On the security of MQ_DRBG

V.O. Drelikhov, G.B. Marshalko [*], A.V. Pokrovskiy

Abstract

MQ_DRBG is a pseudorandom number bit generator proposed for international standardization by the French national organization for Standardization (AFNOR). It makes use of a specific instantiation of a one-way function $S : \mathbb{F}_2^n \to \mathbb{F}_2^{n+r}$ based on quadratic multivariate polynomials. We describe two methods for constructing function S, satisfying requirements of the proposed draft, but having less security level.

1 Introduction

In 2010 AFNOR proposed a deterministic pseudorandom number bit generator based on quadratic multivariate polynomials. The security of the generator could be described in terms of complexity of solving of the corresponding system of multivariate quadratic equations. In [4] the specialists from AFNOR present the following arguments in favor of MQ_DRBG:

- well-studied security, based on MQ-problem, which is known to be NP-hard;
- based on elementary operations AND XOR;
- MQ_DRBG is slower than generators based on block ciphers and hash functions, but provably secure;
- MQ_DRBG is much faster than public-key primitives.

In [3] the security of the generator is described in terms of provable security. This approach allows to show that system of functions defining MQ_DRBG satisfies the requirements of ISO/IEC 18031 [7] and well-known BSI's reference document for evaluating pseudorandom number generators AIS20 [8] in case of a random choice of the system and a random choice of the initial state.

The defense against the malicious developer, who tries to deliberately implement "weak" instances of cryptographic algorithms, is one of the challenges for cryptographers nowadays. In case of MQ_DRBG this problem could be rather serious.

[*]gmarshalko@gmail.com

Identification schemes based on the Isomorphism of Polynomials are widely studied lately. Such schemes ([10],[11],[12],[13]) exploit the idea of selecting the easily-invertible systems of nonlinear polynomials and hide them with secret affine (linear) transformations. The resulting system, being supposedly indistinguishable from random, should be hard to solve.

One of the questions concerning MQ_DRBG is the possibility of constructing "weak" instances of multivariate quadratic equations (with less security than stated in [5]), which could be mapped later to "random" ones.

It should be mentioned that in [11] this problem of "hiding" polynomials is described in the following way. Let $F(x) = (f_1(x), \ldots, f_r(x))$ and $G(x) = (g_1(x), \ldots, g_r(x))$ – two systems of quadratic equations. We have to check the existence of two invertible affine (linear) transformations S and U satisfying the following property:

$$G(x) = U(F(S(x))).$$

If U – is a identity matrix, the problem is called the Isomorphism of Polynomials with one secret (IP1S), and IP2S in the other case. In [15] an algorithm to solve IP1S is presented in case that F and G are known. The complexity of the algorithm is $O(n^6)$, where n – is the number of variables.

In case of MQ_DRBG implemented by malicious developer we can't suppose that "weak" G is known, so the mentioned algorithm is not applicable directly.

Further, we will study the possibility of constructing "weak" instances of multivariate quadratic equations, satisfying the requirements of [5], which could be solved with less complexity than stated in [5].

Organization of the paper. In section 2 we recall basic notations and facts about multivariate quadratic equations. In section 3 we describe the specifications of MQ_DRBG. In section 4 we propose two methods for constructing multivariate quadratic equations defining MQ_DRBG, which could be solved using guess-and-determine and meet-in-the-middle techniques and their combination with less complexity than stated in [5].

2 Basic notations and definitions

We denote $\mathcal{F}_{\mathrm{n}}$ – the set of all boolean functions of n variables and AGL – the group of affine transformations (affine group) over $V_n = \{0,1\}^n$. Let $\mathrm{wt}(f)$ be the weight of function $f \in \mathcal{F}_{\mathrm{n}}$, and (α, x) is the scalar multiplication of binary vectors $\alpha, x \in V_n$.

Definition 1. *Functions $f_1, f_2 \in \mathcal{F}_n$ are called affine equivalent, if there exists a pair $(A, \gamma) \in$ AGL such, that $f_1(x \cdot A \oplus \gamma) = f_2(x)$ for all $x \in \mathbb{F}_2^n$. We will use the notation $f_1 \underset{\mathrm{AGL}}{\sim} f_2$ to denote affine equivalence.*

The following theorem describes three classes of quadratic functions (see [1],[2]).

