S. Pac. J. Nat. Sci., (2002), Vol. 20, pp. 34 - 44
Section B: Featured Article

Walks on Graphs as Symmetric or Asymmetric Tools to Encrypt Data

[1]Vasiliy A. Ustimenko and [2]Youry M. Khmelevsky
[1]*Department of Mathematics, Sultan Qaboos University, Sultanate of Oman*
[2]*Department of Mathematics and Computing Science, University of the South Pacific, Suva, FIJI,*
E-mail: vasyl@squ.edu.om, khmelevsky_y@usp.ac.fj

Abstract

New results on graph theoretical method of encryption will be presented. The general idea is to treat vertices of a graph as messages, and walks of a certain length as encryption tools. We will construct one-time pad algorithms with a certain resistance to attacks when the adversary knows plaintext *and* ciphertext. *Special linguistic graphs of high girth whose vertices (messages) and walks (encoding tools) could be both naturally identified with vectors over the finite field, and with the so-called parallelotopic graphs, which turn out to be efficient tools for* symmetric *encryption. We will formulate criteria when parallelotopic graph (or the more general graph of tactical configuration) is a graph of absolutely optimal encryption scheme, producing asymptotic one-time pad algorithm. We will show how to convert one-time pads, which are related to geometries of rank 2 of simple groups of Lie type, to a real-life encryption scheme involving potentially infinite text and flexible passwords.*

We will discuss families of linguistic and parallelotopic graphs of increasing girth as the source for the generation of asymmetric cryptographic functions and related open key algorithms. *We will construct new families of such graphs via group theoretical and geometrical technique.*

The software for symmetric and asymmetric encryption (prototype model of the package) is ready for demonstration.

Keywords: cryptography, symmetric encryption, public key, digital signatures, virtual campus, networks, e-commerce

I. GENERAL INTRODUCTION

Security of data is one of the most important problems in Applied Informatics. The current security market is not satisfactory to fulfill the demand. In particular, it is important for geographically remote countries of the Pacific region to have safe and secure communication lines to overcome “tyranny of distance”. Security of data and communication processes need encryption - invertible procedure for converting of initial data which is a certain string of characters (the plaintext) to unreadable, chaotic like string (the ciphertext).

The goal of the paper is to present new security algorithms based on the idea to look at what kind of finite automaton (roughly, graph) we need for encryption. An assumption first codified by Kerckhoff (one may recall Kerckhoff’s rules in Physics) is that the encryption algorithm is known and security rests entirely on the security of key (password).

Proposed iterative nonlinear algorithms are convenient tools of symmetric encryption, when procedure and its inverse take same time. They have flexible size of keys, procedure is robust and compares well with the performance of some existing algorithms. If double size of key is less then length of the plaintext, different keys convert it into different ciphertext. A prototype model of software package is ready for the demonstration. Experiment support some theoretical conjecture on the resistance of proposed encryption to attacks when adversary already knows several plaintext - ciphertext pairs anf trying to get a password to control the channel of communication.

The procedure can be given by close formula which is a certain multivariable polynomial. The use of close formula instead of iteration makes the procedure assymetric, decryption is practically much harder then encryption. This idea requires further studies: it may be used as so called public keys.

II. INTRODUCTION

Security of data is one of the most important issues to consider in the daily operation of a modern university. Confidential data of different format (text, image, sound) need to be transmitted securely to and from different sections, departments and schools or faculties of the university. This issue becomes more pressing if the university operates a distance mode teaching.

Research in security of data is very important and urgently needed for developing countries. The main reason is that current security software is usually not available for use in these countries, as it is only recent that international cooperation in the area of network security is becoming possible because certain legal and political barriers have been lifted recently. For example, in September 16, 1999, the US Government relaxed ban for the export of encryption technology. Thus, such software will have to be developed locally by university researchers and computer system developers to cater for the local needs. This paper reports a first attempt at creating a local security software.

The work reported in this paper was initially motivated by security concerns on transmitting data across the University of the South Pacific (USP) intranet, called USPNet (http://www.usp.ac.fj), designed and implemented as an e-commerce application to cater for a distance mode education and related administrative work. An initial development was that of an algorithm for text and image encryption, which is fast and (like RSA) has strong resistance to attacks by adversaries who may utilized image data and ciphertext. The algorithm was used at USP as a private key algorithm. The algorithm is based on the method which explores the possibilities of a straightforward approach to look at what kind of finite automaton (roughly, graph) is needed for encryption. It turns out that graphs of large girth are effective tools for the encryption of images.

Since 2001, a team of researchers and software developers from USP, Sultan Quaboos University (SQU, Oman) and University of Kyiv-Mohyla Academy have been working on the development of a new version of the original USP system. The new package contains modified private key algorithms for the encryption not only of txt files but other text and image data in other format - special programs to work with the html, pdf, gif files as well as universal program to work with any data in WINDOW's 9x/2000 alphabet. The algorithms have variable sizes of keys, are robust and compare well with the performance of some existing algorithms. The most important innovation is an algorithm where special families of graphs will be used for public keys, as well as for electronic signatures. A prototype model of the software package is ready for demonstration.

This paper presents a first description of graph theoretical approach used in the development of public key algorithms and some new results on symmetric encryption.

III. TWO TYPES OF ATTACKS

Assume that an unencrypted message, *plaintext*, which can be image data, is a string of bits. It is to be transformed into an encrypted string or *ciphertext*, by means of a cryptographic algorithm and a *key*. In order for the recipient to read the message, encryption must be *invertible*.

Conventional wisdom holds that in order to defy easy decryption, a cryptographic algorithm should produce a seemingly chaotic pattern: that is, the ciphertext should look chaotic. In theory, an eavesdropper should not be able to determine any significant information from an intercepted ciphertext. Broadly speaking, attacks to a cryptosystem fall into 2 categories: *passive attacks*, in which adversary monitors the communication channel and *active attacks*, in which the adversary may transmit messages to obtain information (e.g. ciphertext of chosen plaintext).

Passive attacks are easier to mount, but yields less. Attackers hope to determine the plaintext from the ciphertext they capture; an even more successful attack will determine the key and thus compromise the whole set of messages.

An assumption first codified by Kerckhoffs in the nineteenth century is that the algorithm is known and the security of algorithm rests entirely on the security of the key.

Cryptographers have been improving their algorithms to resist the following list of increasingly aggressive attacks:

i) *ciphertext only* – the adversary has access to the encrypted communications;

ii) *known plaintext* – the adversary has some plaintext and its corresponding ciphertext.

IV. BASIC APPROACH

One of the classical models of the procedure for encoding data is to present the information to be sent as a variety of n-tuples over the finite Galois field $GF(q)$. We have to "encode" our message x by taking an affine transformation $y = Ax + b$, where A is a certain matrix and b is another n-tuple.

Our proposal is based on the combinatorial method of construction of linear and nonlinear codes, which has a certain similarity with the classical scheme above. It is different from graph theoretical tools used in [5], [9].

