On some algebraic aspects of data security in cloud computing

Vasyl Ustimenko, Aneta Wroblewska
Maria Curie-Sklodowska University in Lublin (Poland)

awroblewska@hektor.umcs.lublin.pl

Abstract

The paper is dedicated to ideas of homomorphic encryption and multivariate key dependent cryptography. We observe recent theoretical results on the above-mentioned topics together with their applications to cloud security. Post Quantum Cryptography could not use many security tools based on Number Theory, because of the factorization algorithm developed by Peter Shor. This fact and fast development of Computer Algebra make multivariate cryptography an important direction of research. The idea of key dependent cryptography looks promising for applications in Clouds, because the size of the key allows to control the speed of execution and security level. Finally, special classes of finite rings turned out to be very useful in homomorphic encryption and for the development of multivariate key.

Cloud computing provides clients with a virtual computing infrastructure on top of which they can store data and run applications. While the benefits of cloud computing are clear, it introduces new security challenges since cloud operators are expected to manipulate client data without necessarily being fully trusted. We are designing cryptographic primitives and protocols tailored to the setting of cloud computing, attempting to strike a balance between security, efficiency and functionality. The current generation of cloud storage services do not provide any security against untrusted cloud operators making them unsuitable for storing sensitive information such as medical records, financial records or high impact business data. To address this we are pursuing various research projects that range from theory to practice.

Homomorphic encryption. The most common use of encryption is to provide confidentiality by hiding all useful information about the plaintext. Encryption, however, renders data useless in the sense that one loses the ability to operate on it. To address this we are designing cryptosystems that support a variety of computations on encrypted data, ranging from general-purpose computations (i.e., fully-homomorphic encryption) to special-purpose computations (e.g., voting and search).

Searchable structured encryption. A searchable encryption scheme encrypts data in such a way that a token can be generated to allow a third party to search over the encrypted data. Using a searchable encryption scheme, a client can safely store its data with an untrusted cloud provider without losing the ability to search over it. There is a need of structured encryption which allows a client to encrypt various types of data (e.g., social networks or web graphs) in such a way that complex queries can be performed over the encrypted data. Structured encryption and various constructions for graph data is known.

Some security issues raised by cloud computing are motivated by virtualization. Dynamic scalability or elasticity will help generalize high-performance computing and very large data sets in applications. But the real gains in performance depend heavily on the predictability of physical and virtualized resources. It means that the balancing of performance against security and the adaptation of HPC or VLDB techniques to cloud computing are important issues and will have long-lasting scientific content. The direction of Key Dependent Message (KDM) secure encryption in Cryptography can bring an appropriate security tools for Cloud Computing.

The goal of the presented paper is discussion of new KDM cryptosystems, which have some potential to be used in the era of Postquantum Cryptography. The Quantum Computer is a special random computational machine. Recall that computation in Turing machine can be formalised with the concept of finite automaton as a walk in the graph with arrows labelled by special symbols. "Random computation" can be defined as a random walk in the random graph. So we are looking for the deterministic approximation of random graphs by extremal algebraic graphs. It is known that the explicit solutions for an optimization graphs have properties similar to random graphs. The probability of having rather short cycle in the walking process on random graph is zero. So the special direction of Extremal Graph Theory

of studies of graphs of order v (the variable) without short cycles of maximal size (number of edges) can lead to the discovery of good approximation for random graphs.

Keywords
multivariate cryptography, cloud computing, symbolic computations, graphs of large girth

1 Introduction

The plainspace of the algorithm is K^n, where K is the chosen commutative ring. Graph theoretical encryption corresponds to walk on the bipartite graph with partition sets which are isomorphic to K^n. We conjugate chosen graph based encryption map, which is a composition of several elementary polynomial automorphisms of a free module K^n with special invertible affine transformation of K^n. Finally we compute symbolically the corresponding polynomial map g of K^n onto K^n. We say that the sequence g_n, $n \geq 3$, $n \to \infty$ of polynomial transformation bijective maps of free module K^n over commutative ring K is a sequence of stable degree if the order of g_n is growing with n and the degree of each nonidentical polynomial map of kind ${g_n}^k$ is an independent constant c. A transformation $b = \tau {g_n}^k \tau^{-1}$, where τ is affine bijection, n is large and k is relatively small, can be used as a base of group theoretical Diffie-Hellman key exchange algorithm for the Cremona group $C(K^n)$ of all regular automorphisms of K^n. The specific feature of this method is that the order of the base may be unknown for the adversary because of the complexity of its computation. The exchange can be implemented by tools of Computer Algebra (symbolic computations). The adversary can not use the degree of righthandside in $b^x = d$ to evaluate unknown x in this form for the discrete logarithm problem.

