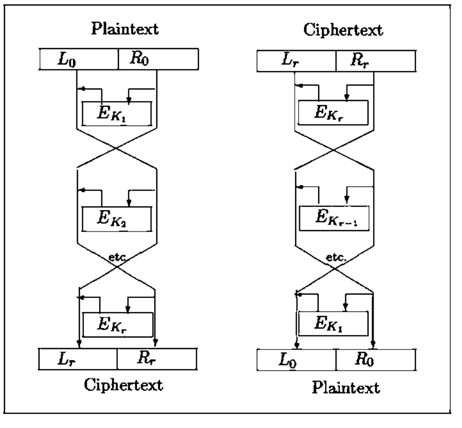
WHAT IS...

a cryptographic Boolean function?

Lilya Budaghyan and Pantelimon Stănică

As one of the most fundamental objects investigated in pure and applied mathematics and computer science, the notion of a Boolean function was introduced about 150 years ago in the context of fundamental mathematics and mathematical logic by an English mathematician George Boole (1815-1864). Boole's treatment of algebra of logic (now known as Boolean algebra) in his *The laws of thought* [3] laid the foundation for the design of modern digital computer circuits. For positive integers n and m, a vectorial, or $(n, m)-$Boolean function is a map from the finite field $\mathbb{F}_{2^n}$ (or the vector space $\mathbb{F}_2^n$) to $\mathbb{F}_{2^m}$ (or $\mathbb{F}_2^m$) ($m = 1$ corresponds to a Boolean function). Since the middle of the twentieth century, with the rapid development of information and communication technology, Boolean function theory has become an important tool for solving problems of analysis and synthesis of discrete devices which transform and process information, in particular, in cryptography. In this article, we give an overview of the main concepts and problems in the area of cryptographic Boolean functions from the last 40 years.

To respond to a need for ensuring security of electronic data, in 1973 the US National Bureau of Standards (now, National Institute of Standards and Technology – NIST) issued a call for strong encryption primitives that would become a government-wide standard. Since all submissions were unsuitable that year, NBS re-issued the call and in 1974 a team from IBM submitted a modified version of their cipher Lucifer (designed in 1971). After a year of intense collaboration between IBM and NSA, Data Encryption Standard (DES) was approved, and then published in 1975 in the Federal Register [20]. The call, submission and collaborative effort to approve and publish DES marks the start of intense work on cryptography in academia.



Feistel Cipher with r rounds; Reprinted from [10].

In modern society, exchange and storage of information in an efficient, reliable and secure manner is of fundamental importance. Cryptographic primitives are used to protect information against eavesdropping, unauthorized changes and other misuse. The security of any symmetric cipher relies on components like the substitution S-boxes.

Lilya Budaghyan is a research director of the project "Optimal Boolean Functions" and a head of Selmer Center—Reliable Communication Group at the University of Bergen, Norway.

Her email address is `Lilya.Budaghyan@uib.no`.

Pante Stănică is professor of Applied Mathematics and Associate Chair for Research at Naval Postgraduate School in Monterey, California, USA. His email address is `pstanica@nps.edu`.

Communicated by Notices *Associate Editor Emilie Purvine and Cesar E. Silva.*

For permission to reprint this article, please contact: `reprint-permission@ams.org`.

DOI: https://doi.org/10.1090/noti/1780

As an example, let the Feistel scheme (one of the two major schemes in addition to Substitution-Permutation Networks) [10] for DES depicted in Fig. 1: we split the input block into two halves L_{i-1} and R_{i-1}, $1 \le i \le r$, and go through a sequence of "rounds" with $L_i = R_{i-1}$, $R_i = L_{i-1} \oplus E_{K_i}(R_{i-1}) = R_{i-2} \oplus E_{K_i}(R_{i-1})$, where $\oplus$ is the binary addition and $E_K(R)$ is an encryption function using a key K on the block R).

A substitution or an S-box (part of the E_K's above) is nothing else than a *cryptographic Boolean function* (CBF), that is, a Boolean function possessing some optimal (or near optimal) cryptographic properties. The S-boxes introduce confusion (as defined by Shannon in 1949 in his seminal "Communication Theory of Secrecy Systems" [21], where he also considered diffusion – "mixing" the properties of the plaintext into the ciphertext) in the system. Among the most important classes of CBF are the so-called bent, almost perfect nonlinear, almost bent, crooked, correlation immune and resilient functions [4, 7, 8, 10, 16, 23]. Defined for cryptographic purposes starting in the 70's, these turned out to be known for many decades in the context of other domains of mathematics and information theory such as coding theory, sequence design, commutative algebra, combinatorics and finite geometry. During recent years more applications of these beautiful objects have been found, and, we believe, many more are still to be discovered. In spite of the universality of these functions and long history of study not much is known about some of these classes (the case of almost perfect nonlinear and almost bent functions, for instance) and just a few families of such functions have been constructed.