Let Γ be a k-regular graph and $V(\Gamma)$ is the set of its vertices. Let us refer to the sequence $\rho = (v_1, v_2, \cdots, v_n)$, where $v_i \in V(\Gamma)$, $v_i \neq v_{i+2}$, $i = 1, \cdots, v_{n-2}$, and $v_i \Gamma v_{i+1}$, $i = 1, \cdots, n-1$, and $v_\rho = v_n$ as *encoding sequence* and *encoded vertex* of $v = v_1$. Clearly for $u = v_\rho$ there is sequence μ of length s such that $u_\mu = v$. We refer to μ as decoding sequence for v_ρ and write $\mu = \rho^{-1}$.

In the case of vertex transitive graphs set of all *encoding sequences* of certain length starting from the chosen vertex v_0 may be considered as the set of possible keys. To apply the key μ from this set to the vertex v means taking the last vertex of walk μ^g where g is the graph automorphism moving v_0 to v. In case of *parallelotopic graphs* defined below there exists a combinatorial way of description keys in a uniform way, which does not depend on starting vertex (or message).

The girth $g = g(\Gamma)$ of a graph Γ is the length of the shortest cycle in the graph.

If the length of the encoding sequence ρ of the k-regular graph Γ of girth $g = g(\Gamma)$ is less then g, then $v_\rho \neq v$ for any vertex v.

If one knows the length $t \leq g/2$ of the decoding sequence the probability of generating the correct message applying the encoding sequence at random is $1/(k(k-$

$1)^{t-1})$. In this case the algorithm is $k(k-1)^t$ secure. We will use the term *graph encryption scheme* for the pair (Γ, t).

V. ON GRAPH THEORETICAL ONE-TIME PADS

The revolutionary classical result on private key algorithm was obtained by C. Shannon at the end of the 1940s (see [23]). He constructed the so-called one-time pads, whose keys and strings of random bits at least as long as a message itself, achieved the seemingly impossible: an eavesdropper was not able to determine any significant information from an intersected ciphertext. The simplest classical example is as follows: if p_i is the i-th bit of the plaintext, k_i is the i-th bit of the key, and c_i is the first bit of the ciphertext, then $c_i = p_i + k_i$, where $+$ is exclusive or, often written XOR, and is simply addition modulo 2. One-time pads must be used exactly once: if a key is ever reused, the system becomes highly vulnerable.

It is clear that the encryption scheme above cannot resist attacks of type (ii) - one simply subtracts p_i from c_i and gets the key. The construction of Shannon's one-time pad is theoretically sound. In practice, however, we have a size of information which is exponentially larger than the size of the key. Also, a one-time usage of the password is impossible because it makes communication expensive. Nevertheless construction of new one-time pad is important to show the possibility of the chosen method of encryption.

Theoretically, if we have a family of one-time pad $p(n)$ for the encryption of the text of the length n, we can use it safely $c(n)$ times, where $c(n) \geq 1$ is bounded by a constant independent from n, according to Shannon's result. Such an encryption schemes may have even a resistance to attacks with known plaintext and ciphertext (type (ii)), but again the complexity of breaking key is bounded above by some constant.

First question: Can we find a graph of girth g such that for an encryption bypass of length $< [g/2]$, is the graph a one-time pad?

Examples of such graphs can give us an idea on what objects are good tools for encryptions in practical situations where the size of the key is essentially smaller than the size of the plaintext. Besides, one-time pads can be used as blocks in real life encryption algorithms.

Let us consider our encryption bypasses of length $t \leq [g/2]$ for the graph Γ of girth g.

If Γ is a *one-time pad* then the ratio $p_{\text{key}}(i)/p_{\text{mes}}(i)$ of probabilities $p_{\text{key}}(i)$ and $p_{\text{mes}}(i)$ to guess the encoding sequence and to guess the message (plaintext) in the scheme (Γ_i, t_i), respectively, equals 1.

We will introduce below one-time pads with a certain resistance to an attack of type (ii), which are bipartite graphs. These are graphs related to tactical configurations. A *tactical configuration*, so-termed by E. H. Moore nearly a century ago, is a rank two incidence structure $\Delta = \Delta(l, p, a, b)$ consisting of l lines and p points in which each line is incident to a points and each point is incident to b lines. We denote the incidence graph of Δ by $\Gamma = \Gamma(\Delta)$, though when no confusion arise, we may abuse terminology and refer to Γ as a tactical configuration as well. The incidence graphs of incidence structures are called bipartite graphs. If structure has a tactical configuration, then the incidence graphs are called biregular with bidegree a, b.

Graph $\Gamma(\Delta)$ has order $v = l + p$ (number of vertices), and size $e = la = pb$ (number of edges). As usual, the girth of the graph is the length of its minimal cycle.

We will prove the following statement:

Lemma 1 *If the tactical configuration with bidegrees $r+1$ and $s+1$ and parameters P=p, l=l has girth $g \geq 2k+2$, $E = p(r+1) = l(s+1)$, then the following inequalities hold:*

1) If $k = 2t+1$, then

$$\left.\begin{array}{l} 1 + r + rs + r^2 s + r^2 s^2 + \cdots + r^{t+1} s^t \leq p, \\ 1 + s + sr + s^2 r + s^2 r^2 + \cdots + s^{t+1} r^t \leq l \end{array}\right\} \qquad (1)$$

2) If $k = 2t$, then

$$\left.\begin{array}{l} 1 + r + rs + r^2 s + r^2 s^2 + \cdots + r^t s^t \leq p, \\ 1 + s + sr + s^2 r + r^2 s^2 + \cdots + s^t r^t \leq l \end{array}\right\} \qquad (2)$$

Proof: Let us consider a chosen point P. The pass of length $h \leq k$ between two chosen vertices is unique. Thus counting of vertices at distance h can be done by the branching process. Thus, we have $l_1 = r+1$ lines at distance 1 from P, $p_1 = (r+1)s$ is the number of points at distance 2 from $P \ldots$, $l_3 = (r+1)rs$ is the number of points at distance 3 from P. Let $k = 2t+1$. Then

$$l_{2h+1} = (r+1)r^h s^h \text{ and } p_{2h+2} = (r+1)r^h s^{h+1},$$

where $h = 0, 1, \ldots, t$.

Obviously $l_1 + l_2 + \cdots + l_{2t+1} \leq l$ and this inequality is equivalent to (1).

If we change the points and lines in the computation above , we will get (2) by branching process starting from a chosen line L.

The $k = 2t$ inequalities,

$$p_0 + p_2 + \cdots + p_{2t} \leq p, l_0 + l_2 + \ldots l_{2t} \leq l, \qquad (3)$$

are equivalent to (2).