In the paper we introduce the explicit constructions of sequences of elements of stable degree c for each commutative ring K containing at least 3 elements and each $c \geq 2$. Special cases of $c = 3$ and $c = 2$ were obtained in [11] and [10]. We discuss the implementation of related key exchange and public key algorithms. It is interesting that in the case of $c \geq 4$ use of special affine bijections lead to sparse polynomial transformation with $O(n^3)$ monomial expressions. Those results are based on the construction of the family $D(n, q)$ of graphs with large girth and the description of their connected components $CD(n, q)$. The existence of infinite families of graphs of large girth had been proven by Paul Erdös' (see [1]). Together with famous Ramanujan graphs introduced by G. Margulis [4] and investigated in [3] graphs $CD(n, q)$ is one of the first explicit constructions of such a families with unbounded degree. Graphs $D(n, q)$ had been used for the construction of LDPS codes and turbocodes which were used in real satellite communications ([2]), for the development of private key encryption algorithms ([9], [5]), the option to use them for public key cryptography was considered in [8], [7] and in [6], where the related dynamical system had been introduced.

2 Preliminaries

Let $\mathbb{K}$ denote commutative ring.

Set Q of the ring $\mathbb{K}$ is **the multiplicative set** of ring $\mathbb{K}$, if it is closed under operation of multiplication ($x, y \in Q \Rightarrow x \cdot y \in Q$) and does not contain 0.

Elements $t_1, t_2, \ldots, t_l$, $l \geq 1$ z $\mathbb{K}$ are called **multiplicative generators**, if there is a multiplicative set Q containing all t_i, $i = 1, 2, \ldots, l$.

2.1 Graphs and incidence system

The missing definitions of graph-theoretical concepts which appear in this paper can be found in [1]. All graphs we consider are simple, i.e. undirected without loops and multiple edges. Let $V(G)$ and $E(G)$ denote the set of vertices and the set of edges of G, respectively. Then $|V(G)|$ is called the *order* of G, and $|E(G)|$ is called the *size* of G. A path in G is called *simple* if all its vertices are distinct. When it is convenient, we shall identify G with the corresponding anti-reflexive binary relation on $V(G)$, i.e. $E(G)$ is a subset of $V(G) \times V(G)$ and write vGu for the adjacent vertices u and v (or neighbors). The sequence of distinct vertices $v_1, \ldots, v_t$, such that $v_i G v_{i+1}$ for $i = 1, \ldots, t-1$ is the pass in the graph. The length of a pass is a number of its edges. The distance $\mathrm{dist}(u, v)$ between two vertices is the length of the shortest pass between them. The diameter of the graph is the maximal distance between two vertices u and v of the graph. Let C_m denote the cycle of length m i.e. the sequence of distinct vertices $v_1, \ldots, v_m$ such that $v_i G v_{i+1}$,

$i = 1, \dots, m-1$ and $v_m G v_1$. The girth of a graph G, denoted by $g = g(G)$, is the length of the shortest cycle in G. The degree of vertex v is the number of its neighbors (see [15] or [1]).

The incidence structure is the set V with partition sets P (points) and L (lines) and symmetric binary relation I such that the incidence of two elements implies that one of them is a point and another is a line. We shall identify I with the simple graph of this incidence relation (bipartite graph). If number of neighbours of each element is finite and depends only on its type (point or line), then the incidence structure is a tactical configuration in the sense of Moore (see [12]). The graph is k-regular if each of its vertex has degree k, where k is a constant. In this section we reformulate results of [13], [14] where the q-regular tree was described in terms of equations over finite field F_q.