For most cryptographic attacks on block ciphers (like DES, or the current standard, Advanced Encryption Standard – AES) there are certain properties of functions which measure the resistance of the S-box to these attacks. One of the most efficient cryptanalysis tools for block ciphers, the differential attack introduced by Biham and Shamir [2], is based on the study of how differences in an input can affect the resulting differences in the output. An (n, m)-function F is called differentially δ-uniform if the derivative equation $D_aF(x) = F(x+a) - F(x) = b$ has at most δ solutions for every $a \in \mathbb{F}_{2^n}^*$ and $b \in \mathbb{F}_{2^m}$. Functions with the smallest possible differential uniformity contribute an optimal resistance to the differential attack. In this sense, differentially 2^{n-m}-uniform functions, called perfect nonlinear (PN), are optimal. However, PN functions exist only for n even and $m \le n/2$. For the important case of $n = m$, differentially 2-uniform functions, called almost perfect nonlinear (APN), have the smallest possible differential uniformity. Another powerful attack on block ciphers is the linear cryptanalysis proposed by Matsui [15], which is based on finding affine approximations to the action of a cipher. The nonlinearity of an (n, m)-function F is the minimum Hamming distance between all components of F and all affine Boolean functions in n variables. The nonlinearity quantifies the level of resistance of the function to the linear attack: the higher the nonlinearity, the better the resistance of F. Functions whose nonlinearity reaches the universal upper bound are called bent. All bent functions are also PN and vice versa, that is, these functions have optimal resistance against both linear and differential attacks. If $m = n$, PN (or bent) functions do not exist, but if n is also odd, functions with the best possible nonlinearity are called almost bent (AB). When n is even the tight upper bound on nonlinearity is still to be determined. All AB functions are APN, but the converse is not true in general.

There are very few known classes of bent functions in the almost 40 years they have been studied (or more, as difference sets in elementary Abelian 2-groups), and we mention here the constructions of Rothaus, Maiorana, McFarland, Dillon and Carlet (see the excellent survey [9]). The actual count is known for all even dimensions ≤ 8. For example, for $n = 8$, there are approximately $2^{106.2}$ bent functions (in the class of all $2^{2^8} = 2^{256}$ Boolean functions), but the known bent constructions account only for about 2^{85} of those, so there are many more constructions we have not been able to find yet. In spite of not being balanced, there are many properties on bent functions that make them so desirable. For example, the derivative of a bent function is always balanced in any direction, which is quite important for differential cryptanalysis; they have excellent propagation characteristics, since if one changes any nonzero number of bits in the input of a bent function, the output changes with probability $1/2$, a desirable property for (fast) correlation attacks. These are just a handful of cryptographic properties connected to bentness, but bent functions are also useful for coding theory. Since they generate difference sets, they can be used to construct a symmetric design. They also give rise to the Kerdock codes, and some other robust error detecting codes. For even dimension, bent functions attain the covering radius of the first-order Reed–Muller code. They have flat absolute values with respect to the Walsh-Hadamard transform (a discrete Fourier transform). One could consider different transforms, and impose flatness of the corresponding values. For example, the choice of the nega-Hadamard transform is motivated by local unitary transforms that play an important role in the structural analysis of pure n-qubit stabilizer quantum states, as argued by Riera and Parker [18] in 2006.

The nonlinearity and the differential uniformity (and, therefore, bentness, PNness, APNness and ABness), are invariant under affine, extended affine and CCZ-equivalences (in increasing order of generality). Two functions F and F' are affine equivalent (EA), if $F' = B \circ F \circ A$ (A, B are affine permutations) and extended affine equivalent (EA-equivalent), if $F' = B \circ F \circ A + C$ (C is just affine). They are called CCZ-equivalent if their graphs are affine equivalent. CCZ-equivalence is the most general known equivalence relation of functions which preserves the nonlinearity and differential uniformity. Although CCZ-equivalence is a very powerful method for constructing functions, very little is currently known about CCZ-equivalence classes of the known APN functions (in particular, of all power APN functions except Gold case) whether they are larger than EA-equivalence classes (and EA-equivalence classes of their inverses when a function is a permutation). In general, identifying situations in which CCZ-equivalence reduces to EA-equivalence is useful. It is already known that for all Boolean functions and for all bent functions CCZ-equivalence and EA-equivalence coincide, while for quadratic power APN and AB functions CCZ-equivalence is strictly more general.