QED

If $t+1 = s+1 = k$, then the order of the graph is $v = 2p = 2l$. The associated inequalities are equivalent to the well- known Tutte's inequality

$$v \geq 2(1 + (k-1) + \ldots (k-1)^{(g-2/2)} \qquad (4)$$

The well-known transport problem in Operation Research is equivalent to finding the tactical configuration of given size (number of edges) E with minimal number of vertices. There is a well-known efficient algorithm to solve this transport problem. In many cases this algorithm can be modified to solve efficiently the transport problem with additional restrictions, when one is looking at the tactical configurations with minimal number of vertices among graphs satisfying the list of restrictions. One of the natural list of restrictions is an absence of cycles of length 4, 6, $\dots 2k-2$. One can notice that the incidence graph of tactical configuration does not have cycles of odd length and the last requirement is equivalent to inequality $g \geq 2k$.

We refer to a tactical configuration with bidegrees $s+1$ and $r+1$ of girth $g \geq 2k$ and minimal possible order $p+l$ as a *cage configuration.*

It is clear that if inequalities (1) and (2) above are instead equalities, then we have a cage configuration. In this special situation we will use term a *perfect cage configuration.* It is clear that in the case of the perfect cage configuration $g = 2k$, we have an example of a one-time pad.

If $t = s$, then the cage configuration is a bipartite "cage" (see [9]) of degree $t+1$. A cage is a $t+1$-regular graph of given girth with the minimal number $v(k,g)$ of vertices. The cage with number of vertices on the Tutte's bound above and odd girth is the so-called Moore graph. The only Moore graphs of degree 2 are $2n+1$-gons. An m-gon is just a totality of vertices (points) and edges (lines) of ordinary cycle of length m with the natural incidence. A Moore graph of degree $k \geq 3$ has diameter 2 and $k \in \{3,7,51\}$.

We are interested in the case of even girth because our graphs are bipartite and have no odd cycles. In case of degree 2, a $2n$-gon is an example of perfect cage configuration. In fact, the $(2,g)$-cage is the g-circuit, and $v(g,2) = g$.

Let us list some known families of cages of even girth.

i) the $(k,4)$-cage is the complete bipartite graph $K_{k,k}$ and $v(k,4) = 2k$.

If $k = q+1$ for a prime power q, then

ii) 1) a $(k,6)$-cage is the incidence graph of a projective plane $PG(2,q)$, and $v(k,g) = 2(q^2+q+1)$;

2) a $(k,8)$-cage is the incidence graph of a generalized quadrangle $CQ(q,q)$, and $v(k,g) = 2(q^3+q^2+q+1)$;

3) a $(k,12)$-cage is the incidence graph of a generalized hexagon $GH(q,q)$, and $v(k,q) = 2(q+1)(q^4+q^2+1)$

The $(3,8)$-cage is the Tutte - Coxeter graph ($v = 30$) [22].

One has $v(7,6) = 90$ and the unique $(7,6)$ cage was independently found in [9], [5]. Finally, there are 3 distinct $(3,10)$- cages, all of them a biparite [10], - and $v(3,10) = 70$.

The problem of determining $v(k,g)$ was posed in 1959 by F. Kartesi who noticed that $v(3,5) = 10$ was realized by the Petersen graph. Sachs showed that $v(k,g)$ is finite and Erdös and Sashs gave the upper bound. This bound was improved in [11], and for the best known general bound, see [20]. For the case of bipartite graphs, similar problem had been considered in [18]. It is clear that a lower bound was given by Tutte's formula.

The nontrivial examples of families of cages ((ii) - (3)) are special cases of the generalized m-gons, defined by J. Tits in 1959 (see [21]) as tactical configurations of bidegrees $s+1$ and $t+1$ of girth $2m$ and diameter m. The pair (s,t) is known as the order of generalized m-gon

The following statement is well-known (see [21]):

Theorem 2 *A finite generalized n-gon of order (s,t) has $n \in \{3,4,6,8,12\}$ unless $s = t = 1$. If $s > 1$ and $t > 1$, then*

1) $n \neq 12$;

2) if $n = 4$, then $s \leq t^2$, $t \leq s^2$;

3) if $n = 6$, then st is a square and $s \leq t^3$, $t \leq s^3$;

4) if $n = 8$, then $2st$ is a square and $s \leq t^2$, $t \leq s^2$;

Without the second statement involving the inequalities, we have the original Feit-Higman theorem.

The known examples are incidence graphs of rank 2 finite simple groups of type Lie. The regular incidence graphs are cages which have been listed above.

The biregular but not regular generalized n-gons have parameters $s = q^\alpha$ and $t = q^\beta$, where q is some prime power. The list is below.

1) $n = 4$

(i) $s = q, r = q^2$ and q is arbitrary prime power

(ii) $s = q^2, r = q^3$ and q is arbitrary prime power

2) $n = 6$

$s = q^2$, $t = q^3$ and $q = 3^{2k+1}$, $k > 1$

3) $n = 8$

$s = q$, $t = q^2$ and $q = 2^{2k+1}$

Theorem 3 *Finite generalized polygons are perfect cage configurations.*

Proof: The order of regular generalized m-gons of degree $q+1$ is $1+q+q^2+\cdots+q^{m-1}$ and reaches the Tutte's bound for graphs of girth $m-2$. The finite irregular tactical configurations which are generalized polygons have to be of even diameter $m = 2k$. If their degrees are $r+1$ and $s+1$ then the numbers of points p and number of lines l can be computed by formulas

$$\begin{aligned} p &= 1+r+rs+r^2s+r^2s^2+\cdots+r^ks^k+r^{k+1}s^k\ , \\ l &= 1+s+sr+s^2r+\cdots+s^kr^{k+1}+s^{k+1}r^{k+1}\ , \end{aligned}$$

where k has to be an element of $\{2, 3, 4, 6\}$. They are at the bounds of (1) and (2). Thus finite generalized m-gone is a perfect cage configuration.

QED

It means that the generalized $m-$gons, related to simple Lie groups $G(F_q)$ with chosen Dynkin diagramm over the finite field F_q, $q = p^n$, $n \geq 1$, where p is prime, produce an infinite family of one-time pads. They have a certain resistance to attacks of type (ii). The best resistance given by the constant of complexity would be in the case of $G(q) = (F_4)^2(q)$, $q = 2^{2k+1}$. Here, the problem of finding the pass between 2 vertices of general position in generalized octagon is a well-known difficult problem in Algebraic Combinatorics.

The set of points (lines, respectively) of generalized m-gon can be considered as a disjoint union of vector spaces over the F_q. It is convenient to treat elements of F_q as tuples over the fixed alphabet F_p, so we may encrypt "potentially infinite" text over F_p. We may consider, say, a real-life encryption schemes with flexible keys if we restrict our passes to the set of passes for the m-gon related to $G(F_t)$ where $p \leq t \leq q$ is chosen power of p. We may vary the resistance $f(n)$ of such a scheme to attacks of type (i) (known ciphertext), or we may let it be as close to one-time pad as we want, or we may chose an increasing $f(n)$. However, the resistance to attack of type (ii) is bounded by some constant.

We need families of increasing girth to construct theoretical graph schemes of encryption for the case of increasing resistance to attacks of type (ii).

VI. PARALLELOTOPIC GRAPHS

Let Γ be a bipartite graph with partition sets P_i, $i = 1, 2$ (inputs and outputs) . Let M be a disjoint union of finite sets M_1 and M_2.

