**COMPUTER SCIENCE | RESEARCH ARTICLE**

**Design of substitution nodes (S-Boxes) of a block cipher intended for preliminary encryption of confidential information**

**Ardabek Khompysh**, **Nursulu Kapalova**, **Kunbolat Algazy**, **Dilmukhanbet Dyusenbayev** and **Kairat Sakan**

**Abstract:** This paper considers a new method for obtaining an S-box, which is one of the nonlinear transformations used in modern block-symmetric cipher systems. This method is based on modular arithmetic, that is, exponentiation modulo polynomial in extended Galois fields GF(2^8). The indicators and criteria of efficiency of the obtained S-box (balance, Hamming distance, distribution criteria, autocorrelation, algebraic immunity, cyclic structure of the S-box) are analyzed. The cryptographic characteristics are presented in comparison with the substitution boxes of known modern block ciphers. In addition, the resulting S-box was investigated by the methods of linear and differential cryptanalysis. In the future, the proposed S-box will be used in the developed encryption algorithm designed for the precryption of confidential information.

**Subjects:** Algorithms & Complexity; Cryptology; Computing & IT Security

**Keywords:** block ciphers; S-box; Boolean functions; nonlinear transformation; strict avalanche criterion (SAC); algebraic immunity; linear and differential cryptanalysis

---

**ABOUT THE AUTHOR**

**Ardabek Khompysh** — Research Scientist at the Institute of Information and Computational Technologies, PhD in specialty «Information Security Systems». Theme of scientific research development and research of reliable algorithms for generating pseudorandom sequences used to generate key sequences for symmetric block and stream encryption, as well as the development of cryptographic information protection algorithms using modular arithmetic. Published 20 works in foreign and domestic publications. Received two certificates of state registration of the intellectual property of the Republic of Kazakhstan.

**Kunbolat Algazy** — Research Scientist at the Institute of Information and Computational Technologies, PhD in specialty «Information Security Systems». Published 37 works in foreign and domestic publications. Received four certificates of state registration of the intellectual property of the Republic of Kazakhstan. Theme of scientific research development and research of symmetric block cipher algorithms.

---

**PUBLIC INTEREST STATEMENT**

To protect algorithms from various cryptanalysis methods, S-boxes must have a number of cryptographic properties and satisfy a number of criteria. Currently, the main attacks for which the properties of S-boxes used in cryptographic algorithms are important are attacks based on linear, differential, and algebraic cryptanalysis methods.
1. Introduction
Symmetric cryptographic primitives are widely used due to their high performance and low implementation complexity (Schneier, 1996). When guaranteeing confidentiality using block ciphers, symmetric primitives are used to ensure the integrity of information based on message authentication codes and hash functions, as components of an electronic digital signature to protect authenticity, generate pseudo-random sequences, as part of authentication protocols (Menezes et al., 1997), etc. In accordance with the well-known Shannon principles (Shannon, 1949), such algorithms use nonlinear operations for confusion and linear transformations for diffusion. Consistent, multiple applications of confusion and diffusion allow for a high level of cryptographic strength. Nonlinear substitution nodes for modern symmetric primitives are typically implemented as substitution tables or S-boxes. Considering that most modern block algorithms (AES, Kuznyechik, Kalyuna, BelT, etc.) use a single linear modulo 2 addition operation to introduce round keys and combine inter-round values, S-box is the only element that determines the nonlinearity of the encryption transformation and the level of its resistance to cryptanalytic attacks (Kuznetsov et al., 2015). The required number of rounds of block ciphers is calculated based on ensuring resistance to known types of cryptographic analysis under the condition of the specified properties of nonlinear substitution nodes. Many stream algorithms, cryptographic hash functions, and pseudo-random sequence generators are based on block ciphers or their structural elements. Thus, the cryptographic strength of most modern symmetric primitives largely depends on the properties of the S-boxes used (Kazimirov & Oleinikov, 2010).

The most advanced key cryptosystems are based on the idea of cipher production, which has already become traditional (Shannon, 1949) and represent a class of cryptosystems that repeat a complex operation transforming plain text into ciphertext. Each such repetition (iteration) is known as a cipher cycle (Connor, 1994). A complex (compound) operation performed in each cycle is usually a combination of a set of primitive operations, such as shift, linear transformation, modulo addition, and substitution. In an appropriate combination, these transformations should implement the concept of building such ciphers, which, as is known, is that a combination of permutation and substitution operations of separately weak transformations can lead to a cryptographically strong nonlinear transformation if it is applied a sufficient number of times. Substitution operations in many ciphers act in this case as the main nonlinear element of the cyclic transformation (Dolgov et al., 2009).

