sqrt{h}+3)+1]\right]$, where E is the maximum upper bound noise of the multiplier

[1] The noise added during a homomorphic multiplication can further be analyzed in two components the first is the noise added by the multiplication process and the second is the noise added by the rounding process. So, $E_{mult} = E_m + E_{round}$, where:
$E_m = E_1 \leq p[2\phi(m)(6\sqrt{h}+4)+1]E + \left[p^2\phi^2(m)[2(6\sqrt{h}+3)+1]\right]$, and
$E_{round} = E_2 \leq \frac{1}{2}\phi(m)(1+12\sqrt{h}+16h)$

and multiplicand ciphertexts.

- *Scalar Multiplication*: The scalar multiplication of a ciphertext with a number $\kappa \in Z/pZ$ results to an increase of its noise by a factor of $|\kappa|$.

Based on the analysis of [GHS12] a *z-bit* security level holds when:

$$\phi(m) = N \geq \frac{(z + 110) \log_2 \left(\frac{q}{\sigma}\right)}{7.2}$$

So, to ensure a *128-bit* security level, we have:

$$\phi(m) = N \geq 33.1 \log_2 \left(\frac{q}{\sigma}\right) \tag{1}$$

Moreover, we adopt the conventions of [GHS12] and [WH12] about the parameters σ, h and b. So, we set $\sigma = 3.2$, $h = 64$ and $b = 2^{24}$. In addition, any computation is performed with respect on primes $prime_1, prime_2, prime_t : \prod_{i=1}^{t} prime_i \geq 2^{128}$.

Replacing the values in the above inequalities, the bound of noise added during the encryption takes the form:

$E_{clean} \leq 8 \cdot \sqrt{2} \cdot 3.2 \cdot N + 6 \cdot 3.2 \cdot (2 \cdot N + 3) \cdot \sqrt{N} + 16 \cdot 3.2 \cdot (2 \cdot N + 3)\sqrt{64 \cdot N} \approx 1287 \cdot \sqrt{N} + 37 \cdot N + 858 \cdot N^{\frac{3}{2}} = 2^{10.3} \cdot \sqrt{N} + 2^{5.2} \cdot N + 2^{9.7} \cdot N^{\frac{3}{2}} \leq 2^{10} \cdot N^{\frac{3}{2}} \Rightarrow$

$$\boxed{E_{clean} \leq 2^{10} \cdot N^{\frac{3}{2}}}$$

where the inequality holds for $N \geq 9$ [2].

The amount of noise added during the multiplication is bounded by:

$E_{mult} \leq (2 \cdot N + 3)[2 \cdot N \cdot (6 \cdot \sqrt{64} + 4) + 1] \cdot E + \Big[\frac{1}{2} \cdot N \cdot (1 + 12 \cdot \sqrt{64} + 16 \cdot 64)\Big] + \Big[(2 \cdot N + 3)^2 \cdot N^2 \cdot [2 \cdot (6 \cdot \sqrt{64} + 3) + 1]\Big] = (2 \cdot N + 3)(104 \cdot N + 1) \cdot E + 560.5 \cdot N + 103 \cdot (2 \cdot N + 3)^2 \cdot N^2 \leq (2 \cdot N + 3)(2^{6.8} \cdot N) \cdot E + 2^{9.2} \cdot N + 2^{6.7} \cdot (2 \cdot N + 3)^2 \cdot N^2 \leq (2^{8.4} \cdot N + 2^{7.9} \cdot N^2) \cdot E + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4 \Rightarrow$

$$\boxed{E_{mult} \leq (2^{8.4} \cdot N + 2^{7.9} \cdot N^2) \cdot E + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4}$$

The amount of noise added during the switch key is bounded by:

$E_{KS} \leq 6 \cdot 3.2 \cdot 2^{24} \cdot (4 + 1)\lceil \log_b q \rceil N^{\frac{3}{2}} \leq 2^{30.6} \cdot \lceil \log_b q \rceil N^{\frac{3}{2}} = 2^{30.6} \cdot \left\lceil \frac{\log_2 q}{\log_2 2^{24}} \right\rceil N^{\frac{3}{2}} = 2^{30.6} \cdot \left\lceil \frac{\log_2 q}{24} \right\rceil N^{\frac{3}{2}} \Rightarrow$

$$\boxed{E_{KS} \leq 2^{30.6} \cdot \left\lceil \frac{\log_2 q}{24} \right\rceil N^{\frac{3}{2}}}$$

Having determine the upper bounds of noise added by the fundamental homomorphic operations, we continue our analysis in order to determine the upper bounds of noise that added during the computation of homomorphic formulas which compose the k-means. Namely, we focus on the formulas of Manhattan distance d_M[3], Euclidean distance d_E, the centroids data in Technique I and the centroids data in Technique II.