We say that Γ is a *bipartite parallelotopic graph* over (M_1, M_2) if there exists a function $\pi : V(\Gamma) \to M$ such that if $p \in P_i$, then $\pi(p) \in M_i$ and for every pair (p, j), $p \in P_i$, $j \in M_i$, there is a unique neighbor u with given $\pi(u) = j$.

It is clear that the bipartite parallelotopic graph Γ is a $(|M_1|, |M_2|)$ - biregular graph.

So a parallelotopic graph is just a bipartite graph with special colorings for inputs and outputs into $|M_1|$ and M_2 colors, respectively, such that for each vertex there exists a unique neighbor of any given color.

We refer also to the function π in the definition of bipartite parallelotopic graph as a *labelling.* We will often omit the term "bipartite", because all our graphs are bipartite. In case of encryption scheme of bipartite graph we will use one of the partition sets (inputs) as the plaintext space.

Linguistic graphs:

Let M be the Cartesian product of t copies of the set M. We say that the graph Γ is a *linguistic graph* over the set M with parameters m, k, r, s if

Γ is a bipartite parallelotopic graph over (V_1, V_2), $M_1 = M^r$, $M_2 = M^s$ with the set of points $I = M^m$ (inputs) and set of lines $O = M^k$ (outputs). (i.e. M^m and M^k are the partition sets of Γ). It is clear that $m + r = k + s$.

We use the term *linguistic coding scheme* for a pair (Γ, n), where Γ is linguistic graph and $n < g$ is the length of encoding sequences.

We choose a bipartite graph in the definition above because regular trees are infinite bipartite graphs and many bi-regular finite graphs of high girth can be obtained as their quotients (*homomorphic images*).

Using linguistic graphs, our messages and coding tools are words over the *alphabet* M and we can use the usual matching between real information and vertices of our graph. In case of $M = GF(q)$ the similarity with the linear coding is stronger, because of our messages and keys are tuples over the $GF(q)$.

VII. ABSOLUTELY OPTIMAL SCHEMES, ELEMENTS OF EXTREMAL GRAPH THEORY

One-time pads, whose keys and strings of random bits at least as long as the message itself, achieve the seemingly impossible: an eavesdropper is not able to determine any significant information from an intersected ciphertext. The simplest classical example is as follows: if p_i is the i-th bit of the plaintext, k_i is the i-th bit of the key, and c_i is the first bit of the ciphertext, then $c_i = p_i + k_i$, where $+$ is exclusive or, often written XOR, and is simply addition modulo 2. One-time pads must be used exactly once: if a key is ever reused, the system becomes highly vulnerable.

It is clear that the encryption scheme above cannot resist attacks of type (ii).

Families of one-time pads can be constructed for the case, when the key space and the message space have the same magnitude. For theoretical studies of cryptographic properties of graph Γ, we will always look at encryption scheme (Γ, t) , where $t = [g/2]$ and g is the girth of Γ.

Let Γ_i be an *absolutely optimal* family of graphs, i.e., family of graphs such that the ratio $p_{\text{key}}(i)/p_{\text{mes}}(i)$ of probabilities $p_{\text{key}}(i)$ and $p_{\text{mes}}(i)$ to guess the encoding sequence and to guess the message in the scheme (Γ_i, t_i), respectively, goes to 1 when i is growing.

The constructions of *absolutely optimal* families of schemes of high girth of increasing degree are connected with studies of some well-known problems in Extremal Graph Theory (see [2]). Let $e = ex(v, n)$ be, as usual, the greatest number of edges (size) in a graph on v vertices, which contains no cycles C_3, C_4, ..., C_n.

From Erdös' Even Cycle Theorem and its modifications [2], it follows that

$$ex(v, 2k) \leq Cv^{1+1/k} \qquad (5)$$

where $C \leq 90k$ is a positive constant.

It is easy to see that the magnitude of the extremal family of regular graphs of given girth and of unbounded degree have to be on the Erdös upper bound (5). This bound is known to be sharp precisely when $k = 2, 3$, and 5. Thus the problem of constructing absolutely optimal families of high girth is a difficult one. It has been shown in [13] that the incidence graphs of simple groups of Lie type of rank 2 can be used as absolutely optimal encryption schemes with certain resistance to attacks of kind (i), and examples of families of absolutely optimal coding schemes of parallelotopic graphs of girth 6, 8, 12 were considered. Let us look at one of them.

Example 1

Let $P = \{(x_1, x_2, x_3, x_4, x_5)|x_i \in GF(q)\}$, $L = \{[y_1, y_2, y_3, y_4, y_5]|y_i \in GF(q)\}$. Let us define a bipartite graph I as: $(a, b, c, d, e)I[x, y, z, u, v]$ if and only if

$y - b = xa$

$z - 2c = -2xb$

$u - 3d = -3xc$

$2v - 3e = 3zb - 3yc - ua$

Input (a, b, c, d, e) and output $[x, y, z, u, v]$ are connected by edge in graph I iff the conditions above hold.

From the equations above, it follows that π : $\pi((x_1, x_2, x_3, x_4, x_5)) = x_1$ and $\pi([y_1, y_2, y_3, y_4, y_5]) = y_1$ is a labelling for the parallelotopic graph I.

It can be shown that for char$GF(q) > 3$ the girth of this graph is at least 12. Directly from the equations above we can get that I defines the linguistic coding scheme with parameters $(1, 1, 5, 5)$ of affine type over $GF(q)$. It is clear that in case of encoding tuples of length 5, we get $p_{\text{key}} = 1/q(q-1)^4$, $p_{\text{mes}} = 1/q^5$ and $I = I_5(q)$ is an absolutely optimal family of linguistic graphs.

In fact, we are working with graphs of tactical configuration; in this case, we can get stronger bounds than those obtained by the Even Cycle Theorem.

Let us consider some corollaries of Lemma 1. Without loss of generality, we will assume $r = a^m, s = a$, where $m \geq 1$

In case of $k = 2t$, we may omit all terms of the left-hand-side 0f (2) except the highest terms:

$a^{mt}a < p$, $a^ta^{mt} < l$.

Adding new inequalities, we will get $a^{(m+1)t} < v/2$, or $a < (v/2)^{1/((m+1)t)}$.

We have $l(a+1) = e$ or $la = e - l$. Thus $e - l < l(v/2)^{1/((m+1)t)}$.

Putting v instead of l, we will get $e < v(v/2)^{(1/((m+1)t)} + v$, which is equivalent to

$$e \leq (1/2)^{(1/((m+1)t)}v^{(1+1/(m+1)t)} + v \qquad (6)$$

Remark: If $m = 1$, the magnitude of right-hand-side is the same as that in the Erdös' Even Cycle Theorem, but the constant is better, has monotonic dependence on m, and always < 1.

If $m > 1$ then (4) is stronger then Erdös' inequality in the sense of magnitude. Of course (4) is applicable only to bipartite biregular graphs.