The nonlinearity of the elements that make up symmetric block crypto schemes has a great impact on the strength of the information security subsystem in modern telecommunication systems. The selection of quality substitutions (S-boxes) is considered the most difficult aspect of block cipher design. To date, the existing tools for assessing the strength of substitutions do not contain any method that determines the best S-box in terms of counteracting various cryptanalytic attacks and techniques for its hardware and/or software implementation (Panasenko, 2009; Zenzin & Ivanov, 2002). In modern algorithms AES and Kuznyechik, instead of fixed substitutions, linear transformations are used to map an established S-box (often the only one) to another box of a sufficiently large set of equivalent ones (Grigoryan & Ngi, 2019).

Therefore, significant efforts of researchers are aimed at studying properties and constructing substitutions with high cryptographic indicators. One of the most popular for describing and studying the properties of S-boxes is the mathematical apparatus of linear algebra and, in particular, the apparatus of Boolean functions (Pasalic et al., 2001).

The article presents a novel way to design an S-box based on modular arithmetic, that is, exponentiation modulo polynomial in extended Galois fields $GF(2^8)$.

2. General information about substitution boxes
An S-box (Substitution-box) underlies any block cipher and is a source of nonlinearity. The study of a cipher for strength, as a rule, begins with a study of the properties of its S-boxes.
In general, an $m \times n$ S-box maps $Z_2^m \rightarrow Z_2^n$. The algorithms GOST R 34.13-2015 and AES use $8 \times 8$ S-boxes, and DES uses 8 different S-boxes of $6 \times 4$.

Tabular presentation of S-boxes. The simplest way to define S-boxes is a table of values, which indicates where each bit combination goes. Memory sizes required to store such a table are as follows:

\[
\begin{align*}
4 \times 4 &= 16 \\
8 \times 8 &= 256 \\
16 \times 16 &= 65536 \\
32 \times 32 &= 4294967296 \approx 4 \times 109 \\
64 \times 64 &= 18446744073709551616 \approx 2 \times 1019
\end{align*}
\]

Tables are convenient for software implementation, but restrictions are imposed on their size.

There are many competing approaches to the selection of S-boxes, among them four main ones can be distinguished (Babenko & Ischukova, 2006).

Random sampling. It is clear that small random S-boxes are not reliable, but large random S-boxes can be good enough. Random S-boxes with eight or more inputs can be quite strong. The strength of S-boxes increases when they are both random and key-dependent.

Sampling followed by testing. In some ciphers, random S-boxes are generated first, and then their properties are tested for compliance with requirements.

Manual development. Here, the mathematical apparatus is rarely used: S-boxes are created using intuitive techniques. Bart Preneel stated that “...theoretically interesting criteria are not sufficient (for choosing Boolean S-box functions) ...” and “...special design criteria are needed.” (Preneel, 1993).

Mathematical development. S-boxes are created under the laws of mathematics; therefore, they have guaranteed resistance to differential and linear cryptanalysis and good diffusion properties (Kapalova et al., 2020). There have been proposals to combine the “mathematical” and “manual” approaches, but in practice, randomly selected S-boxes and S-boxes with certain properties compete. The advantages of the latter approach include optimization against known attack methods—differential and linear cryptanalysis. In recent years, many approaches to obtaining substitution tables have appeared, for example, a linear fractional transformation (LFT), a cubic fractional transformation (CFT), a heuristic approach, a modular approach, and others. It is also worth noting the work of the following scientists (Chew & Ismail, 2020; Hussain et al., 2012; Naseer et al., 2019; Ozkaynak, 2017; Shah et al., 2011; Shah & Shah, 2018; Zahid et al., 2020; Zahid & Arshad, 2019; Zahid et al., 2019; Y. Zhang, 2018), who are engaged in obtaining dynamic quality S-boxes. This article (Naseer et al., 2019) shows a way to obtain a dynamic S-box by applying substitution-permutation transformations on the input value of the S-box. This (considered) S-box is structurally different from the developed S-box; our proposed algorithm represents a method for obtaining a fixed S-box.

An S-box, in other words, is a mapping of $m$-bit inputs to $n$-bit outputs. S-boxes are part of the transform function and are important for the strength of the encryption algorithm. Any changes to the S-box input should result in similarly random changes in the output. The dependence of the output values on the input should not be linear or easily approximated by linear functions (this property is used when applying linear cryptanalysis; Kapalova & Haumen, 2018). S-boxes are currently used in many symmetric encryption algorithms, such as AES, GOST 28147–89, DES, Twofish, etc., (Biyashev et al., 2021).
3. The proposed method for obtaining nonlinear substitution nodes for symmetric cryptoalgorithms

Let us consider a three-step method for obtaining an S-box.