Theorem 1. *Let f – be a quadratic boolean function from $\mathcal{F}_n$, then f is affine equivalent to one of the following types:*

$$q_1^k(x_1,\ldots,x_n) = x_1x_2 \oplus \ldots \oplus x_{2k-1}x_{2k} \oplus 1,$$
$$q_2^k(x_1,\ldots,x_n) = x_1x_2 \oplus \ldots \oplus x_{2k-1}x_{2k},$$
$$q_3^k(x_1,\ldots,x_n) = x_1x_2 \oplus \ldots \oplus x_{2k-1}x_{2k} \oplus x_{2k+1}.$$

The number $2k$, $0 \leqslant k \leqslant [n/2]$, is called the rank of quadratic function. Each of the forms q_i^k $i = \overline{1,3}$ has the weight $2^{n-1} \pm 2^{n-k-1}$, 2^{n-1} correspondingly and belongs to separate orbits (see [2]), induced by AGL. The set of all quadratic and linear boolean functions is closed relatively to the operation of addition and form a Reed-Muller linear code of order 2 which is denoted as $\mathrm{RM}(2,n)$.

Let us also denote

$$A_i = |\{f \in \mathrm{RM}(2,n) \mid \mathrm{wt}(f) = i\}|,$$

then the following theorem is correct.

Theorem 2 ([1])**.** *For $\mathrm{RM}(2,n)$ the weights $A_i, i = \overline{0,2^n}$ takes the following values: $A_0 = A_{2^n} = 1$,*

$$A_{2^{n-1}-2^{n-k-1}} = A_{2^{n-1}+2^{n-k-1}} = 2^{k(k+1)}\frac{(2^n-1)(2^{n-1}-1)\cdot\ldots\cdot(2^{n-2k+1}-1)}{(2^{2k}-1)(2^{2k-2}-1)\cdot\ldots\cdot(2^2-1)}$$

for all $1 \leqslant k \leqslant [n/2]$. Other A_i are equal to zero, except A_{2^n-1}, which could be found from the equation $\sum_{i=0}^{2^n} A_i = 2^{\frac{n^2}{2}+\frac{n}{2}+1}$.

3 Description of MQ_DRBG

Let us consider the system of boolean quadratic functions of MQ_DRBG. The system S consists of $n+r$ functions of n variables and degree 2. The system defines the mapping $V_n \to V_{n+r}$ and i-th step of the generator could be described as follows:

- the value of initial state $x_i \in V_n$ is set up (the value is unknown to a cryptanalyst);
- we apply the mapping S on vector x_i, and get $S(x_i) = (z||y)$, where $y \in V_n$, $z \in V_r$;
- $x_{i+1} = y$;
- z is a pseudorandom vector produced by the generator at the current step.

The last n functions define the transition function of internal states of the generator. We will denote this system $G(x)$, $G : V_n \to V_n$. The first r functions define the output function $F(x)$, $F : V_n \to V_r$. The described functions could be written in the following

way:

$$F(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_r(x) \end{pmatrix}, \tag{1}$$

$$G(x) = \begin{pmatrix} f_{r+1}(x) \\ f_{r+2}(x) \\ \vdots \\ f_{r+n}(x) \end{pmatrix}. \tag{2}$$

where each of coordinate functions is a quadratic boolean function over n variables.

The system S should satisfy the following property.

Definition 2. *The system of functions $f_i(x), i = \overline{1, n+r}$ is said to satisfy the AFNOR property with parameters l_{max} and $\rho_{min}(n)$, if*

- all multivariate quadratic functions $f_i(x), i = \overline{1, n+r}$ have the rank greater or equal to $\rho_{min}(n)$,

- all sums of at most l_{max} functions have the rank greater or equal to $\rho_{min}(n)$.

For different security levels described in the informative part of [7] the authors of [5] suggest different values of parameters n, r, $\rho_{min}(n)$, l_{max} and different finite fields, determined be the complexity of solving the corresponding system of multivariate quadratic equations. We recall these values in the table below.