Let us consider the case $k = 2t + 1$. If we discard some summations from the left-hand-side of (1) we get $r^ts^t + r^{t+1}s^t < p$.

As we set before, $r = a^m$, $s = a$. Thus $a^{mt+1}(a^m+1) < p$.

From $l(p+1) = p(a^m+1) = e$, we get $a^{mt+t}(l/p)(a+1)l < p$ or $a^{mt+t}(a+1)l < p^2 = l^2(a+1)^2/(a^m+1)^2$.

Simplifying the last inequality, we obtain

$a^{mt+t}(a^m+1)^2/(a+1) < l$.

We can notice that function $f(a) = (a^m+1)^2/(a+1)$ is increasing. Thus

$f(a-1)a^{mt+t} < l$ or $a^{mt+t-1}[(a-1)^2+1] < l$.

From the last inequality, we get

$(a-1)^{mt+2m+t-1} < l$ or $a - 1 < e^(1/(mt+m+t-1))$.

We know that $l(a+1) = e$. So $l(a-1) = e - 2l$ and multiplication of two sides of the last inequality by l produces

$$e < l^{1+1/(m(t+2)+t-1)} + 2l.$$

Finally, substitution of order v instead of l gives us a slightly weaker inequality

$$e \leq v^{1+1/(m(t+2)+t-1)} + 2v \qquad (7)$$

Remark If $m = 1$, then the bound above has the same magnitude with that of Erdös' bound in the Even Cycle Theorem, but the constant is better then in (5). In fact, we can improve the constant by substituting $l = v/2$ into inequality (3).

$$e \leq (1/2)^{1+1/(2t+1)}v^{1+1/(2t+1)} + v \qquad (8)$$

If $m > 1$ then magnitude of (7) is better than in the Erdös' bound.

A family of graphs of tactic configurations is *absolute optimal*, if the magnitude or the size of graphs is same as in that in right-hand-side of the inequalities (i) and (ii). It is clear that a linguistic graph cannot be a one-time pad, but family of linguistic can be absolutely optimal as in the example above. We may generalize graphs $I_5(q)$, which are special induced subgraphs of the generalized gexagon related to special induced subgraphs in the in-group $G_2(q)$ in the following way:

Example 2. Let the point P and the line L form a chosen edge of a tactical configuration of generalized m-gon. We can consider a totality V_P (V_L) of points (lines, respectively) at maximal distance d ($d = m/2$ or $d = [m/2] + 1$) with the restriction of incidence relation on $V_P \cup V_L$. It will be a linguistic absolutely optimal family of graphs.

If the generalized m-gon is a geometry of finite simple group G(q), $q = p^n$, where p is a prime, then we may treat V_p , V_L and the totality of walks as a vector spaces over F_p and use them for the encryption of potentially infinite text.

For known absolutely optimal schemes of high girth with the resistance to attacks of type (ii), the girth is ≤ 16, which is the girth of generalized octagon. The problem of breaking the key during the attack of type (ii) is equivalent to the solution of a system of nonlinear equations of degree $d(g)$ depending on the girth g. Graphs with better resistance to attacks of this type will be considered in the next section.

VIII. OPTIMAL SCHEMES OF UNBOUNDED GIRTH

It is known that one-time pads are impractical because in real life we need to deal with large amounts of information. A reasonable strategy is to consider the weaker requirement then equality of dimensions d_{key} of key space and dimension d_{pt} of plain text space. Let us consider the family of graphs Γ_i of increasing girth g_i such that for corresponding coding scheme $(\Gamma_i, t = [g_i/2])$ $\lim (p(i)_{\text{key}})^c / p(i)_{\text{mes}} = 1$, $i \to \infty$ where c is the constant which does not depend on i.

In this situation we say that the schemes of Γ_i form an *optimal family* of schemes. It is easy to check that in case of the optimal family of schemes corresponding to graphs of degree l_i and unbounded girth g_i, we have

$$g_i \geq \gamma \log_{l_i - 1}(v_i) \tag{9}$$

The last formula means that Γ_i, $i = 1, \ldots$ form an infinite family of graphs of large girth in the sense of N. Biggs [1].

A few examples of such families are known (see [1] and [8], [19]).

We have $\gamma \leq 2$, because of (1), but no family has been found for which $\gamma = 2$. Bigger γs correspond to more secure coding schemes. A. Lubotzky (see [11]) conjectured that $\gamma \leq 4/3$.

IX. EXPLICIT CONSTRUCTIONS, COMPARISON WITH OTHER METHODS

Explicit constructions of optimal families of linguistic graphs over $M = GF(q)$ with good complexity of computation of walks have been considered in [13], [24].

We are exploring one of them, which is the family of q-regular linguistic graphs $L_n(q)$ such that the input $(x_1, x_2, \ldots, x_n) = (x)$ and the output $[y_1, y_2, \ldots, y_n] = [y]$ are neighbors if $x_i - y_i = x_{k(i)} y_{s(i)}$ for $2 \leq i \leq n$, where $k(i) < i$, $s(i) \leq i$ and n can be any number. In fact, the parallelotopic morphism of $L_n(q)$ onto $L_m(q)$, $n > m$ is induced by canonical projecture of n-dimensional vector space onto m-dimensional. Each graph $L_n(q)$ is similar to the graph from Example 1 above.

Of course, we are not computing the adjacency matrix, but we have two affine operators $N(\alpha, (x))$ and $N(\alpha, [y])$, and we compute the neighbor of (x) and $[y]$ with the first component α.

If n and the dimension d of key space are "sufficiently large", then our encryption resists to attack of kind (ii). Why?

We have the following argument: Family $L_n(q)$, $q > 2$ is a good approximation of the q regular tree T_q. More precisely T_q is the projective limit of $L_n(q)$. $n \to \infty$. The plaintext p and the ciphertext c are vertices in $L_n(q)$ roughly T_q at a distance d. The key is the uniquely determined pass between p and c. Let $h_q(d)$ be the complexity of the determination of the pass between p and c. Then h_q is an increasing function in variable d. So, if the girth is growing, then we may attain the level security we want. There is no proof that h_q is polynomial function.

Remark. The encryption and decryption procedures, both depending on key, have the same complexity (symmetric encryption), but the key is hard to compute.

Let us compare our encryption with the following popular scheme of linear encryption:

We treat our message as a polynomial $f(x)$ over $GF(q)$ (our tuple is an array of coefficients of $f(x)$). The linear coding procedure is just a multiplication of our $f(x)$ of degree $n - 1$ by a polynomial $g(x)$, $\deg(g(x)) = t$, $t > 0$. Thus, y is just an array of coefficients of the polynomial $F(x) = f(x)g(x)$, $m = \deg F(x) = n + t - 1$.

It is clear that this symmetric encoding is cannot resist attacks of type (ii) and sizes of plaintext and ciphertext are different. Counting of operation in case of equal dimensions of the plaintext and the ciphertext for the classical scheme as above and our scheme corresponding to $L_n(q)$, where q is a prime, shows that our encryption is faster.