At the first step, we choose an irreducible polynomial that generates a multiplicative group in the Galois field \(GF(2^8)\) and an irreducible polynomial called the base. The modular exponential operation is performed on the selected polynomial \(P(x)\):

\[ S_i(x) = A(x)^i \mod P(x) = 0.255, \]  
(1)

where \(A(x)\) is an irreducible polynomial called the base, \(P(x)\) is the module (an irreducible polynomial).

At the second step, we consider the coefficients of the polynomial \(S'(x)\) as vectors of length 8, that is, \(S' = (s_0, s_6, s_5, s_4, s_3, s_2, s_1, s_0)\) if \(S'(x) = s_0x^7 + s_1x^6 + s_2x^5 + s_3x^4 + s_4x^3 + s_5x^2 + s_6x + s_7\). Then the operation of addition \(S\) modulo 2 (XOR) with a fixed vector \(B\) is performed:

\[ S' = S \oplus B \]  
(2)

At the third step, we multiply by the matrix \(M\) \((S'_j = M \times S_j)\):

\[
\begin{bmatrix}
  s_0 \\
  s_1 \\
  s_2 \\
  s_3 \\
  s_4 \\
  s_5 \\
  s_6 \\
  s_7
\end{bmatrix} =
\begin{bmatrix}
  1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
  1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
  1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
  1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
  1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
  1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
  1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
  0 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix} \times
\begin{bmatrix}
  s_0 \\
  s_1 \\
  s_2 \\
  s_3 \\
  s_4 \\
  s_5 \\
  s_6 \\
  s_7
\end{bmatrix}
\]

Thus, the result of the calculation by formula (3) is a vector of S-box values.

The developed S-box is calculated once and is not dynamic, the calculation of the computational complexity is considered insignificant. To generate an S-box according to the proposed method, an average of 97,921 mathematical operations are performed.

Example:

Let the following parameters be chosen: \(A(x) = x^3 + x^2 + 1, P(x) = x^8 + x^5 + x^3 + x + 1, B(x) = x^7 + x^5 + x^3 + x\). Substituting the parameters into formulas (1) and (2), we calculate \(S_0(x)\) and \(S'(x)\). For \(i = 0\) we get the values \(S_0(x) = (x^3 + x^2 + 1)^0 \mod (x^8 + x^5 + x^3 + x + 1) = 1 \mod (x^8 + x^5 + x^3 + x + 1)\) and calculate \(S = 00001000101010101 = 101010101\)

Then we multiply the resulting vector \(S\) by the matrix according to the formula (3):

\[
\begin{bmatrix}
  1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
  1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
  1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
  1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
  1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
  1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
  0 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\times
\begin{bmatrix}
  s_0 \\
  s_1 \\
  s_2 \\
  s_3 \\
  s_4 \\
  s_5 \\
  s_6 \\
  s_7
\end{bmatrix} =
\begin{bmatrix}
  1 \\
  0 \\
  1 \\
  0 \\
  1 \\
  1 \\
  0 \\
  1
\end{bmatrix}
\]

\[ = 101010002 \rightarrow 0a_{16} \]
The obtained values after conversion from binary to hexadecimal form are presented in Table 1.

3.1. Software implementation of the method for designing substitution nodes of a block cipher

The proposed method for designing block cipher substitution nodes for the encryption algorithm under development was implemented in software. Figure 1 shows the work window of the program for obtaining an S-box.

4. Evaluation of the strength of nonlinear substitution nodes of an S-box

The use of the mathematical apparatus of vector Boolean functions provides means to simplify the description of the basic elements of symmetric algorithms. Such a representation makes it possible to generalize a set of criteria, including those applied to substitutions, at the same time enabling us to evaluate the correlation, algebraic, and other properties of S-boxes (Maier & Staffelbach, 1990; Maitra & Pasalic, 2001; Sergienko & Moskovchenko, 2007).

If we briefly characterize the possibilities of this approach, then we can note that its basis is the representation of the S-box in the form of a composition of component Boolean functions with the subsequent study of their properties. True, these criteria had to be supplemented with additional restrictions on the maximum permissible values of the elements of the tables of differences and linear approximations, which, however, are also present when using the apparatus of Boolean functions (Millan et al., 1998).

Let $F^\mathbb{F}_2^n$ be the vector space of all binary vectors of length $n$, where $F^\mathbb{F}_2$ is a Galois field with two elements (0,1). Let $n$ and $m$ be two natural numbers, then by an $(n,m)$-function we mean the vector Boolean function $F : F^\mathbb{F}_2^n \rightarrow F^\mathbb{F}_2^m$. Such functions are used in cryptography as nonlinear mappings in pseudo-random generators (stream ciphers) or as substitutions (S-boxes) in symmetric block ciphers (Carlet, 2010). Boolean functions $f_1, \ldots, f_m$ such that $F(x) = (f_1, \ldots, f_m)$ are called coordinate functions of $F$. Linear combinations of coordinate functions with nonzero coefficients are called component functions of $F$.

