Information Sciences 402 (2017) 91–104

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Efficient probabilistic algorithm for estimating the algebraic properties of Boolean functions for large n

Yongzhuang Wei[a,b,1], Enes Pasalic[c,2], Fengrong Zhang[d,3,*], Samir Hodžić[e]

[a] Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin 541004, P.R. China
[b] Guangxi Cooperative Innovation Center of Cloud Computing and Big Data, Guilin University of Electronic Technology, Guilin 541004, P.R. China
[c] University of Primorska, FAMNIT and IAM, Koper, Slovenia
[d] School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, P.R. China
[e] University of Primorska, FAMNIT, Koper, Slovenia

ARTICLE INFO

Article history:
Received 2 July 2016
Revised 18 November 2016
Accepted 21 March 2017
Available online 22 March 2017

Keywords:
Stream ciphers
Fast algebraic attacks
Time complexity
Algebraic immunity

ABSTRACT

Although several methods for estimating the resistance of a random Boolean function against (fast) algebraic attacks were proposed, these methods are usually infeasible in practice for relatively large number of input variables n (for instance $n \geq 30$) due to increased computational complexity. An efficient estimation of the resistance of Boolean functions, with relatively large number of inputs n, against (fast) algebraic attacks appears to be a rather difficult task. In this paper, the concept of partial linear relations decomposition is introduced, which decomposes any given nonlinear Boolean function into many linear (affine) subfunctions by using the disjoint sets of input variables. Based on this result, a general probabilistic decomposition algorithm for nonlinear Boolean functions is presented which gives a new framework for estimating the resistance of Boolean function against (fast) algebraic attacks. It is shown that our new probabilistic method gives very tight estimates (lower and upper bound) and it only requires about $O(n^2 2^n)$ operations for a random Boolean function with n variables, thus having much less time complexity than previously known algorithms.

1. Introduction

Boolean functions play an important role in the design of symmetric encryption algorithms, more precisely in certain designs of stream ciphers. For instance, nonlinear filter generator and combination generator are two typical representative of hardware oriented design schemes, which consist of single or multiple linear feedback shift registers (LFSRs) and a nonlinear Boolean function. The security of these LFSR-based stream ciphers heavily relies on the algebraic properties of the

* Corresponding author.
E-mail addresses: walker_wei@msn.com (Y. Wei), enes.pasalic6@gmail.com (E. Pasalic), zhfl203@cumt.edu.cn, zhfl203@163.com (F. Zhang), samir.hodzic@famnit.upr.si (S. Hodžić).

[1] Yongzhuang Wei is supported in part by the Natural Science Foundation of China (61572148), in part by the Guangxi Natural Science Found (2015GXNSFGA139007), in part by the project of Outstanding Young Teachers Training in Higher Education Institutions of Guangxi.
[2] Enes Pasalic is partly supported by the Slovenian Research Agency (research program P3-0384 and research project J1-6720).
[3] Fengrong Zhang is supported in part by National Science Foundation of China(61303263), and in part by the Fundamental Research Funds for the Central Universities (Grant No. 2015XKMS086), and in part by the China Postdoctoral Science Foundation funded project (Grant No. 2015T80600).

http://dx.doi.org/10.1016/j.ins.2017.03.025
0020-0255/

used Boolean function. Over the last decades, Boolean functions satisfying some particular cryptographic properties (such as high nonlinearity, high algebraic immunity (AI) etc.) have been studied [3,9–12,17–19].

Algebraic attacks (AA) and fast algebraic attacks (FAA) were respectively proposed in [4,5], which are two famous and powerful attacks that are easily applied to LFSR-based stream ciphers. The core idea behind the two attacks can be summarized as follows. The first step is to set up a low degree algebraic system of multivariate equations in the secret key/state bits, where the degree of these equations is closely related to the algebraic properties of the used nonlinear Boolean function. The second step is to solve the system of equations and recover the secret key/state bits. Whereas the second step is well elaborated and understood, the first step of finding low degree multivariate equations for relatively large number of input variables n is still an open problem due to complexity issues.