3.1 The guess-and-determine technique

Let us consider quadratic function

$$q = x_1 (\alpha_1, y) \oplus \ldots \oplus x_k (\alpha_k, y), \quad y, \alpha_i \in V_k, \ i = \overline{1, k},$$

of $2k$ variables where the functions $x_1, \ldots, x_k, (\alpha_1, y), \ldots, (\alpha_k, y)$ are linearly independent over $\mathbb{F}_2$.

By applying nonsingular transformation of the variables $z_{2i-1} = x_i, \ z_{2i} = (\alpha_i, y) \quad i = \overline{1, k}$ we could map the quadratic function q to the following form:

$$q'(z_1, \ldots, z_{2k}) = z_1 z_2 \oplus \ldots \oplus z_{2k-1} z_{2k}$$

of the rank $2k$.

Let us define the map $\tau_q \ : \ V_k \to \mathbb{F}_{2^k}$:

$$\tau_q(y) = (\alpha_1, y)\, \xi \oplus \ldots \oplus (\alpha_k, y)\, \xi^{2^{k-1}},$$

80 2-TDEA	$n = r = 112$ $GF(2)$,$l_{max} = 4$ $\rho_{min}(n) = 106$	$n = r = 128$ $GF(2^4)$,$l_{max} = 5$ $\rho_{min}(n) = 30$	$n = r = 192$ $GF(2^6)$,$l_{max} = 5$ $\rho_{min}(n) = 30$	$n = r = 256$ $GF(2^8)$,$l_{max} = 5$ $\rho_{min}(n) = 30$
112 3-TDEA	$n = 120, r = 112$ $GF(2)$,$l_{max} = 4$ $\rho_{min}(n) = 114$	$n = r = 128$ $GF(2)$,$l_{max} = 5$ $\rho_{min}(n) = 122$	$n = r = 192$ $GF(2^4)$,$l_{max} = 5$ $\rho_{min}(n) = 44$	$n = r = 256$ $GF(2^4)$,$l_{max} = 5$ $\rho_{min}(n) = 60$
128 AES-128	–	$n = r = 128$ $GF(2)$,$l_{max} = 4$ $\rho_{min}(n) = 122$	$n = r = 192$ $GF(2^3)$,$l_{max} = 5$ $\rho_{min}(n) = 60$	$n = r = 256$ $GF(2^4)$,$l_{max} = 5$ $\rho_{min}(n) = 60$
192 AES-192	–	–	$n = 200, r = 192$ $GF(2)$,$l_{max} = 4$ $\rho_{min}(n) = 192$	$n = r = 256$ $GF(2^2)$,$l_{max} = 4$ $\rho_{min}(n) = 124$
256 AES-256	–	–	–	$n = 272, r = 256$ $GF(2)$,$l_{max} = 4$ $\rho_{min}(n) = 264$

Table 1: Values of parameters for MQ_DRBG

where $\left\langle \xi, \xi^2, \ldots, \xi^{2^{k-1}} \right\rangle$ is the normal basis of $\mathbb{F}_{2^k}$. Let us also define the map $\pi_i : \mathbb{F}_{2^k} \to \mathbb{F}_{2^k}$, $i = \overline{0, k-1}$:

$$\pi_i(v) = \xi^{2^i} v, \quad v \in \mathbb{F}_{2^k}.$$

The map τ_q is linear and surjective. It's linearity is obvious, and surjectivity follows from

$$\operatorname{rank} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = k.$$

More than that, π_i is a nonsingular linear map. Let $v = \tau_q(y) \in \mathbb{F}_{2^k}$, then

$$\pi_i(v) = (\alpha_1, y)\, \xi^{2^i+1} \oplus \ldots \oplus (\alpha_k, y)\, \xi^{2^i+2^{k-1}} = \left(\alpha_1^{(i)}, y\right) \xi \oplus \ldots \oplus \left(\alpha_k^{(i)}, y\right) \xi^{2^{k-1}}. \quad (3)$$