Obviously, for $m = 1$, a vector Boolean function has one output bit and is equivalent to an ordinary Boolean function. To find the algebraic structure, the vector space often has the structure of a finite field $F_2^n$ with some irreducible polynomial.

Let $f(x) : F^\mathbb{F}_2^n \rightarrow F_2$, where $x = (x_0, x_1, \ldots, x_{n-1})$, is a Boolean function with $n$ variables. Then, the function of finding the Hamming weight is given as

$$ hw(f) = \sum_{x=0}^{2^n-1} f(x). $$

(4)

Let $f(x)$ and $g(x)$ be Boolean functions with $n$ variables. Then, the Hamming distance between the two functions is calculated by the formula:

$$ hd(f,g) = \sum_{x=0}^{2^n-1} f(x) \oplus g(x). $$

(5)

The Algebraic Normal Form (ANF) of a Boolean function has the form:

$$ f(x_0, x_1, \ldots, x_{n-1}) = \alpha \oplus \alpha_0 x_0 \oplus \alpha_1 x_1 \oplus \ldots \oplus \alpha_{n-1} x_{n-1} \oplus \alpha_{01} x_0 x_1 \oplus \alpha_{02} x_0 x_2 \oplus \ldots \oplus \alpha_{(n-2)(n-1)} x_{n-2} x_{n-1} \oplus \ldots \oplus \alpha_{012 \ldots n-1} x_0 x_1 \ldots x_{n-1}. $$

(6)
## Table 1. Generated substitution box

|   | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | A  | B  | C  | D  | E  | F  |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 0 | a8 | b0 | 08 | 6b | cc | 63 | 0a | df | f3 | a0 | 76 | b4 | 92 | b3 | e6 | 1e |
| 1 | 3b | f4 | d4 | db | 69 | 78 | 5c | c9 | a3 | 98 | c1 | 65 | 24 | 38 | 1a | a1 |
| 2 | 2c | fe | a6 | 58 | 53 | 7b | b2 | bc | 54 | 0f | 1f | 61 | be | e0 | 30 | dc |
| 3 | 1d | d5 | 81 | 23 | 4c | b7 | 7c | c6 | 11 | 89 | e5 | f0 | 4e | 03 | 43 | 05 |
| 4 | 6d | e2 | 84 | e3 | de | a9 | ea | 42 | 5f | 27 | d6 | 6f | 56 | bb | 20 | a2 |
| 5 | c2 | 8b | 51 | cf | 8d | 7f | 28 | d4 | 64 | 7e | 72 | 2e | 4a | 99 | 9b | 2f |
| 6 | d3 | af | c4 | a5 | b6 | 26 | 8c | 25 | 62 | 50 | 95 | c7 | 4b | c3 | d1 | 1b |
| 7 | fb | 66 | ca | 4d | ed | 36 | f2 | fa | 3c | 80 | 79 | 06 | 83 | 97 | 73 | 74 |
| 8 | 00 | ad | 70 | 9a | 75 | 5a | e7 | 44 | 71 | c0 | 3f | 6e | 0c | f1 | 14 | 49 |
| 9 | 77 | ee | d8 | 87 | 0d | ab | 5e | 7d | 9c | 5b | bd | 0e | 45 | 2b | 8a | 0b |
| A | 85 | b9 | 94 | 9d | 01 | f7 | 3a | ae | 9e | ef | 82 | cd | 39 | 40 | eb | 18 |
| B | 15 | 13 | 3d | da | 33 | 32 | 68 | 22 | 16 | fd | 48 | 2d | a4 | ec | 6c | b8 |
| C | ce | d7 | 35 | 1c | 8f | cb | 17 | a7 | 02 | 19 | 4f | 59 | 09 | 31 | 86 | 57 |
| D | e1 | 6a | 96 | 29 | 3e | 34 | 46 | c5 | ff | fc | 12 | 67 | 90 | 07 | d9 | dd |
| E | 47 | 9f | b5 | c8 | f9 | d2 | f5 | 9e | 91 | 5d | 93 | e9 | ac | 2a | d0 | 41 |
| F | b1 | 52 | 21 | f8 | 88 | bf | ba | 7a | e8 | f6 | 60 | e4 | aa | 04 | 37 | 55 |
The algebraic degree of a Boolean function is understood as the maximum degree of a monomial with a coefficient different from 0. It is denoted as \( \text{deg}(f) \) (Oliynykov, 2011).