The concept of algebraic immunity for an arbitrary Boolean function f was introduced in [15] and it reflects the resistance of a Boolean function f against AA. More precisely, this criterion measures the minimum algebraic degree of its annihilators, i.e., $AI_f = \min_{\deg(g)}\{A(f), A(f\oplus 1)\}$, where $A(f) = \{g : fg = 0, g \neq 0\}$ and $A(f\oplus 1) = \{g : (f\oplus 1)g = 0, g \neq 0\}$. It was shown that an optimal resistance of a Boolean function f against AA is achieved if $AI_f = \lceil n/2\rceil$. On the other hand, a Boolean function with an optimal AI still cannot adequately ensure a good resistance against FAA that use the existence of the function pairs (g, h) (with algebraic degree $\deg(g)$ and $\deg(h)$ respectively) such that $fg = h$ and $\deg(g) + \deg(f)$ is not large [6,16]. The value of $\deg(g) + \deg(h)$ measures the resistance of a Boolean function against FAA. An optimal resistance of Boolean functions (used in LFSR-based stream ciphers) against FAA implies that the minimum values of $\deg(g) + \deg(h)$ is always equal to n for any function pairs (g, h) such that $fg = h$, though such functions are very rare. In addition, it was shown that for balanced Boolean functions $\deg(g) + \deg(h) \geq n$ if and only if either $n = 2^k$ or $n = 2^k + 1$ for some positive integer k [13].

During the past decade, an efficient evaluation of the resistance of nonlinear Boolean functions against AA and FAA has been addressed in many works due to a great significance of these estimates from both the design and cryptanalysis point of view. At EUROCRYPT 2003, the first algorithm for determining the existence of annihilators of degree d of a Boolean function with n variables was proposed in [4]. Its time complexity is about $O(D^3)$ operations, where $D = \sum_{i=0}^{d}\binom{n}{i}$. At FSE 2006, an algorithm for checking the existence of annihilators or multiples of degree less than or equal to d was introduced in [7] with time complexity of about $O(n^d)$ operations for an n-variable Boolean function. At EUROCRYPT 2006, based on the multivariate polynomial interpolation, Armknecht et al. [1] proposed an algorithm for computing $AI = d$ of a Boolean function with n variables [1] requiring $O(D^2)$ operations, where $D = \sum_{i=0}^{d}\binom{n}{i}$. Moreover, an algorithm for determining the immunity against FAA was also presented running in time complexity of about $O(D^2E)$ operations for an n-variable Boolean function, where $E = \sum_{i=0}^{e}\binom{n}{i}$ and d is generally much smaller than e, $(\deg(g), \deg(h)) = (d, e)$. At ACISP 2006, an algorithm to evaluate the resistance of Boolean functions against FAA was developed in [2], whose time complexity is about $O(DE^2 + D^2)$ operations for an n-variable Boolean function. At INDOCRYPT 2006, based on the Wiedemann's algorithm, Didier proposed a new algorithms to evaluate the resistance of an n-variable Boolean functions against AA and FAA in [8] with time complexity of about $O(n^{2n}D)$ operations and a memory complexity of about $O(n^{2n})$. Finally, Jiao et al. [14] revised the algorithm of [1] to compute the resistance against AA and FAA, reducing the complexity to $O(D^{2\pm\varepsilon})$ operations, where $\varepsilon \approx 0.5$ and D is the same as above.

Despite the development of the above mentioned algorithms, the exact evaluation of the algebraic properties of a Boolean function remains infeasible for relatively large input variables n (for instance $n \geq 30$). For instance, in order to estimate exactly the resistance of a random Boolean function with 30 variables against AA and FAA, the best known algorithm of [14] still requires $\binom{n}{d}^{2.5} = \binom{30}{15}^{2.5} \approx 2^{68}$ operations, for $n = 30$ and $d = 15$. It appears to be a rather difficult task to efficiently estimate the resistance of Boolean function (with relative large input variables n) against AA and FAA. The purpose of this paper is to present an efficient probabilistic algorithm for determining the resistance of a random Boolean function against AA and FAA. A suitable choice of input parameters gives a high success rate of the algorithm so that the estimates are correct with probability very close to one. The algorithm employs partial linear relations, derived form the decomposition of an arbitrary nonlinear Boolean function into many small partial linear subfunctions by using the disjoint sets of input variables. A general probabilistic decomposition algorithm for nonlinear Boolean functions is given along with the sufficient conditions regarding the existence of low degree annihilators (or multipliers). This probabilistic algorithm provides a new framework for estimating the resistance of Boolean function against AA and FAA requiring only about $O(n^2 2^n)$ operations (for an n-variable Boolean function), thus offering much less complexity at the price of being probabilistic. The lower and upper bound on AI and FAA that we derive appears to be very tight for randomly selected Boolean functions thus giving a close estimate of the algebraic properties for large n where due to computational complexity the deterministic algorithms cannot be applied. Several examples are provided justifying the tightness of our bounds when compared to the actual algebraic properties of a given function for relatively small values of n for which the deterministic algorithms could be applied.

The rest of the paper is organized as follows. In Section 2, some basic definitions and notations are recalled. In Section 3, a new concept of partial linear relations decomposition is introduced, and then a general dissection algorithm for nonlinear Boolean functions is proposed. An efficient algorithm for determining the resistance of Boolean functions (with relatively large input variables n) against AA and FAA is descried in Section 4. Finally, some concluding remarks are given in Section 5.