SCIENCE CHINA
Information Sciences

• INSIGHT •
Special Focus on Advanced Technology of Software
May 2016, Vol. 59 050104:1–050104:2
doi: 10.1007/s11432-016-5548-2

Solving Boolean equation systems and applications in cryptanalysis

Xiaoshan GAO[1] & Zhenyu HUANG[2*]

[1]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China;*
[2]*State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093*, China*

Received November 9, 2015; accepted December 30, 2015; published online April 8, 2016

Citation Gao X S, Huang Z Y. Solving Boolean equation systems and applications in cryptanalysis. Sci China Inf Sci, 2016, 59(5): 050104, doi: 10.1007/s11432-016-5548-2

Solving Boolean equation systems plays a fundamental role in many important fields such as coding theory, cryptology, and analysis of computer hardware. To find efficient algorithms for solving Boolean equations and estimate their complexities are central issues in theoretic computer science. In this note, we will survey recent results on the complexity of using the approximate algorithm to solve Boolean equation systems and on new characteristic set algorithms for solving Boolean equation systems and their applications in cryptanalysis.

Approximate algorithms for solving Boolean equations. It is known that finding the solutions of a Boolean equation system is NP-hard. A natural question is whether there exists effective approximate algorithms for solving such equation systems.

Consider the optimization problem MAX-MQ: given a set of m quadratic polynomial equations $F = \{f_1, f_2, \ldots, f_m\}$ with n variables, where the coefficients of f_i belong to a finite field $\mathbb{F}_q$, find a solution in $\mathbb{F}_q^n$ satisfying the maximal number of equations in F. Obviously, equations with higher degrees can be transformed into quadratic equations by introducing new variables and solving Boolean equations is a special case of MAX-MQ when $q = 2$. For problem MAX-MQ, the following results were proved.

Theorem 1 (See [1]). It is NP-hard to approximate MAX-MQ with an approximation ratio of $q - \epsilon$ for $\epsilon > 0$.

Theorem 2 (See [2]). Random assignment is a $(q + \frac{q^2}{q^{n/2}-q})$-approximation algorithm for the non-degenerate MAX-MQ in $\mathbb{F}_q$.

From Theorem **??**, one can observe that for large n, say $n = 128$, $(q+\frac{q^2}{q^{n/2}-q}) \approx q$. This implies that random assignment is the best polynomial-time approximation algorithm for MAX-MQ unless P=NP. Then q is the minimal achievable approximation ratio for MAX-MQ over $\mathbb{F}_q$. This gives the complexity of using approximation algorithms to solve MAX-MQ.

Characteristic set method and applications in cryptanalysis. Efficient algebraic algorithms for solving Boolean equation systems have been developed, such as the Gröbner basis method [3] and the XL algorithm [4]. Moreover, SAT-solvers are also efficient approaches for solving Boolean equation systems [5]. Although these algorithms have good performances for some practical problems, their asymptotic complexities are not better than that of the exhaustive search. For the problem of solving quadratic Boolean equation systems, the fast exhaustive search algorithm proposed in [6] has

* Corresponding author (email: huangzhenyu@iie.ac.cn)
The authors declare that they have no conflict of interest.

the complexity $O(\log_2(n)2^n)$. In [7], by combining exhaustive search and solving linear equations, a hybrid algorithm is proposed, whose complexity is $O(2^{0.841n})$ for quadratic Boolean equations under certain assumptions.

In [8], improved characteristic set (CS) algorithms for solving Boolean equation systems were proposed. The idea of the CS method is reducing an equation system in general form to equation systems in the form of triangular sets. With this method, solving an equation system can be reduced to solving univariate equations in cascaded form. In the case of finite fields, univariate equations can be solved with Berlekamp's algorithm.

To solve Boolean equations, we need to develop efficient CS algorithms for polynomial systems in the ring $\mathbb{R}_2 = \mathbb{F}_2[x_1, \ldots, x_n]/(\boldsymbol{H})$ where $\boldsymbol{H} = \{x_1^2 - x_1, \ldots, x_n^2 - x_n\}$. Due to the special property of $\mathbb{R}_2$, the proposed CS methods are more efficient and have better properties than the general CS method.