The correlation value between a Boolean function \( f(x) \) and the set of all linear functions is defined as the Walsh transform (Carlet, 2010):

\[
W(\omega) = \sum_{x=0}^{2^n-1} (-1)^{f(x)\cdot L_\omega(x)},
\]

where \( L_\omega(x) = \omega \cdot x = \omega_0 x_0 + \omega_1 x_1 + \ldots + \omega_{n-1} x_{n-1} \).

A function \( f(x) \) of \( n \) variables is called balanced if

\[
h_w(f) = 2^n - 1.
\]

The nonlinearity of a Boolean function \( \langle \text{NL}(f) \rangle \) is understood as the minimum Hamming distance to all affine functions consisting of \( n \) variables. The following is the relationship between the nonlinearity of a Boolean function and the Walsh transform:

\[
\text{NL}(f) = \frac{1}{2} \cdot (2^n - \max(|W(\omega)|)).
\]

The autocorrelation (AC) function \( r_f(\alpha) \) of the truth table of the Boolean function \( f(x) \) is the derivative of the function for all variables in the direction of \( \alpha \in GF(2^n) \), which is given in the form:

\[
r_f(\alpha) = \sum_{x=0}^{2^n-1} (-1)^{f(x)\oplus f(x;\alpha)}.
\]

Let \( |\text{AC}|_{\max} \) be the maximum absolute value of the autocorrelation function, then:

\[
|\text{AC}|_{\max} = \max_\alpha |r_f(\alpha)|.
\]

Let \( \sigma \) denote the Global Avalanche Characteristics (GAC) “sum-of-squares indicator,” then:

\[
\sigma = \sum_{\alpha=0}^{2^n-1} r_f^2(\alpha).
\]
It is said that some Boolean function $f(x)$ satisfies the Strict Avalanche Criterion (SAC) if the following system of equations is valid for all $s$ (X.-M. Zhang, 1995):

$$\begin{align*}
&\left\{ \begin{array}{c}
hw(s) = 1 \\
\sum_{x=0}^{2^n-1} (f(x) \oplus f(x \oplus s)) = 2^{n-1}
\end{array} \right. \\
&\text{if } s \text{ satisfies } k \leq \text{PC} \leq x \text{ satisfies } GF_2 = s
\end{align*}$$

(12)

A Boolean function $f(x)$ satisfies the propagation criterion of order $k(\text{PC}(k))$ if and only if for a nonzero vector $a \in GF(2^n)$.

$$\begin{align*}
&\left\{ \begin{array}{c}
1 \leq hw(a) \leq k \\
\sum_{x=0}^{2^n-1} (f(x) \oplus f(x \oplus a)) = 2^{n-1}
\end{array} \right.
\end{align*}$$

(13)

A Boolean function $f(x)$ has correlation immunity of order $m(\text{CI}(m))$, if the system of equations is valid for all $w$:

$$\begin{align*}
&\left\{ \begin{array}{c}
1 \leq hw(w) \leq k \\
W(w) = 0
\end{array} \right.
\end{align*}$$

(14)

If a Boolean function $f(x)$ is balanced and at the same time possesses correlation immunity of order $t$, then such a function is called $t$-stable (Kazimirov, 2013).

Let the function $g(x)$ be the annihilator of the function $f(x)$, that is $f(x) \cdot g(x) = 0$. Then, the minimal algebraic degree of the function $g(x) \neq 0$ is called the algebraic immunity of the function $f(x)$ and is denoted as $\text{AI}(f)$ if

$$\begin{align*}
&\left[ \begin{array}{c}
f(x) \cdot g(x) = 0 \\
(f(x) \oplus 1) \cdot g(x) = 0.
\end{array} \right.
\end{align*}$$

(15)

Currently, work is underway to establish a connection between these approaches, especially since recently, work has intensified to study the properties of reduced versions of block symmetric ciphers.

We consider S-boxes of common symmetric block algorithms, including the S-box obtained by our proposed method.

An S-box fulfills SAC criterion if an alteration in one bit in the input bit can cause an avalanche change in the output bits that is nearly half of the output bits must be altered. The comparison of overall SAC analysis of proposed S-box with AES and S-p-box is shown in Tables 2–4, while the average outcomes are

