Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

Electronic Notes in Discrete Mathematics 49 (2015) 31–35

www.elsevier.com/locate/endm

An application of Combinatorics in Cryptography

Peter Horak

School of Interdisciplinary Arts & Sciences
University of Washington
Tacoma, USA

Igor Semaev

Department of Informatics
University of Bergen
Bergen, Norway

Zsolt Tuza

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences
University of Pannonia
Veszprém, Hungary

Abstract

Nowadays sparse systems of equations occur frequently in science and engineering. In this contribution we deal with sparse system that are common in cryptanalysis. Given a cipher system, one converts it into a system of sparse equations, and then the system is solved to retrieve either a key or a plaintext. Raddum and Semaev proposed a new method for solving such sparse systems. It turns out that a combinatorial MaxMinMax problem provides bounds on the case where the average computational complexity of their method is maximum. We focus on this MaxMinMax problem and present results over finite and infinite fields.

Keywords: sparse system of equations; a computational complexity; algebraic cryptanalysis

http://dx.doi.org/10.1016/j.endm.2015.06.006

1 Introduction

Sparse objects such as sparse matrices and sparse systems of (non-)linear equations occur frequently in science or engineering. Nowadays sparse systems are frequently studied in algebraic cryptoanalysis. First, given a cipher system, one converts it into a system of equations. Second, the system of equations is solved to retrieve either a key or a plaintext. As pointed out in [1], this system of equations will be sparse, since efficient implementations of real-word systems require a low gate count.

There are plenty of papers on methods for solving a sparse system of equations. In [4] a so called Gluing Algorithm was designed to solve such systems over a finite field $GF(q)$. If the set S_k of solutions of the first k equations together with the next equation $f_{k+1} = 0$ is given then the algorithm constructs the set S_{k+1}. It is shown there that the average complexity of finding all solutions to the original system is $O(mq^{\max|\cup_1^k X_j|-k})$, where m is the total number of equations, and $\cup_1^k X_j$ is the set of all unknowns actively occurring in the first k equations. Clearly, the complexity of finding all solutions to the system by the Gluing Algorithm depends on the order of equations. Hence one is interested to find a permutation π that minimizes the average complexity, and also to describe the worst case scenario, i.e., the system of equations for which the average complexity is maximum. Therefore in [5] Semaev suggested to study the following combinatorial MaxMinMax problem.

Let $\mathcal{S}_{n,m,c}$ be the family of all collections of sets $\mathcal{X} = \{X_1, ..., X_m\}$, where the X_i are subsets of an underlying n-set X, and $|X_i| \leq c$ holds for all $i \in [m]$; we allow that some set may occur in $\mathcal{X}$ more than once. Then we define

$$f_c(n,m) := \max_{\mathcal{X}} \min_{\pi} \max_{1\leq k\leq m} \left(\left|\bigcup_{i=1}^{k} X_{\pi(i)}\right| - k\right) \tag{1}$$

where the minimum runs over all permutations π on $[m]$, and the maximum is taken over all families $\mathcal{X}$ in $\mathcal{S}_{n,m,c}$.

In [2] the authors confined themselves to the case $|X_i| \leq 3$ for all $i \in [m]$. It was shown there that, for $n \geq 2$ and all m, $f_2(n,m)$ equals the maximum number of non-trivial components in a simple graph on n vertices with m edges; in particular, $f_2(n,m) = 1$ for $m \geq n-1$. The main result of that paper claims that $f_3(n,n)$ grows linearly. More precisely, the following estimates are valid.

Theorem 1.1 *For all n sufficiently large, $f_3(n,n) \geq \frac{n}{12.2137}$ holds, while $f_3(n,n) \leq \lceil \frac{n}{4} \rceil + 2$ for all $n \geq 3$.*

Later, an asymptotically better upper bound was proved in [5]; moreover, the proof is algorithmic.

Theorem 1.2 *For all n, $f_3(n,n) \leq \frac{n}{5.7883} + 1 + 2\log_2 n$.*

As a corollary we get: Let $\mathcal{X}$ be fixed. If $|X_i| \leq 3$, $m = n$, then the average complexity of finding all solutions in $GF(q)$ to polynomial equation system $f_i(X_i) = 0\,(1 \leq i \leq m)$ is at most $q^{\frac{n}{5.7883}+O(\log n)}$ for arbitrary $\mathcal{X}$ and q.

In [3] a new method for solving systems of algebraic equations common in cryptanalysis has been proposed. This method differs from the others in that the equations are not represented as multivariate polynomials, but as a system of Multiple Right Hand Sides (MRHS) linear equations. The results overcome significantly what was previously achieved with Gröbner Basis related algorithms. We point out that equations describing the Advanced Encryption Standard (AES) can be expressed in MRHS form as well. AES is likely the most commonly used symmetric-key cipher; AES became effective as a federal government standard on May 26, 2002 after approval by the Secretary of Commerce. It is the first publicly accessible and open cipher approved by the National Security Agency (NSA) for top secret information when used in an NSA approved cryptographic module.