The classification of bent, APN and AB functions is a hard open problem. Complete classification for APN and AB functions over $\mathbb{F}_{2^n}$ is known only for $n = 5$. For bent functions the classification is done nowadays up to $n = 8$. The reason the computation is difficult is because the space of all Boolean functions in n variables is doubly exponential, that is, its cardinality is 2^{2^n}. There are only a few infinite classes of APN and AB functions known (e.g., four classes of AB power functions, six classes of APN power functions, eleven classes of quadratic APN and AB functions constructed recently; for $6 \leq n \leq 9$ there is a large list of quadratic APN and AB functions, but infinite classes are still to be determined). It was conjectured by Dobbertin in 1999 that classification of APN and AB power functions is complete, but the proof is yet to be determined.

Recent advances in APN functions have made a prominent impact on the theory of commutative semifields. In spite of considerable work, only two previously known infinite families of commutative semifields (that are not finite fields) of order p^n, p an odd prime, were constructed by Dickson [11] and Albert [1]. A few new such infinite families have been constructed using families of quadratic APN functions (see for instance [6]). Hence, results on the classification of APN and PN functions have important consequences on our understanding of commutative semifields and projective planes.

Quadratic AB permutations are of interest for combinatorial analysis. A mapping F from $\mathbb{F}_{2^n}$ to itself is called crooked if $\{D_aF(x) : x \in \mathbb{F}_{2^n}\}$ is the complement of a hyperplane, for all $a \in \mathbb{F}_{2^n}^*$. Every crooked function gives rise to a distance regular rectagraph (a graph without triangles in which every pair of vertices at distance 2 lies in a unique 4-cycle) of diameter 3, and every quadratic AB permutation (taking 0 value at 0) is crooked. The converse is not known, that is, whether a crooked function is necessarily a quadratic AB permutation. There are not too many constructions of rectagraphs known, especially rectagraphs of small diameter. Hence such functions provide not only interesting building blocks for symmetric cryptosystems but also provide new distance regular rectagraphs. Nowadays only two families of crooked functions are known. One constructed in 1968 by Gold [13] in his investigation of sequence designs and rediscovered in 1993 by Nyberg [17] in the context of cryptography, giving rise to Preparata graphs. The other one is the family of binomials constructed in 2008 by Budaghyan, Carlet and Leander [5].

For n odd all power APN functions are permutations. Certainly, one wonders whether there exist APN permutations over $\mathbb{F}_{2^n}$ in the case when n is even, and in [12] Dillon et al. constructed an APN permutation for $n = 6$ applying CCZ-equivalence to a quadratic APN function. It is a surprising result since quadratic APN functions themselves cannot be permutations, and it is one of the few results showing that CCZ-equivalence changes properties of functions considerably. Now it is a big challenge to construct APN permutations for larger n, or better yet, construct infinite families of such functions.

We end our Boolean functions snapshot with an extension that has gained increasing attention in the last few years, stepping off outside of the binary world. We mention here the generalized Boolean functions, defined on $\mathbb{F}_{2^n}$ (or $\mathbb{F}_2^n$) with values in integers modulo $q \geq 2$, or p-ary Boolean functions, that is, functions defined on $\mathbb{F}_{p^n}$ with values in $\mathbb{F}_{p^m}$, where p is a prime number. An important topic here is the study of generalized bent functions, which are those whose generalized (using powers of a complex root of 1) Walsh-Hadamard transform absolute values are constant. Multicarrier communications and modulation schemes is a major research topic that has been around for slightly over two decades, rooted in the explosive growth of the Internet, which in turn has increased the demand for wired and wireless high data rates. A major problem with the multicarrier modulation in general and the OFDM (Orthogonal Frequency Division Multiplexing) system in particular is the high peak-to-average power ratio (PAPR) that is inherent in the transmitted signal. Schmidt in 2009 [19] showed that one can achieve a perfect modulation (with respect to the PAPR) if one uses the sequence associated to a generalized bent Boolean function. Researchers' attempts in constructing or describing these generalized bent functions culminated in 2017 with their complete characterization in terms of classical bent Boolean functions in two independent works [14, 22].

In this article we gave a brief overview of some of the main concepts, results and the problems we face in cryptographic Boolean functions' research. It is by no means all inclusive, rather it is an attempt to whet the appetite of the mathematician as well as the practitioner to delve more in this area.