| Bit | $f_0$ | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
| 0   | 132   | 132   | 116   | 114   | 116   | 124   | 116   | 128   |
| 1   | 120   | 124   | 144   | 128   | 124   | 116   | 128   | 136   |
| 2   | 132   | 132   | 128   | 120   | 144   | 128   | 136   | 128   |
| 3   | 136   | 136   | 120   | 116   | 128   | 128   | 128   | 140   |
| 4   | 116   | 128   | 116   | 132   | 128   | 128   | 140   | 136   |
| 5   | 116   | 132   | 132   | 120   | 120   | 140   | 136   | 136   |
| 6   | 136   | 136   | 120   | 132   | 120   | 136   | 136   | 124   |
| 7   | 128   | 140   | 136   | 132   | 144   | 120   | 132   | 120   |
shown in Table 5 and graphical representation of analysis comparison is describe Figure 2. It can be observed from Table 5 that the proposed S-box has attained a maximum value = 0.526, minimum value = 0.437, average value = 0.499, and square deviation = 0.0039. The average SAC values for the proposed S-box are good. These results are consistent with the avalanche effect. Table 6 shows the comparative characteristics of S-boxes of symmetric block ciphers SM4, AES (A), Kuznyechik (K), BelT (B), Camellia (C), and the proposed S-box (M). The values in Table 6 were obtained by known methods for checking the properties of S-boxes (links are given) and using a computer program developed in the laboratory and designed to study the properties of Boolean functions (Duysenbayev & Algazy, 2021).

It is known that the cryptographic strength of block cipher algorithms also depends on the cryptographic strength of the S-boxes used in the algorithm (Kapalova et al., 2020). Therefore, a linear and differential analysis was carried out for the S-boxes used in the proposed algorithm, and the results were compared with the S-boxes of other algorithms (Table 7). The results of linear and differential analysis shown in Table 7 were obtained using a computer program developed in our laboratory. In particular, Table 7 shows the minimum and maximum values of the results of linear and differential analysis. The nonlinearity of the proposed S-box is 102, as shown in Table 7. This value is higher than that of Kuznyechik and coincides with the BelT value.
Table 6. Comparison of the cryptographic characteristics of the generated substitution with S-boxes of known algorithms

| Properties                  | S-boxes |   |   |   |   |
|-----------------------------|---------|---|---|---|---|
|                             | SM4     | A | K | B | C |
| Hamming weight              | 128     | 128 | 128 | 128 | 128 | 128 |
| Balance                     | True    | True | True | True | True | True |
| Hamming distance            | 128     | 128 | 128 | 128 | 128 | 128 |
| | Nonlinearity \( |_{\text{min}}\) | 112 | 112 | 104 | 106 | 112 | 102 |
| | Nonlinearity \( |_{\text{max}}\) | 144 | 144 | 154 | 152 | 144 | 152 |
| Correlation value (min)     | -32     | -32 | -52 | -48 | -32 | -48 |
| Correlation value (max)     | 32      | 32 | 48 | 44 | 32 | 52 |
| | Value \( |_{\text{max}}\) | 32 | 32 | 64 | 72 | 32 | 72 |
| | Value \( |_{\text{max}}\) | 133120 | 133120 | 86656 | 161152 | 133120 | 174592 |
| | Value \( |_{\text{max}}\) | 133120 | 133120 | 228352 | 201088 | 133120 | 201472 |
| SAC                         | False   | False | False | False | False | False |
| Propagation criterion       | N/A     | N/A | N/A | N/A | N/A | N/A |
| CI                          | N/A     | N/A | N/A | N/A | N/A | N/A |
| t-stability                 | N/A     | N/A | N/A | N/A | N/A | N/A |
| AI                          | 3       | 2 | 3 | 3 | 3 | 3 |
| S-box cyclic structure      | 0;113;120, 1;108;24, 4;22;56, 13;172;35, 36;979;45;216;3, 54;121;6, 171;171;1, 180;222;2, 0;82;59, 1;98;1, 4;48;87, 11;158;27, 115;43;2, 0;165;24;3;21;96;13, 3;156;78, 22;147;7, 0;10;136;1;167;35, 3;156;78, 22;147;7, 0;146;249;80;127;5, 201;245;2, 0;128;12;1;164;104, 2;200;82;4;253;29, 7;221;3, 24;175;6, 40;86;18, 125;151;2, |

Khompysh et al. Cogent Engineering (2022), 9:2080623
https://doi.org/10.1080/23311916.2022.2080623
5. Conclusion

This article describes a developed and software-implemented method for designing block cipher substitution nodes that can be used to create encryption algorithm substitution boxes for pre-encrypting confidential information. This method can be used in the design and implementation of cryptographic information protection facilities to protect unclassified information of limited access transmitted over open communication channels. The use of the mathematical apparatus of logical vector functions, and also indicators and criteria of the efficiency of the obtained S-box make it possible to simplify the description of the main elements of symmetric algorithms. The studies carried out have shown that the application of the theoretical approach does not always meet the practical needs, in particular, for the generation of substitutions with given unsaturated values. Thus, despite the many existing solutions in the field of symmetric cryptography, the search for approaches that provide protection against existing and proposed types of attacks on encryption
algorithms is relevant. It is necessary to substantiate the criteria for the development and improvement of methods for creating nodes of nonlinear transformations, as well as to conduct research on cryptanalysis methods and the theory of vector Boolean functions.