We have the following equation in the matrix form:

$$\begin{pmatrix} \left(\alpha_1^{(i)}, y\right) \\ \vdots \\ \left(\alpha_k^{(i)}, y\right) \end{pmatrix} = C \cdot \begin{pmatrix} (\alpha_1, y) \\ \vdots \\ (\alpha_k, y) \end{pmatrix} = C \cdot \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} \alpha_1^{(i)} \\ \vdots \\ \alpha_k^{(i)} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}, \quad (4)$$

where C is a nonsingular matrix of size $k \times k$ over $\mathbb{F}_2$, $\alpha_j^{(i)} \in V_k$, $y_j \in \{0, 1\}$, $j = \overline{1, k}$.

The vectors $\alpha_1^{(i)}, \ldots, \alpha_k^{(i)}$ are linearly independent over $\mathbb{F}_2, i = \overline{1,k}$.

The map $\pi_i(\tau_q(y)), i = \overline{0,k-1}$ defines the following quadratic function:

$$q'_i = x_1(\alpha_1^{(i)}, y) \oplus \ldots \oplus x_k(\alpha_k^{(i)}, y).$$

Due to the linear independence of the functions $x_1, \ldots, x_k, (\alpha_1^{(i)}, y), \ldots, (\alpha_k^{(i)}, y)$, the rank of q'_i is equal to $2k$. Consider the linear combination

$$\pi_{i_1}(\tau_q(y)) \oplus \ldots \oplus \pi_{i_d}(\tau_q(y)) = \xi^{2^{i_1}}\tau_q(y) \oplus \ldots \oplus \xi^{2^{i_d}}\tau_q(y) = (\xi^{2^{i_1}} \oplus \ldots \oplus \xi^{2^{i_d}})\tau_q(y).$$

This transformation is linear and nonsingular, because $\left\langle \xi, \ldots, \xi^{2^{k-1}} \right\rangle$ is the normal basis of $\mathbb{F}_{2^k}$.

The coefficients of element $\oplus_{j=1}^{d} \pi_{i_j}(\tau_q(y))$ in the normal basis expansion define the quadratic function

$$q'_{i_1 \ldots i_d} = q'_{i_1} \oplus \ldots \oplus q'_{i_d}.$$

The rank of $q'_{i_1 \ldots i_d}$ is equal to $2k$ due to the nonsingularity of the transformations (the proof of this fact is the same to the case $d = 1$ considered above). Thereby, all non-zero combinations $c_1 q'_1 \oplus \ldots \oplus c_k q'_k$ produce quadratic functions of the rank $2k$, $c_j \in \{0, 1\}$, $j = \overline{1,k}$.

We associate the quadratic form $q_{i_1} \oplus \ldots \oplus q_{i_d}$ with the following binary vector in V_k:

$$\begin{array}{ccccccccccc} (0\ldots0 & & 1 & & 0\ldots0 & & 1 & & 0\ldots0 & & 1 & & 0\ldots0) \\ & & i_1 & & & & i_2 & & & & i_d & & \end{array}$$

Consider linear code C of length k and rank m, such that $2^m - 1 \geq n + r$, and all linear combinations of at most 4 codewords are not equal to zero. If there exists the orthogonal $[k, k-m, d]$code $C^{\perp}$, where $d > 4$, we can construct the system of $2^m - 1$ quadratic functions, each of the following form:

$$x_1(\alpha_1^{(i)}, y) \oplus \ldots \oplus x_k(\alpha_k^{(i)}, y) \oplus (\beta, z) = \gamma,$$

where $\gamma \in \{0, 1\}$, $\beta, z \in V_t$ and coordinates of vectors $(x_1, \ldots, x_k)$, y, z are variables of the system.

Such systems could be solved in the following way. We assign variables $x_1, \ldots, x_k$ with all possible values (this will require 2^k operations). For each vector of values all quadratic functions became linear, except the case when all variables are assigned with zeros (the probability of that is 2^{-k} and negligible). As a result we get the linear system of k variables, which could be solved with complexity k^3, and restore the initial vector. The overall complexity is $2^k k^3$ binary operations.