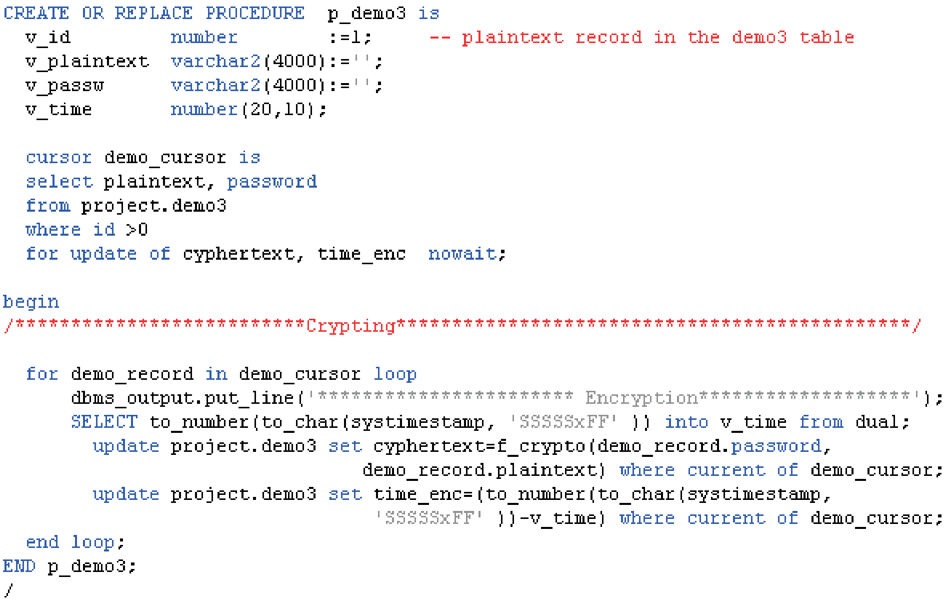
The development of the prototype model (a fragment PL/SQL code is shown in Figure 1) allows us to test the resistance of the algorithm above to attacks of different time (see Figure.2. Prototype Model, Based on Oracle Portal 9iAS, Release 1).

Our initial results from such tests show that the results are encouraging ([4]). Let us consider, for example, the case of $p = 127$ (size of ASCII alphabet minus "delete" character). Let $t(k, l)$ be time (in seconds) we need to encrypt (or decrypt because of symmetry) file, the size of which is k kilobytes with a password of length l (key space roughly 2^{7l})) by a Pentium IV (Linux Red Hat 7.2 Operational System, Oracle Portal 9iAS, PL/SQL language). Then some values of $t(k, l)$ can be presented by the following matrix, represented graphically in Figure 3:

$l \setminus k$	1Kb	2Kb	3Kb	3.5Kb
9	0.194187	0.3737	0.560048	0.653088
13	0.276417	0.551537	0.830635	0.966417
17	0.365576	0.731214	1.099837	1.281369
21	0.454007	0.910893	1.368683	1.596617
25	0.542276	1.090975	1.63913	1.909816
35	0.766706	1.53664	2.312793	2.697898
55	1.209971	2.438037	3.656054	4.270853
75	1.659585	3.331478	5.005951	5.84399

Encryption and Decryption functions were coded by PL/SQL language too. By using C++ for encryption and decryption functions, we get faster results for algorithm evaluation.

The following result was reported recently in [16].



```
CREATE OR REPLACE PROCEDURE  p_demo3 is
  v_id         number         :=1;     -- plaintext record in the demo3 table
  v_plaintext  varchar2(4000):='';
  v_passw      varchar2(4000):='';
  v_time       number(20,10);

  cursor demo_cursor is
  select plaintext, password
  from project.demo3
  where id >0
  for update of cyphertext, time_enc  nowait;

begin
/**************************Crypting**********************************************/

  for demo_record in demo_cursor loop
      dbms_output.put_line('*********************** Encryption*******************');
      SELECT to_number(to_char(systimestamp, 'SSSSSxFF' )) into v_time from dual;
        update project.demo3 set cyphertext=f_crypto(demo_record.password,
                                  demo_record.plaintext) where current of demo_cursor;
        update project.demo3 set time_enc=(to_number(to_char(systimestamp,
                                   'SSSSSxFF' ))-v_time) where current of demo_cursor;
  end loop;
END p_demo3;
/
```

FIG. 1: PL/SQL code

Theorem 4 *The complexity of the above algorithm for the encryption of text of length* n *over some alphabet with the password of length* m *is* $O(nm)$.

It means that the time of the encryption is Cnm, where the constant C depends on chosen alphabet and parameters of the computer only.

X. PARALLELOTOPIC GRAPHS OF LARGE GIRTH AND ASYMMETRIC ALGORITHMS

Let Γ be a parallelotopic graph without loops and multiple edges of girth g , i. e., the graph without cycles of length $< g$.

We can consider the neighbor $w = N_a(v)$ of vertex v in graph Γ such that the color of edge w is $a \in M$. Let $N(x_1, x_2, \dots, x_t) = N_{x_t}(N_{x_{t-1}}(\dots(N_{x_1}(v))\dots))$ in variables $x_i \in M$, $v \in V(\Gamma)$. As before, we will treat an element v of $V(\Gamma)$ as a plaintext, and the sequence $v = v_0, v_1, v_2, \dots, v_t$, where $v_i \Gamma v_{i+1}$ as the encryption tool. If $N_{a_{i+1}(v_i)=v_{i+1}}$, then the pass is uniquely defined by string (or word) a_1, $a_2, \dots, a_t$ over the alphabet of "colors" M and "inverse" string $a_{t-1}, a_{t-2}, \dots, a_0$. Here, a_0 is the color of plaintext defines "decrypting" sequence of vertices. Thus we identify walks on Γ with strings over M.

The RSA algorithm demonstrated that the information for encryption (number pq) can be just part of the information for decryption (at least, numbers p and q).

Let us consider such a situation ("encryption with secret") in case of graph encryption.