To estimate the bound of noise added during the computation of the Manhattan distance, $E_{P_k}(d_M(p_1, p_2)) = c_1 - c_2$, we have already notified that one scalar multiplication and one addition are taken place. So,

[2] On that point, we follow the convention of [WH12]. Specifically, this means that the encryption scheme uses polynomials with degree $N \geq 9$, which does not affect the k-means.

[3] From a noise bound estimation perspective, the computation of d_M has no difference between the 1st and 2nd representation method. This results from the independent computation of each coefficient of d_M, when 2nd representation method takes place. The same occurs in case of centroids data computation.

$$\Theta_{res} = \Theta_1 + |\kappa|\Theta_2 = \Theta_1 + |-1|\Theta_2 = \Theta_1 + \Theta_2 = 2E_{clean} \leq 2 \cdot 2^{10} \cdot N^{\frac{3}{2}} \Rightarrow$$

$$\boxed{\Theta_{res} \leq 2^{11} \cdot N^{\frac{3}{2}}}$$

To estimate the bound of noise added to the result of Euclidean distance, $E_{P_k}((d_E(p_1,p_2))^2) = \sum_{j=1}^{d}(c_j^1 - c_j^2)(c_j^1 - c_j^2)$, lets first determine the noise bound, Θ_{pair}, of the difference between a pair of coefficients,$c_j^1 - c_j^2$. So, the noise bound of the difference is,

$$\Theta_{pair} = \Theta_1 + |\kappa|\Theta_2 = \Theta_1 + |-1|\Theta_2 = \Theta_1 + \Theta_2 = 2E_{clean} \leq 2 \cdot 2^{10} \cdot N^{\frac{3}{2}} \Rightarrow$$

$$E = \Theta_{pair} \leq 2^{11} \cdot N^{\frac{3}{2}}$$

Having determine the noise bound of the difference, we can estimate the noise bound of the square of this difference, Θ_{square}, which is equal of the multiplication noise bound with $E \leq 2^{11} \cdot N^{\frac{3}{2}}$ plus the noise added by a key switch operation and it is given by:

$$\Theta_{square} \leq (2^{8.4} \cdot N + 2^{7.9} \cdot N^2) \cdot E + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4 + 2^{30.6} \cdot \left\lceil \frac{\log_2 q}{24} \right\rceil N^{\frac{3}{2}}$$

In order to simplify the above inequality, we use the inequality provided by [WH12], $2^8 N^2 \geq 2^{8.4} \cdot N + 2^{7.9} \cdot N^2$ and the fact that we can further bound the formula $\left\lceil \frac{\log_2 q}{24} \right\rceil \leq 128 = 2^7$, because we can assume that always $q < 2^{3072}$.

$$\Theta_{square} \leq 2^8 \cdot N^2 \cdot E + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4 + 2^{37.6} \cdot N^{\frac{3}{2}} \Rightarrow$$

$$\Theta_{square} \leq 2^{19} \cdot N^{\frac{7}{2}} + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4 + 2^{37.6} \cdot N^{\frac{3}{2}} \Rightarrow$$

$$\Theta_{square} \leq 2^{20} N^{\frac{7}{2}}$$

Finally, we observe that the result of the Euclidean distance formula is the sum of d squared differences. Therefore, $\Theta_{res} = d \cdot \Theta_{square} \Rightarrow$

$$\boxed{\Theta_{res} \leq d \cdot 2^{20} N^{\frac{7}{2}}}$$

To estimate the bound of noise added during the computation of new centroids data in Technique I, $centroid\_data_j = \sum_{i=1}^{n} c_i[p_i \in clusterj]$[4], in the worst case, when all points clustered to one cluster, we estimate the noise of a ciphertext that results from the addition of at most n ciphertexts. So,

$$\Theta_{res} = \Theta_1 + ... + \Theta_d = nE_{clean} \leq n \cdot 2^{10} \cdot N^{\frac{3}{2}} \Rightarrow$$

$$\boxed{\Theta_{res} \leq n \cdot 2^{10} \cdot N^{\frac{3}{2}}}$$

To estimate the bound of noise added in the results of the new centroids data formula in Technique II, $centroid\_data_j = \sum_{i=1}^{n} c_i cipher_{[p_i \in clusterj]}$[5], we observe that the produced ciphertext is the summation of n products of the form $c_i \cdot cipher_{[p_i \in clusterj]}$ each followed by a key switch operation. So, we have with $E = E_{clean} \leq 2^{10} \cdot N^{\frac{3}{2}}$:

$$\Theta_{res} = n \cdot (E_{mult} + E_{KS}) \Rightarrow$$

$$\Theta_{res} \leq n \cdot \Big[(2^{8.4} \cdot N + 2^{7.9} \cdot N^2) \cdot E + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4 +$$