For example, the complexity of such an approach for MQ_DRBG with parameters $n = r = 112$ (the corresponding linear code exists [2]) is $2^{53}(59)^3 \approx 2^{71}$ binary operations, instead of 2^{80} suggested by [5].

3.2 The meet-in-the-middle technique

In this section we are going to construct a system of multivariate quadratic functions such, that corresponding systems of equations could be solved with the meet-in-the-middle technique. In order to apply this technique, the output function should be equal to the sum of two functions of independent variables.

Let us choose s quadratic functions $h_i(z_1, \ldots, z_{n/2}), i = \overline{1, s}$ of $n/2$ variables, satisfying the AFNOR property with parameters l_{max} and $\rho_{min}(n/2) \geq \rho_{min}(n)/2$, and construct the system defining generator in the following way.

$$\begin{pmatrix} F(x) \\ G(x) \end{pmatrix} = \begin{pmatrix} h_1(x_1, \ldots, x_{n/2}) \oplus h_1(x_{n/2+1}, \ldots, x_n) \\ h_2(x_1, \ldots, x_{n/2}) \oplus h_2(x_{n/2+1}, \ldots, x_n) \\ \vdots \\ h_s(x_1, \ldots, x_{n/2}) \oplus h_s(x_{n/2+1}, \ldots, x_n) \\ f_{s+1}(x) \\ \vdots \\ f_{n+r}(x) \end{pmatrix}. \tag{5}$$

The first s functions in the system are the sums of the chosen functions of independent variables and satisfy the AFNOR property with parameters l_{max} and $\rho_{min}(n)$. The rest functions could be chosen at random with restrictions in order to satisfy the AFNOR property with parameters l_{max} and $\rho_{min}(n)$.

We have to assess the complexity of solving such a system. The first step consists of evaluating all the output vectors of function $F' : V_{n/2} \to V_s$

$$F'(x_1, \ldots, x_{n/2}) = \begin{pmatrix} h_1(x_1, \ldots, x_{n/2}) \\ h_2(x_1, \ldots, x_{n/2}) \\ \vdots \\ h_s(x_1, \ldots, x_{n/2}) \end{pmatrix},$$

which are the memory addresses, where we have to store the corresponding values of the variables. This step requires $2^{n/2}$ operations of evaluating of function $F'(x)$, and $(n/2) * 2^s * 2^{n/2-s} = (n/2) * 2^{n/2}$ bits of memory.

Suppose we know the output vector of MQ_DRBG, $y = (y_1, \ldots, y_s, y_{s+1}, \ldots, y_r)$. At the second step we have to evaluate vectors $(v_1, \ldots, v_s) = (y_1, \ldots, y_s) \oplus F'(x^i_{n/2+1}, \ldots, x^i_n)$ for all values of the variables $(x^i_{n/2+1}, \ldots, x^i_n), i = \overline{0, 2^{n/2} - 1}$. Further, we search through all values $(\hat{x}_1, \ldots, \hat{x}_{n/2})$ which are stored at address $(v_1, \ldots, v_s)$ and evaluate the output vector of function $F(\hat{x}_1, \ldots, \hat{x}_{n/2}, x^i_{n/2+1}, \ldots, x^i_n)$. In case the resulting value is equal to the output vector of MQ_DRBG, we found the initial value. The second step requires about $2^{n/2}2^{n/2-s}$ evaluations of function F, in case, the mapping defining MQ_DRBG is close to a random mapping.

As a result the whole attack requires $2^{n/2} + 2^{n/2}2^{n/2-s}$ evaluations of function F, or $2^{n/2} * s * 2n^2 + 2^{n/2}2^{n/2-s} * r * 2n^2$ binary operations, considering that the multiplication of a vector of length n by a $(n \times n)$-matrix costs about n^2 binary operations.

We conducted experiments, which showed that for several sets of parameters of MQ_DRBG suggested in [5], the proposed systems could be constructed by random choice.

For example, we constructed a system with $s = 95$ for parameters $n = 200,\ r = 192$, which could be solved with complexity at about 2^{129}, instead of 2^{192} suggested by [5].

Note, that one could also use the technique developed in the previous section to construct described systems for arbitrary s.