Let ϕ_w be the binary relation $\phi_w = \{(u, v) | v = N(a_1, a_2, \dots, a_t)(v)\}$, where w is the string $a_1, a_2, \dots, a_t$. It is clear that for the encryption with the key w, we do not need the information about our graph Γ. We need only the graph Γ_w of the binary relation ϕ_w. Let $N^w(v)$ be the operator of taking the neighbor of the vertex v in the graph Γ_v. The usual situation is that the complexity of computation N^w is much worse than N^w if we do not know the decomposition

$N^w = N_{a_1}(N_{a_2}(\cdots(N_{a_t})\cdots))$. (1)

So we may present the function N_w in the form

$$N^w = N^{w_1}(N^{w_2}\cdots(N^{w_s})\cdots)),$$

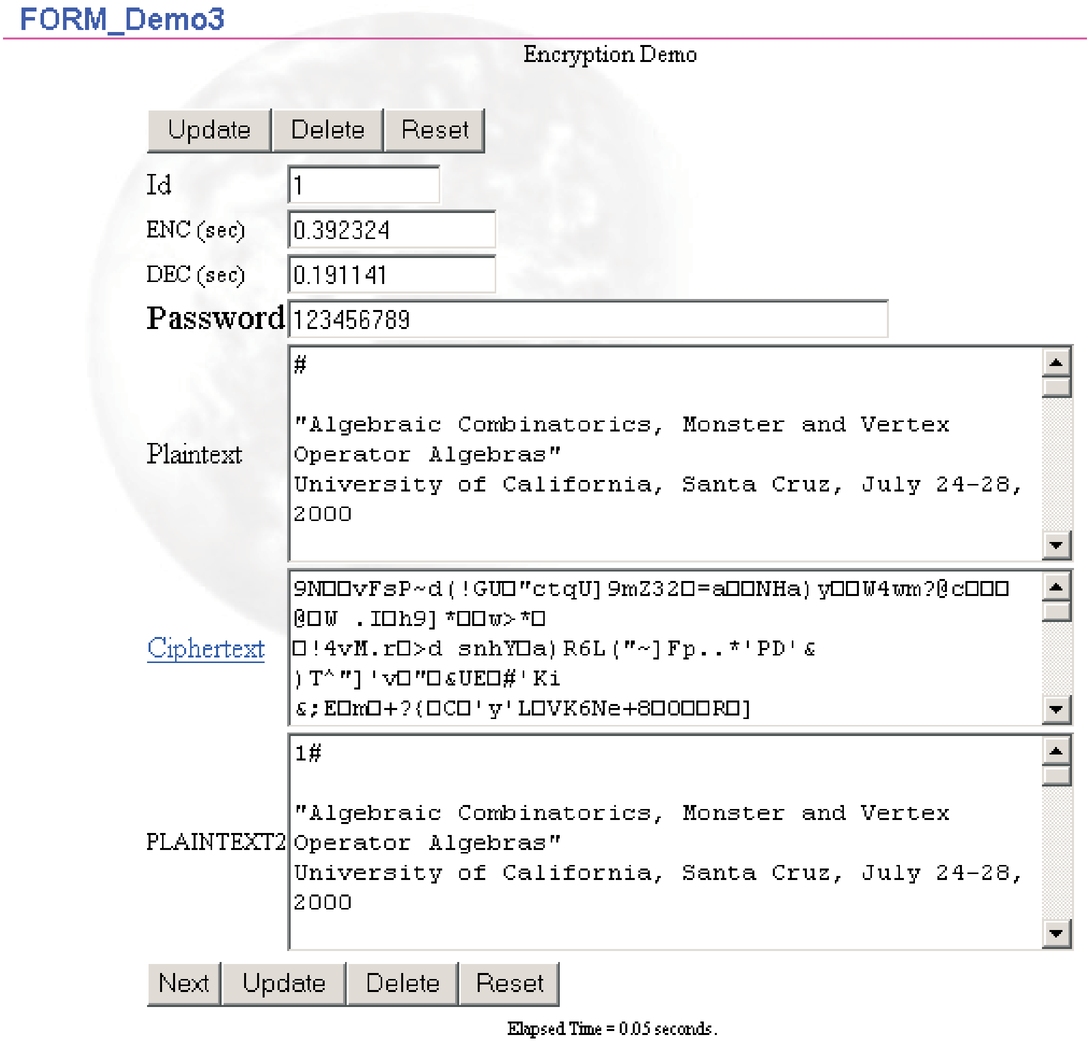
where word w is a product (concatenation) of words w_1, $w_2 \cdots, w_s$ to make computation of N_w faster.

It is clear that to find the decomposition above could be a hard task even in case where the graph Γ is known.

We can give our recipient the "public key" $N^{w_1}, \dots N^{w_s}$. The recipient can encrypt, but can not decrypt if the computation of the superpositions of $(N^{w_s})^{-1}, (N^{w_{s-1}})^{-1}, \dots (N^{w_1})^{-1}$ is sufficiently difficult.

Let us discuss this approach in case of *linguistic graphs* Γ of rational (polynomial) type over commutative ring K. In case of such graphs, $M = K$ and the function $N_a(v)$, obtained by taking the neighbor of vertex $v = (y_1, y_2, \dots, y_t) \in I \cup O$, is a polynomial expression from variables y_i, $i = 1, k$. A degree of polynomial linguistic graph is the maximum degree of polynomial expressions for each $N_a(v)$ in variables y_i.

In this case N^{w_i}, $i = 1, 2, \dots, s$ are polynomial expressions P_i over the commutative field K of degree d_i. For simplicity let us assume that the graph is regular , i.e., $O = I = K^n$, and the recipient has polynomial expression N^w without its decomposition into smaller expressions. If $\deg N^w = d$ then encryption from the given vertex could be done not more than for $O((n^d)$ elementary steps. Thus, if the recipient's "public key" is given as the



FORM_Demo3

Encryption Demo

Id 1

ENC (sec) 0.392324

DEC (sec) 0.191141

Password 123456789

Plaintext

```
#

"Algebraic Combinatorics, Monster and Vertex
Operator Algebras"
University of California, Santa Cruz, July 24-28,
2000
```

Ciphertext

```
9N□□vFsP~d(!GU□"ctqU]9mZ32□=a□□NHa)y□□W4wm?@c□□□
@□W .I□h9]*□□w>*□
□!4vM.r□>d snhY□a)R6L("~]Fp..*'PD'&
)T^"]'v□"□&UE□#'Ki
&;E□m□+?{□C□'y'L□VK6Ne+8□□□□R□]
```

PLAINTEXT2

```
1#

"Algebraic Combinatorics, Monster and Vertex
Operator Algebras"
University of California, Santa Cruz, July 24-28,
2000
```

Elapsed Time = 0.05 seconds.

FIG. 2: Prototype Model, Released for Oracle Portal 9iAS, Release 1.

list of coefficients of monomial expressions for N^w, then the complexity of encoding procedure for the recipient will be proportional to size of this list.

In this case, encryption (or decryption) is faster because the steps N_{a_i} are smaller.

What does a recipient need for the decryption of a given message $(b_1, b_2, \ldots, b_n)$? The recipient has to solve the system of polynomial equations

$$N_w(x_1, x_2, \ldots, x_n) = (b_1, b_2, \ldots, b_n).$$

This task is a classical difficult problem in algebra. The system above can be investigated for $d^{O(n^2)}$ steps, where d is the maximal degree of polynomials. We can do better (d^{Cn}) if we know that the system is consistent.

If we have a family of polynomial linguistic graph of bounded degree, we may choose the dimension n such that the recipient could encrypt but could not decrypt and use graph encryption in "public key fashion", because we would use the gap between computations of polynomial in a given point and the investigation of a given system of equations.

XI. CONCLUSION

Security of network and data are important problem in network learning. Reliable private key algorithm are needed for the data storage, and public key type algorithms are needed for the exchange of passwords and for solving authorization problems (digital signatures).