To date, there is no unambiguous set of criteria for an ideal S-box. Many studies show that perfect substitutions probably don’t exist. Therefore, the term “optimal substitution” is introduced, the criteria of which are determined for a specific encryption algorithm (or a group of algorithms) and are optimal from the point of view of protection against existing types of attacks. Based on the research, a method was proposed for obtaining substitutions used in cryptographic sentences.

As a result of the study, it was found that the characteristics of the efficiency of the obtained S-box are not inferior to those of the known algorithms. It was also shown that the obtained numerical values correspond to the values of the S-boxes of such algorithms as Kuznyechik, Camellia and others.

Acknowledgements
The research work was carried out within the framework of the project AP08856426 “Development and research of an encryption algorithm and the creation of a hardware-software complex for its implementation” at the Institute of Information and Computational Technologies of the Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan.

Funding
The authors received no direct funding for this research.

Author details
Ardabek Khompysh1,2
E-mail: ardabek@mail.ru
Nursulu Kapalova
E-mail: nkopalova@mail.ru
ORCID ID: http://orcid.org/0000-0001-9743-9981
Kunbolat Algazy1
E-mail: kunbolat@mail.ru
E-mail: okoyewole@wpi.edu
Dilmukhanbet Dyusenbayev1
E-mail: dimash_dds@mail.ru
Kairat Sakan1,2
E-mail: kairat_sks@mail.ru
1 Institute of Information and Computational Technologies of the RK MES CS, Kazakhstan.
2 Al-Farabi Kazakh National University, Almaty, Kazakhstan.

Disclosure statement
Theme of scientific research development and research of reliable algorithms for generating pseudorandom sequences used to generate key sequences for symmetric block and stream encryption, as well as the development of cryptographic information protection algorithms using modular arithmetic.

Citation information
Cite this article as: Design of substitution nodes (S-Boxes) of a block cipher intended for preliminary encryption of confidential information, Ardash Khompysh, Nursulu Kapalova, Kunbolat Algazy, Dilmukhanbet Dyusenbayev & Kairat Sakan, Cogent Engineering (2022), 9: 2080623.

References
Bobenkov, L. K., & Ischukova, E. A. (2006). Modern block encryption algorithms and methods of their analysis (pp. 376). Moscow, Helias, ARV.
Biyashev, R. G., Kapalova, N. A., Dyusenbayev, D. S., Algozy, K. T., & Wojcik, W. (2021). Andrzej Smolarz development and analysis of symmetric encryption algorithm qamal based on a substitution-permutation network. International Journal of Electronics and Telecommunications 1, 127–132. http://dx.doi.org/10.24425/iet-2021.135954
Carlet, C. (2010). Vectorial Boolean functions for cryptography. Boolean models and methods in mathematics, computer science, and engineering (pp. 398–469). (Y. Crama & P. Hammer, Ed.). Cambridge University Press.
Carlet, C. (2010). Boolean functions for cryptography and error-correcting codes. Boolean Models and Methods in Mathematics. Computer Science, and Engineering, 2, 257 doi: 10.1017/CBO9780511780448.012.
Chew, L. C. N., & Ismail, E. S. (2020). S-box construction based on linear fractional transformation and permutation function. Symmetry, 12(826), 1–16. https://doi.org/10.3390/sym12050826
Connor, L. D. (1994). An analysis of a class of algorithms for S-box construction. Journal of Cryptology, 7(1), 133–151. https://doi.org/10.1007/BF02318546
Dolgov, V. I., Kuznetsov, A. A., Lisitskaya, I. V., Sergienko, R. V., & Oleshko, O. I. (2009). Study of the cryptographic properties of nonlinear substitution nodes for reduced versions of some ciphers. Applied Radio Electronics, 8(3), 268–276 https://openarchive.nure.ua/bitstream/document/216/1/04.pdf.
Dyusenbayev, D. S., & Algozy, K. T. (2021, June 11). Sakon Study of nonlinear nodes used in symmetric ciphers. International scientific and practical conference Actual problems of information security in Kazakhstan, Kazakhstan, 34–38.
Grigoryan, D. R., & Nji, N. V. (2019). The impact of the S-block inertia group on the differential characteristic of a block cipher. TUSUR Reports, 22(1), 45–49. https://journal.tusur.ru/storage/126704%0D%0A%D1%83%D1%80%0D%0B%0D%BB-12-2019.pdf51559275355
Hussain, I., Shah, T., Gandal, M. A., & Mahmood, H. (2012). Construction of S8 Liu J S-boxes and their applications. Computers & Mathematics with Applications, 64 (8), 2450–2458. https://doi.org/10.1016/j.camwa.2012.05.017
Kapalova, N., & Haumen, A. (2018). The model of encryption algorithm based on non-positional polynomial notations and constructed on an SP-network. Open Engineering, 8(1), 140–146. https://doi.org/10.1515/eng-2018-0013
Kapalova, N. A., Khompysh, A., Müslüm, A., Algozy, K., & Pham, D. (2020). A block encryption algorithm based
on exponentiation transform. Cogent Engineering, 7 (1), 1788292. 2331-1916, 1-12. https://doi.org/10.1080/23311916.2020.1788292