4 Conclusion

The pseudorandom number bit generator proposed by French specialists is based on the (pseudo)random system of multivariate quadratic equations with several properties defined in [5]. The security of the generator is based on the well-known MQ-problem. Since there is no method for generating such pseudo(random) systems described in [5], we managed to propose to different techniques for constructing such systems of equations which could be solved with less complexity than stated in [5]:

- The first technique is based on guessing the part of the variables, which leads to the linearization. The possibility of constructing such systems is based on the existence of a linear code with given parameters. For several parameters of MQ_DRBG such codes are known to exist.
- The second technique exploits the meet-in-the-middle approach. For several parameters of MQ_DRBG one can construct such systems by random choice.
- One can use the combination of both techniques.

These examples show that restrictions for multivariate quadratic systems proposed in [5] could not guarantee the corresponding security level stated in [5]. If there exists a malicious developer, who could force the legitimate user to use "weak" systems, the attack, which exploits such systems could be effective. The systems could be hidden with the Isomorphism of Polynomials.

As a result, we have to state that the main concern is the fact that the authors used, first, the provable security approach while assessing the security of MQ_DRBG and, second, the estimation of the complexity of solving multivariate quadratic equations with universal methods. We believe that this could cause the underestimation while assessing the security of a specific system of multivariate quadratic equations used as MQ_DRBG. We have to mention that the security of standardized primitives is based mainly on the analysis of the specific properties of cryptographic algorithms used within the primitive.

References

[1] Logachev O.A., Salnikov A.A., Yashenko V.V.: *Boolean functions in coding theory and cryptology* – Moscow.: MCNMO, 2004. (in russian)

[2] MacWilliams F., Sloane N.: *The theory of error-correcting codes* Amsterdam, The Netherlands : North Holland, 1977.

[3] Paillier P.: *A comprehensive security analysis of* MQ_DRBG. *Technical report TR-017.*

[4] MQ_DRBG. *An AFNOR proposal for ISO/IEC 18031* SC27 WG2 Meeting. Berlin Oct 4-8 2010.

[5] ISO/IEC JTC 1/SC 27 No 9808. Text for ISO/IEC FDIS 18031 – Information technology – Security techniques – Random Bit Generation.

[6] Berbain C., Gilbert H. Patarin J.: *QUAD: A practical stream cipher with provable security* // EUROCRYPT'06, vol. 4004 LNCS, 109-128.

[7] ISO/IEC 18031:2005 – Information technology – Security techniques – Random Bit Generation.

[8] AIS20 Version 1. Functionality classes and evaluation methodology for deterministic random number generators. BSI. 1999.

[9] Yang B., Chen O., Bernstein D., Chen J.: *Analysis of QUAD* // FSE 2007, vol. 4593 LNCS, pp. 290 – 308, Springer, 2007.

[10] Matsumoto T., Imai H.: *Public quadratic polynomial-tuples for efficient signature-verification and message-encryption*, In. Advances in Cryptology – EUROCRYPT'88. Volume 330 in LNCS. Springer-Verlag (1988) 419-453.

[11] Patarin J.: *Hidden Field Equations (HFE) and Isomorphism of Polynomials (IP): Two new families of asymmetric algorythms.* In EUROCRYPT'96.

[12] Patarin J., Courtois N., Goubin L.: *Flash, a fast multivariate signature algorithm.* In Proceedings CT-RSA 2001, vol. 2020 of LNCS, Springer, 2001.

[13] Patarin J., Courtois N., Goubin L.: *QUARTZ, 128-bit long digital signatures.* In Proceedings CT-RSA 2001, vol. 2020 of LNCS, Springer, 2001.

[14] Perret L.: *A fast cryptanalysis of the isomorphism of polynomials with one secret problem.* In Proceedings EUROCRYPT'98.

[15] Bouillaguet C., Faugere J.-C.: *Practical cryptanalysis of the identification.* In Proceedings PKC'11.

[16] Faugere J.-C., Perret L.: *Polynomial equivalence problem: Algorithmic and theoretical aspects.* In Proceedings EUROCRYPT'06.