The graph theoretical approach allows us to construct a family of fast nonlinear algorithms of encryption with flexible size of keys. Some of the algorithms are faster then linear or affine encryption methods but have better security, in particular they are resistant to attacks when the adversary knows the plaintext and the ciphertext both ($p \cap c$ attacks).

In the case (theoretical, but not practical) when size

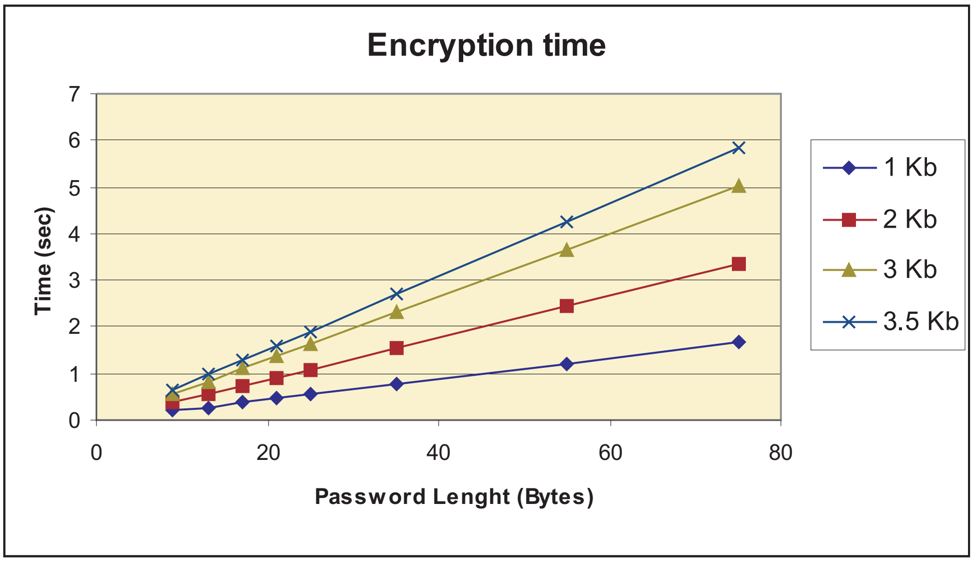


FIG. 3: Encryption time.

of key coincides with size of text, we get the absolute algorithm according to C. Shannon's definition, with a certain resistance to $p \cap c$ attacks (classical examples have no $p \cap c$ resistance). In fact, if the size of text is growing but the size of the key is fixed, $p \cap c$ security is also growing. That is why we have good security in case of large texts (web pages with course descriptions, examination scripts, image data)

Resistance to $p \cap c$ attacks allow us to keep a chosen password for a while, in contrast with "one-time pad" algorithms. Finally we may modified the algorithms based on linguistic graphs to get the asymmetry and use them for the exchange of keys (public key case) and digital signatures.

The prototype model of the package was demonstrated at the Conference on Multi Agents in Wollongong and selected for the presentation at the Congress on Network Learning (Berlin 2001).

Acknowledgments

The author is very grateful to Professors I. Sparlinsky and J. Pieprzyk for useful comments on the topic.

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

[2] B. Bollobás, *Extremal Graph Theory*, Academic Press.

[3] Don Coppersmith, *The Data Encryption Standard (DES) and its strength against attacks, IBM J. Res Dev.*, 38 (1994), 243-250.

[4] V. Ustimenko and D. Sharma, *CRYPTIM : system to encrypt text and Image Data*, Proceedings of International ICSC Congress on Intelligent systems 2000, Wollongong, Australia 2001, 11pp.

[5] R. Baker, *An elliptic semiplane*, J. Comb Theory (A), 25, 1988, 193-195.

[6] S. Landau. *Standing the Test of Time:The Data Encryption Standard*, Notices of the AMS, March 2000, pp 341-349.

[7] F. Lazebnik, V. A. Ustimenko and A. J. Woldar, *A New Series of Dense Graphs of High Girth*, Bull (New Series) of AMS, v.32, N1, (1995), 73-79.

[8] A. Lubotzky, *Discrete Groups, Expanding graphs and Invariant Measures*, Progr. in Math., 125, Birkhoiser, 1994.

[9] M. O'Keefe, P. Wong, *The smallest graph of girth 6 and valency 7*, J. Graph Theory, 5 (1981),79-85.

[10] M. O'Keefe, P. Wong, *The smallest graph of girth 10 and valency 3*, J. Graph Theory 5 (1980), 91-105.

[11] N. Sauer. *Extermaleigenschaften regularer Graphen gegebener Taillenweite*, 1, 2, Osterreich. Acad. Wiss. Math. Natur. Kl. S. -B 2, 176 (1967), 9-25, 27-43.

[12] V. A. Ustimenko, *Random Walks on special graphs and Cryptography*, AMS Meeting, Louisville, March , 1998.

[13] V. A. Ustimenko, *Coordinatization of regular tree and its quotients*, In the volume "Voronoi's Impact in Modern Science": (Proceedings of Memorial Voronoi Conference, Kiev, 1998), Kiev, IM AN Ukraine, July, 1998, pp. 125 - 152.

[14] V. A. Ustimenko, *On the Varieties of Parabolic Subgroups, their Generalizations and Combinatorial Applications*, Acta Applicandae Mathematicae 52 (1998): pp. 223–238.

[15] V. Ustimenko and D. Sharma, *Special Graphs in Cryptography*, in Proceedings of 2000 International Workshop on

Practice and Theory in Public Key Cryptography (PKC 2000), Melbourne, December 1999.

[16] V. A. Ustimenko, *CRYPTIM: Graphs as Tools for Symmetric Encryption*, Springer Lecture Notes in Computer Sci., LNCS 2227: Proceedings of AAECC-14 Symposium, Applied Algebra, Algebraic Algorithms and Error Correcting Codes.

[17] Carter, R.W.,*Simple Groups of Lie Type*, Wiley, New York (1972).

[18] Füredi, Z., Lazebnik, F., Seress, Á., V. A. Ustimenko and A.J.Woldar, *Graphs of prescribed girth and bi-degree*, J. Combin. Theory B 64 (2) (1995), 228–239.

[19] Lazebnik, F., Ustimenko, V.A. and A.J. Woldar, *A new series of dense graphs of high girth*, Bull. AMS 32 (1) (1995), 73–79.

[20] Lazebnik, F., Ustimenko, V.A. and A.J. Woldar, *New upper bounds on the order of cages*, Electronic J. Combin. 14 R13 (1997), 1–11.

[21] Tits J.,*Buildings of Spherical Type and Finite BN-pairs*,Lecture Notes in Math. 386, Springer-Verlag, Berlin (1974).

[22] W. Tutte, *A family of cubical graphs*, Proc. Cambridge Philos. Soc. 43 (1945)

[23] C. Shannon (1949),*Communication theory of secrecy systems*, Bell. Syst. Tech. J., 28, 656-715.

[24] V. Ustimenko. *Families of graphs with Special Arcs and Cryptography*, CITR-TR-84, Technical Reports of CITR, Auckland University, February, 2001,30pp.