Let $\mathrm{Zero}(\boldsymbol{P})$ denote the common zero of a Boolean equation system $\boldsymbol{P}$ in $\mathbb{F}_2^n$. By the algorithms TDCS and MFCS proposed in [8], one can obtain polynomial sets $\mathcal{A}_j$, such that $\mathrm{Zero}(\boldsymbol{P}) = \cup_i \mathrm{Zero}(\mathcal{A}_i)$, where $\mathrm{Zero}(\mathcal{A}_i) \cap \mathrm{Zero}(\mathcal{A}_j) = \emptyset$ for any $i \neq j$. Suppose $\mathcal{A}_i = \{T_1, T_2, \ldots, T_{p_i}\}$. Then each T_j can be written as $x_{c_j} + U_j(x_1, x_2, \ldots, x_{c_j - 1})$, where U_i is a Boolean polynomial and $c_1 < c_2 < \cdots < c_{p_i}$. It is obvious that $|\mathrm{Zero}(\mathcal{A}_i)| = 2^{n-p_i}$, hence $|\mathrm{Zero}(\boldsymbol{P})| = \sum_i 2^{n-p_i}$.

The complexity of TDCS for solving Boolean equation systems is $O(2^{n \log_2(m)})$, where n is the number of variables and m is the number of equations in $\boldsymbol{P}$ [8]. Algorithm MFCS is a multiplication-free CS method, where the size growth problem of the former CS algorithms is well solved. The complexity of MFCS for solving quadratic Boolean equation systems is bounded by $O(n(m+n)2^n)$ [9].

The complexity of MFCS is the same as that of the exhaust search but is much faster for many important practical problems. Several groups of experimental results in [8] show that MFCS is quite efficient in solving Boolean equation systems generated from stream ciphers and automated reasoning. One experiment is about the problem of recovering the internal states of the stream cipher Bivium-A. One can recover the internal states of Bivium-A from 700 bits of keystream by MFCS in average 49.3 s for 100 instances, while on the same platform the Gröbner Basis method (F4 in Magma) fails to solve the equations under the same condition and solve the equations with 300–500 s by using 17000–20000 bits of keystream. Another problem was proposed by Cook in SAT 2004: for two $n \times n$ Boolean matrices A and B, prove $B \cdot A = I$ from $A \cdot B = I$ by solving Boolean equations. The former best result is that the problem of $n = 5$ can be solved by SAT-solvers in about 800–2000 s. By MFCS, one can solve the problem of $n = 6$ in 166 s on a PC with 2.7 GHz i7 CPU.

Compared to the Gröbner Basis method and SAT-solvers, it is more suitable to combine the incremental strategy with MFCS. In [10], based on MFCS, an algorithm called ISBS is proposed to solve Boolean equation systems with noises. The problem of solving Boolean equations with noises is called MAX-PoSSo, which is finding the solution satisfying the maximal number of equations for a given noisy equation system. Most methods of solving MAX-PoSSo, such as the MIP solvers [11], search all the possible values of variables. The idea of ISBS is searching all the possible noises with backtracking and incrementally solving the corresponding systems by MFCS. The experimental results in [10] show that for systems generated from the cold boot key recovery problem of block ciphers AES and Serpent, ISBS is much more efficient than the MIP solvers.

References

1 Håstad J. Some optimal inapproximability results. J ACM, 2001, 48: 798–859
2 Zhao S, Gao X S. Minimal achievable approximation ratio for MAX-MQ in finite fields. Theor Comput Sci, 2009, 410: 2285–2290
3 Faugère J C. A new efficient algorithm for computing Gröner bases without reduction to zero (F5). In: Proceedings of International Symposium on Symbolic & Algebraic Computation (ISSAC), Lille, 2002. 75–83
4 Courtois N, Klimov A, Patarin J, et al. Efficient algorithms for solving over-determined systems of multivariate polynomial equations. In: Advances in Cryptology– EUROCRYPT. Berlin: Springer, 2000. 392–407
5 Mcdonald C, Chernes C, Pieprzyk J. Attacking Bivium With MiniSat. Cryptology ePrint Archive Report 2007/040. 2007
6 Bouillaguet C, Chen H C, Cheng C M, et al. Fast exhaustive search for polynomial systems in $\mathbb{F}_2$. In: Cryptographic Hardware and Embedded Systems. Berlin: Springer, 2010. 203–218
7 Bardet M, Faugére J C, Salvy B, et al. On the complexity of solving quadratic boolean systems. J Complex, 2013, 29: 53–75
8 Gao X S, Huang Z. Characteristic set algorithms for equation solving in finite fields. J Symb Comput, 2012, 47: 655–679
9 Huang Z Y, Sun Y, Lin D D. On the efficiency of solving boolean polynomial systems with the characteristic set method. ArXiv:1405.4596, 2014
10 Huang Z Y, Lin D D. A new method for solving polynomial systems with noise over $\mathbb{F}_2$ and its applications in cold boot key recovery. In: Selected Areas in Cryptography. Berlin: Springer, 2012. 16–33
11 Albrecht M, Cid C. Cold boot key recovery by solving polynomial systems with noise. In: Applied Cryptography and Network Security. Berlin: Springer, 2011. 57–72