Let X be a column n-vector over $GF(q)$. Then MRHS is a system

$$A_i X \in \{b_{i_1}, \ldots, b_{i_{s_i}}\}, \quad i = 1, \ldots, m, \tag{2}$$

where the A_i are matrices over F_q of size $k_i \times n$ and of rank k_i, and the b_{ij} are column vectors of length k_i. An $X = X_0$ is a solution to (2) if it satisfies all inclusions in (2). Methods to solve such equations were introduced in [3] as well.

One of the main goals of our paper is to get asymptotic bounds on the complexity of solving (2). As noted by Semaev, such bounds can be obtained by studying a generalisation of the combinatorial problem described in (2). The idea is based on the following statement that enables to present the given cryptographic problem in combinatorial terms. Let r_k denote the rank of all row-vectors in $A_1, A_2, \ldots, A_k$.

Theorem 1.3 *Suppose that the right hand side column vectors in (2) are zeros of a uniformly random polynomial over $GF(q)$ of degree $< q$ in each*

variable(in other words, each particular column vector appears independently with probability $1/q$*). Then the average complexity of solving* (2) *is at most*

$$m \max_k q^{r_k - k}.$$

So, as in the original problem, the complexity of the solution depends on the order of matrices A_i. Hence, the complexity of the problem is in fact a generalisation of the function $f_c(n,m)$ defined in (1), namely the size of the union of the first k sets is replaced by the rank of vectors belonging to the first k matrices. Formally, let $\mathcal{S}_{n,m,c,V}$ be the family of all collections of sets of **vectors** $\mathcal{X} = \{X_1, ..., X_m\}$ in an n-dimensional vector space V, over $GF(q)$ or over real numbers, under the restriction $|X_i| \leq c$ for all $i \in [m]$. We set

$$F_c(n,m) := \max_{\mathcal{X}} \min_{\pi} \max_{1 \leq k \leq m} (\mathbf{rank}\left(\bigcup_{i=1}^{k} X_{\pi(i)}\right) - k), \tag{3}$$

where the minimum runs over all permutations π on $[m]$, and the maximum is taken over all families $\mathcal{X}$ in $\mathcal{S}_{n,m,c,V}$.

Although functions $f_c(n,m)$ and $F_c(n,m)$ are defined in a similar way, it turns out that their behavior is dramatically different.

Clearly, $f_c(n,n) \leq n - \frac{n}{c}$ is a trivial upper bound. This bound is attained for some vector spaces. In the case of the n-dimensional space $\mathbb{R}^n$ over the real numbers the tightness of this bound follows from the fact that $\mathbb{R}^n$ contains infinitely many vectors such that any n of them are independent.

We now focus on finite fields that are important for the original cryptographic setting of the problem. As mentioned above, $f_2(n,m) = 1$ holds for all $m \geq n$. It turns out that even the case $c = 2$ constitutes a challenge for most vector spaces V. Surprisingly, using Reed-Solomon codes, the trivial upper bound can be attained even for some finite fields.

Theorem 1.4 *Let* $GF(q)$ *be a finite field, where* $q \geq 2n$*. Then for any* n *we have* $F_2(n,n) = \frac{n}{2}$.

At the end we focus on the binary field, the field most important for cryptographic application. We start with an upper bound.

Theorem 1.5 *There is an absolute constant* c *such that* $F_2(n,n) \leq \frac{n}{2} - c \log n$.

As to a lower bound we state first a linear one based on Gilbert-Varshamov type asymptotic lower bound for linear binary code size.

Theorem 1.6 *For all large enough* n*,* $F_2(n,n) \geq \frac{n}{9.0886}$.

At the moment we do not have a conjecture about the asymptotic rate of growth of the function $F_2(n,n)$. To indicate the difficulty of the problem we present a family exhibiting that a linear lower bound on $F_2(n,n)$ can be obtained even by a very special system.

Theorem 1.7 *For sufficiently large n, there is a positive constant c and a family $\mathcal{X} = \{X_1, ..., X_n\}$ of binary vectors, where for all $i \in [n]$, $|X_i| = 2$, and X_i contains a unit vector and a vector with exactly two non-zero coordinates, such that*

$$\min_{\pi} \max_{1 \le k \le n} (\mathbf{rank}\left(\bigcup_{i=1}^{k} X_{\pi(i)}\right) - k) \ge cn,$$

where the minimum runs over all permutations on $[n]$.

Acknowledgement. The authors are indebted to Noga Alon for discussions on expanders and on probabilistic methods, which lead to an improvement of the lower bound in Theorem 1.1.

References

[1] N. T. Courtois, and J. Pieprzyk, "Cryptanalysis of block ciphers with overdefined systems of equations" in Advances of Cryptology, Asiacrypt 2002, LNCS 2501, Springer, 2002, 267–287.

[2] P. Horak and Zs. Tuza, Speeding up deciphering by hypergraph ordering, Designs, Codes and Cryptography 75 (2015), 175 – 185.

[3] H. Raddum, and I. Semaev, Solving multiple right hand side equations, Designs, Codes and Cryptography 49 (2008), 147–160.

[4] I. Semaev, On solving sparse algebraic equations over finite fields, Designs, Codes and Cryptography 49 (2008), 47–60.

[5] I. Semaev, MaxMinMax problem and sparse equations over finite fields, Desings, Codes and Cryptography, DOI 10.1007/s10623-015-0058-6.