Let q be a prime power, and let P and L be two countably infinite dimensional vector spaces over F_q. Elements of P will be called *points* and those of L *lines*. To distinguish points from lines we use parentheses and brackets: If $x \in V$, then $(x) \in P$ and $[x] \in L$. It will also be advantageous to adopt the notation for coordinates of points and lines introduced in [4]:

$$(p) = (p_1, p_{11}, p_{12}, p_{21}, p_{22}, p'_{22}, p_{23}, \dots, p_{ii}, p'_{ii}, p_{i,i+1}, p_{i+1,i}, \dots),$$
$$[l] = [l_1, l_{11}, l_{12}, l_{21}, l_{22}, l'_{22}, l_{23}, \dots, l_{ii}, l'_{ii}, l_{i,i+1}, l_{i+1,i}, \dots).$$

We now define an incidence structure (P, L, I) as follows. We say the point (p) is incident with the line $[l]$, and we write $(p)I[l]$, if the following relations between their coordinates hold:

$$\begin{aligned} l_{11} - p_{11} &= l_1 p_1 \\ l_{12} - p_{12} &= l_{11} p_1 \\ l_{21} - p_{21} &= l_1 p_{11} \\ l_{ii} - p_{ii} &= l_1 p_{i-1,i} \\ l'_{ii} - p'_{ii} &= l_{i,i-1} p_1 \\ l_{i,i+1} - p_{i,i+1} &= l_{ii} p_1 \\ l_{i+1,i} - p_{i+1,i} &= l_1 p'_{ii} \end{aligned} \tag{1}$$

(The last four relations in (1) are defined for $i \geq 2$.) This incidence structure (P, L, I) we denote as $D(q)$. We speak now of the *incidence graph* of (P, L, I), which has the vertex set $P \cup L$ and edge set consisting of all pairs $\{(p), [l]\}$ for which $(p)I[l]$.

2.2 Connected components

Let us consider the description of connected components of the graphs.

Let $n \geq 6$, $t = \lfloor (n+2)/4 \rfloor$, and let $u = (u_1, u_{11}, \cdots, u_{tt}, u'_{tt}, u_{t,t+1}, u_{t+1,t}, \cdots)$ be a vertex of $D(n, \mathbb{K})$. (It does not matter whether u is a point or a line). For every r, $2 \leq r \leq t$, let

$$a_r = a_r(u) = \sum_{i=0}^{r} (u_{ii} u'_{r-i,r-i} - u_{i,i+1} u_{r-i,r-i-1}) \tag{2}$$

,

and $a = a(u) = (a_2, a_3, \cdots, a_t)$. (Here we define

$p_{-1,0} = l_{0,-1} = p_{1,0} = l_{0,1} = 0$, $p_{00} = l_{00} = -1$, $p_{0,1} = p_1$, $l_{1,0} = l_1$, $p'_{00} = l'_{00} = 1$ $l'_{11} = l_{11}$, $p'_{1,1} = p_{1,1}$).

In [13] the following statement was proved.

Proposition 1 *Let u and v be vertices from the same component of $D(k, q)$. Then $a(u) = a(v)$. Moreover, for any $t-1$ field elements $x_i \in F_q$, $2 \leq t \leq [(k+2)/4]$, there exists a vertex v of $D(k, q)$ for which*

$a(v) = (x_2, \dots, x_t) = (x)$.

Corollary 1 *Let us consider a general vertex*

$$x = (x_1, x_{1,1}, x_{2,1}, x_{1,2} \cdots, x_{i,i}, x'_{i,i}, x_{i+1,i}, x_{i,i+1}, \cdots),$$

$i = 2, 3, \cdots$ of the connected component $CD(n, \mathbb{K})$, which contains a chosen vertex v. Then, coordinates $x_{i,i}$, $x_{i,i+1}$, $x_{i+1,i}$ can be chosen independently as "free parameters" from $\mathbb{K}$ and $x'_{i,i}$ could be computed successively as the unique solution of the equations $a_i(x) = a_i(v)$, $i = 2, 3, \dots$.