References

[1] A. A. Albert, On nonassociative division algebras, *Trans. Amer. Math. Soc.* 72 (1952), 296–309; MR0047027.

[2] E. Biham, A. Shamir, *Differential Cryptanalysis of the Data Encryption Standard*, Springer–Verlag, 1993.

[3] G. Boole, *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*, Macmillan, 1854; Reprinted with corrections, Dover Publications, New York, NY, 1958.

[4] L. Budaghyan, *Construction and Analysis of Cryptographic Functions*, Springer–Verlag, 2014; MR3290040.

[5] L. Budaghyan, C. Carlet, G. Leander, Two classes of quadratic APN binomials inequivalent to power functions, *IEEE Trans. Inf. Theory* 54:9 (2008), 4218–4229; MR2450779.

[6] L. Budaghyan and T. Helleseth, New perfect nonlinear multinomials over $\mathbb{F}_p^{2k}$ for any odd prime p, *Proc. Int. Conf. on Sequences and Their Applic. SETA 2008*, LNCS 5203, pp. 403–414, 2008; MR2646419.

[7] C. Carlet, *Boolean functions for cryptography and error correcting codes*, In: Y. Crama, P. Hammer (eds.), Boolean Methods and Models, Cambridge Univ. Press, Cambridge, pp. 257–397, 2010.

[8] C. Carlet, *Vectorial Boolean Functions for Cryptography*, In: Y. Crama, P. Hammer (eds.), Boolean Methods and Models, Cambridge Univ. Press, Cambridge, pp. 398–472, 2010.

[9] C. Carlet, S. Mesnager, Four decades of research on bent functions, *Designs, Codes & Cryptogr.* 78 (2017), 5–50; MR3440222.

[10] T. W. Cusick, P. Stănică, *Cryptographic Boolean Functions and Applications* (Ed. 2), Academic Press, San Diego, CA, 2017; MR3644644.

[11] L. E. Dickson, On commutative linear algebras in which division is always uniquely possible, *Trans. Amer. Math. Soc.* 7 (1906), 514–522; MR1500764.

[12] K. A. Browning, J. F. Dillon, M. T. McQuistan, A. J. Wolfe, An APN Permutation in Dimension Six, *Post-proc. 9-th Int. Conf. Finite Fields and Their Applic. Fq'09*, Contemporary Math., AMS, v. 518, pp. 33–42, 2010; MR2648537.

[13] R. Gold, Maximal recursive sequences with 3-valued recursive crosscorrelation functions, *IEEE Trans. Inf. Theory* 14 (1968), 154–156.

[14] T. Martinsen, W. Meidl, S. Mesnager, P. Stănică, Decomposing generalized bent and hyperbent functions, *IEEE Trans. Inform. Theory* 63:12 (2017), 7804–7812; MR3734198.

[15] M. Matsui, Linear cryptanalysis method for DES cipher, *Adv. Crypt. - EUROCRYPT'93*, 386–397.

[16] S. Mesnager, *Bent functions. Fundamentals and Results*, Springe–Verlag, 2016; MR3526041.

[17] K. Nyberg, Differentially uniform mappings for cryptography, *Adv. Crypt. – EUROCRYPT'93*, LNCS 765 (1994), 55–64; MR1290329.

[18] C. Riera and M. G. Parker, Generalized bent criteria for Boolean functions, *IEEE Trans. Inf. Theory* 52:9 (2006), 4142–4159; MR2298538.

[19] K.-U. Schmidt, Quaternary Constant-Amplitude Codes for Multicode CDMA, *IEEE Trans. Inf. Theory* 55:4 (2009), 1824–1832; MR2582768.

[20] B. Schneier, *Applied Cryptography: Protocols, Algorithms and Source Code in C*, John Wiley & Sons, Inc., 2015; MR3587912.

[21] C. Shannon, Communication Theory of Secrecy Systems, Bell System Technical Journal 28:4, 656–715, 1949.

[22] C. Tang, C. Xiang, Y. Qi, K. Feng, Complete characterization of generalized bent and 2^k-bent Boolean functions, *IEEE Trans. Inform. Theory* 63:7 (2017), 4668–4674; MR3666982.

[23] N. Tokareva, *Bent Functions. Results and Applications to Cryptography*, Academic Press, San Diego, CA, 2015; MR3362707.

Lilya Budaghyan

Pantelimon Stănică

Credits

Figure 1 was reprinted from [10] with permission.

Photos of Lilya Budaghyan and Pantelimon Stănică are courtesy of the authors.