$$2^{30.6} \cdot \left\lceil \frac{\log_2 q}{24} \right\rceil N^{\frac{3}{2}}\Big] \Rightarrow$$

$$\Theta_{res} \leq n \cdot \Big[2^8 \cdot N^2 \cdot E + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4 + 2^{37.6} \cdot N^{\frac{3}{2}}\Big] \Rightarrow$$

$$\Theta_{res} \leq n \cdot \Big[2^{18} \cdot N^{\frac{7}{2}} + 2^{9.2} \cdot N + 2^{9.9} \cdot N^2 + 2^{10.3} \cdot N^3 + 2^{8.7} \cdot N^4 + 2^{37.6} \cdot N^{\frac{3}{2}}\Big] \Rightarrow$$

[4] Resp. $centroid\_data_j = \left(\sum_{i=1}^{n} c_0^i[p_i \in clusterj], ..., \sum_{i=1}^{n} c_d^i[p_i \in clusterj]\right)$ in case of 2nd rep. method

[5] Resp. $centroid\_data_j = \left(\sum_{i=1}^{n} c_0^i cipher_{[p_i \in clusterj]}, ..., \sum_{i=1}^{n} c_d^i cipher_{[p_i \in clusterj]}\right)$ on 2nd rep. method

$$\boxed{\Theta_{res} \leq n \cdot 2^{19} \cdot N^{\frac{7}{2}}}$$

The estimation of formula's noise bounds allows the definition of the maximum noise bound of each Technique. So,

- in Technique I, when Manhattan distance is used, the maximum noise bound is:

$$\Theta_{res} \leq n \cdot 2^{10} \cdot N^{\frac{3}{2}} \tag{2}$$

- In Technique I, when Euclidean distance is used, and in Technique II, the maximum noise bound is[6]:

$$\Theta_{res} \leq n \cdot 2^{19} \cdot N^{3.5} \tag{3}$$

The analysis of [WH12] shows that in order the decryption to be safe, the inequality $\Theta_{res} \leq \frac{\lfloor q/2 \rfloor}{c_m \cdot p} \Rightarrow \Theta_{res} \leq \frac{q}{2 \cdot c_m \cdot (2 \cdot N+3)}$ must be held, where in our implementation $c_m = \frac{4}{\pi} \approx 1.2731 < 2$.
So,in case of (2), the inequality takes the form:

$$n \cdot 2^{10} \cdot N^{\frac{3}{2}} \leq \frac{q}{2 \cdot c_m \cdot (2 \cdot N+3)} \Rightarrow q \geq n \cdot 2^{11} \cdot N^{\frac{3}{2}} \cdot c_m \cdot (2 \cdot N + 3) \overset{8}{\Rightarrow}$$

$$q \geq 3 \cdot n \cdot 2^{12} \cdot N^{\frac{5}{2}} \tag{4}$$

While, in case of (3), the inequality takes the form:

$$n \cdot 2^{19} \cdot N^{\frac{7}{2}} \leq \frac{q}{2 \cdot c_m \cdot (2 \cdot N+3)} \Rightarrow q \geq n \cdot 2^{20} \cdot N^{\frac{7}{2}} \cdot c_m \cdot (2 \cdot N + 3) \overset{8}{\Rightarrow}$$

$$q \geq 3 \cdot n \cdot 2^{21} \cdot N^{\frac{9}{2}} \tag{5}$$

On that point, we solve the systems of inequalities (1), (4) and (1), (5), in order to define the minimum values of N and $\log_2 q$ for different number of data points n, in Technique I, when Manhattan distance is used, and in all other cases, respectively. Table 2 illustrates the respective values of N and $\log_2 q$.

Table 2: Minimum values of N and $\log_2 q$ for a $128 - bit$ security level.

Technique	n	N	$\log_2 q$
Technique I (Manhattan distance)	20	1403	44.043
	100	1486	46.571
All other cases	20	2518	77.748
	100	2602	80.283

In conclusion, the encryption parameters that used as input in the experiments were $\boldsymbol{prime_1 = 2663}$, $\boldsymbol{\log_2 q = 81}$ and $\boldsymbol{g = 7}$. We use the same parameters in all experiments seeking to have equal size ciphertexts, in order to be able the execution time comparison between the techniques.

References

[GHS12] Craig Gentry, Shai Halevi, and Nigel P. Smart. Homomorphic evaluation of the AES circuit. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.

[WH12] David Wu and Jacob Haven. Using Homomorphic Encryption for Large Scale Statistical Analysis (Technical Report from `https://crypto.stanford.edu/~dwu4/papers/FHE-SI_Report.pdf`). 2012.

[6] In case of Euclidean distance this holds, when $n > 2d$.

[8] $c_m < 2$ and $3N \geq 2N + 3, \forall N > 3$.