3 Operators L_{D,n,β_k} and P_{D,n,α_k}

Let L_{D,n,β_k} be the operator of taking the neighbour of point:

$$(p)^{2k-2} = (p_{0,1}, p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}, p'_{2,2}, p_{2,3}, \ldots, p_{i,i}, p'_{i,i}, p_{i,i+1}, p_{i+1,i}, \ldots),$$

of a kind

$$[l]^{2k-1} = [\beta_k, l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2}, l'_{2,2}, l_{2,3}, \ldots, l_{i,i}, l'_{i,i}, l_{i,i+1}, l_{i+1,i}, \ldots],$$

where parameters $l_{1,1}, l_{1,2}, l_{1,2}, l_{2,2}, \ldots, l_{i,i}, l_{i,i+1}, l_{i+1,i}, \ldots$ are computed consequently from the equations (1) in definition of $D(n, \mathbb{K})$ and all $l'_{i,i}$ for $i = 2, 3, \ldots$ are computed using equation describing connected component (2).

Similarly, P_{D,n,α_k} is the operator of taking the neighbour of line

$$[l]^{2k-1} = [l_{1,0}, l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2}, \ldots, l_{i,i}, l_{i,i+1}, l'_{i,i}, l_{i+1,i}, \ldots],$$

of a kind

$$(p)^{2k} = (p_{0,1}^{2k-2} + \alpha_k, p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}, \ldots, p_{i,i}, p'_{i,i}, p_{i,i+1}, p_{i+1,i}, \ldots),$$

where parameters $p_{1,1}$, $p_{1,2}$, $p_{2,1}$, $p_{2,2}$,$\ldots$, $p_{i,i}$, $p_{i,i+1}$, $p_{i+1,i}$, $\ldots$ are computed consequently from the equations (1) in definition of $D(n, \mathbb{K})$ and all $p'_{i,i}$ for $i = 2, 3, \ldots$ are computed using equation describing connected component (2).

Given the vector $(p)^0 = (p_{0,1}, p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}, p'_{2,2}, p_{2,3}, \ldots, p_{i,i}, p'_{i,i}, p_{i,i+1}, p_{i+1,i}, \ldots)$, (of length n) let us take elements $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ from Q^k and composition $F_{n,\alpha,\beta} = L_{D,n,\beta_1} P_{D,n,\alpha_1} L_{D,n,\beta_2} P_{D,n,\alpha_2} \ldots L_{D,n,\beta_k} P_{D,n,\alpha_k}$.

Theorem 1 *(A. Wroblewska) Independently from the choice of $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \in Q^k$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_k) \in Q^k$, the map $F_{n,\alpha,\beta}$ of free module $\mathbb{K}^{n - \lfloor \frac{n+2}{4} \rfloor}$ is bijective map with degree $\lfloor \frac{n+2}{4} \rfloor$.*

Theorem 2 *(V. Ustimenko) The order $F_{n,\alpha,\beta}$ is going to ∞ when $n \to \infty$*

4 Application

4.1 Public key algorithm

Let τ be linear transformation $\tau : x \to Ax$, where A is sparse matrix with condition $\det A \neq 0$ Map $\tau F_{n,\alpha,\beta} \tau^{-1}$ written as a multivariate public rule:

$$x_1 \to h_1(x_1, x_2, \ldots, x_n)$$

$$x_2 \to h_2(x_1, x_2, \ldots, x_n)$$

$$\ldots$$

$$x_n \to h_n(x_1, x_2, \ldots, x_n),$$

can be used in public key cryptography. Alice - the holder of the key - keeps linear transformation and $(\beta_1, \alpha_1, \beta_2, \alpha_2, \ldots, \beta_k, \alpha_k)$ secret. Bob (public user) has the above map.

Combining the transformation $F_{n,\alpha,\beta}$ with two linear transformation, Bob get a formula:

$$y = (h_1(x_1, \ldots, x_n), \ldots, h_n(x_1, \ldots, x_n)),$$

where $h_i(x_1, \ldots, x_n)$ are polynomials of n variables of degree $\lfloor \frac{n+2}{4} \rfloor$. Hence the process of straightforward encryption can be done in polynomial time $O(n^6)$. But the cryptanalyst Catherine, having a only a formula for y, has very hard task to solve the system of n equations in n variables of degree $\lfloor \frac{n+2}{4} \rfloor$. So the general algorithm for finding the solution of system of polynomials equations has exponential time $(\lambda n)^{O(n)}$.

4.2 Diffie-Hellman key exchange protocol

We consider Diffie-Hellman algorithm for $C(K^n)$ for the key exchange in the case of group. Let $AGL_n(F_q)$ be the group of affine transformation of the vector space F_q^n, i.e. maps $\tau_{A,b}$: $\widetilde{x} \to \widetilde{x}A + b$, where $\widetilde{x} = (x_1, x_2, \ldots, x_n)$, $b = (b_1, b_2, \ldots, b_n)$ and A is invertible sparse matrix with $\det A \neq 0$. Let h_n^k be the new public rule obtained via k iterations of $h_n = F_{n,\alpha,\beta} = L_{D,n,\beta_1} P_{D,n,\alpha_1} L_{D,n,\beta_2} P_{D,n,\alpha_2} \ldots L_{D,n,\beta_k} P_{D,n,\alpha_k}$. Correspondents Alice and Bob have different information for making computation. Alice chooses dimension n, element h_n as above, affine transformation $\tau \in AGL_n(K)$. So she obtains the base $b = \tau h_n^k \tau^{-1}$ and sends it in the form of standard polynomial map to Bob.

So Alice chooses rather large number n_A computes $c_A = b^{n_A}$ and sends it to Bob. On his turn Bob chooses his own key n_B and computes $c_B = b^{n_B}$. He and Alice get the collision map c as ${c_A}^{n_B}$ and ${c_B}^{n_A}$ respectively.

Notice that the position of adversary is similar to Bob's position. He (or she) need to solve one of the equations $b^x = c_B$ or $b^x = c_A$. The algorithm is implemented in the cases of finite fields and rings Z_m for family of groups $C(K^n)$.

References

[1] B. Bollobás, *Extremal Graph Theory*, Academic Press, London, 1978.

[2] Jon-Lark Kim, U. N. Peled, I. Perepelitsa, V. Pless, S. Friedland, *Explicit construction of families of LDPC codes with no 4-cycles* , Information Theory, IEEE Transactions, 2004, v. 50, Issue 10, 2378 - 2388.

[3] A. Lubotsky, R. Philips, P. Sarnak, *Ramanujan graphs*, J. Comb. Theory., 115, N 2., (1989), 62-89.

[4] G. A. Margulis, *Explicit construction of graphs without short cycles and low density codes*, Combinatorica, 2, (1982), 71-78.

[5] V. A. Ustimenko, *Graphs with Special Arcs and Cryptography*, Acta Applicandae Mathematicae, 2002, vol. 74, N2, 117-153.

[6] V. A. Ustimenko, *Linguistic Dynamical Systems, Graphs of Large Girth and Cryptography*, Journal of Mathematical Sciences, Springer, vol.140, N3 (2007) pp. 412-434.

[7] V. A. Ustimenko, *On the graph based cryptography and symbolic computations*, Serdica Journal of Computing, Proceedings of International Conference on Application of Computer Algebra, ACA-2006, Varna, N1 (2007).

[8] V. A. Ustimenko, *Maximality of affine group, and hidden graph cryptosystems*, J. Algebra and Discrete Math., 10 (October 2004), 51-65.

[9] V. A. Ustimenko, *CRYPTIM: Graphs as Tools for Symmetric Encryption*, in Lecture Notes in Computer Science, Springer, 2001, v. 2227, 278-287.

[10] V. A. Ustimenko, A. Wrblewska, *Dynamical systems as the main instrument for the constructions of new quadratic families and their usage in cryptography*, Annales UMCS Informatica AI, ISSN 1732-1360.

[11] A. Wroblewska *On some properties of graph based public keys*, Albanian Journal of Mathematics, Volume 2, Number 3, 2008, 229-234, NATO Advanced Studies Institute: "New challenges in digital communications".

[12] E. H. Moore, *Tactical Memoranda*, Amer. J. Math., v.18, 1886, 264-303.

[13] F. Lazebnik, V. A. Ustimenko, A. J. Woldar, *A Characterization of the Components of the graphs $D(k,q)$*, Discrete Mathematics, 157 (1996) 271–283.

[14] F. Lazebnik F. and V. Ustimenko, *Explicit construction of graphs with an arbitrary large girth and of large size*, Discrete Appl. Math. , 60, (1995), 275 - 284.

[15] N.L. Biggs, *Graphs with large girth*, Ars Combinatoria, 25C (1988), 73–80.