Kazimirov, A. V., & Oleinikov, R. V. (2019). Selection of S-boxes for symmetric cryptographic algorithms based on the analysis of algebraic properties. Bulletin of Kharkiv National University, 95, 79–85.

Kazimirov, A. V. (2013). Methods and tools for generating nonlinear substitution nodes for symmetric cryptoolgorithms (Ph.D. thesis). Engineering Science. 190.

Kuznetsov, A. A., Belazertsev, I. N., & Andrushkevich, A. V. (2015). Analysis and comparative studies of nonlinear substitution nodes of modern block symmetric ciphers. Applied Radio Electronics, 14(4), 343–350 https://openarchive.nure.ua/bitstream/document/6470/1/A15.pdf.

Maier, W., & Staffelbach, O. (1990). Nonlinearity criteria for cryptographic functions. Advances in Cryptology, 434, 549–562. EUROCRYPT'89, Lecture Notes in Computer Science, Springer-Verlag https://doi.org/10.1007/3-540-46885-4_53

Maitra, S., & Pasalic, E. (2001, May). Further constructions of resilient Boolean functions with very high nonlinearity. Accepted in SETA, 48(7) http://ieeexplore.ieee.org.ludwig.lub.lu.se/iel5/18/21818/10103128.pdf?tp=&arnumber=1013128&isnumber=21818. Norway.

Menezes, A. J., van Oorschot, P. C., & Vanstone, S. (1997). Handbook of applied cryptography (pp. 810). CRC Press.

Millan, Y., Clark, A., & Dawson, E. (1996). Heuristic design of cryptographically strong balanced boolean functions. Advances in Cryptology EUROCRYPT'98, Springer Verlag LNCS 1403. 489–499.

Naseer*, Y., Shah, T., Shah, D., & Hussain, S. (2019). A novel algorithm of constructing highly nonlinear S-boxes. Cryptography, 3(1), 2–13. https://doi.org/10.3390/cryptography3010006

Oliynykov, R. (2013). An impact of S-Box Boolean function properties to strength of modern symmetric block ciphers. Radio Engineering, All-Ukrainian Interdepartmental Scientific and Technical Bulletin, 166, 11–16. Information Security.

Ozkaynak, F. (2017). Construction of robust substitution boxes based on chaotic systems. Neural Computing and Applications, 31(8), 3317–3326. https://doi.org/10.1007/s00521-017-3287-y

Panasesko, S. P. (2005). Encryption algorithms. Special reference book (pp. 576). BHV-Petersburg.

Pasalic, E., Johansson, T., Maitra, S., & Sarkar, P. (2001, January). New constructions of resilient and correlation immune Boolean functions achieving upper bounds of nonlinearity. Workshop of Coding and Cryptography, Electronic Notes in Discrete Mathematics. Elsevier.

Preneel, B. (1993, January). Analysis and design of cryptographic hash functions [Ph.D. dissertation]. Katholieke Universiteit Leuven.

Sergienko, R. V., & Moskovchenko, I. V. (2007). Study of the cryptographic properties of nonlinear substitution nodes of the symmetric encryption algorithm GOST 28147-89. Information Processing Systems, 8 (66), 91–95.

Shah, T., Hussain, I., Gondal, M. A., & Mahmood, H. (2011). Statistical analysis of S-box in image encryption applications based on majority logic criterion. International Journal of Physical Sciences, 6(18), 4110–4127 https://doi.org/10.5897/IJPS11.531.

Shah, T., & Shah, D. (2018). Construction of highly nonlinear S-boxes for degree 8 primitive irreducible polynomials over Z2. Multimedia Tools and Applications, 78, 1–16 https://doi.org/10.1007/s11042-018-6250-8. [CrossRef].

Shannon, C. E. (1949). Communication theory of secrecy systems. Bell System Technical Journal, 28(4(6)), 656-715. https://doi.org/10.1002/j.1538-7305.1949.tb00928.x

Zahid, A. H., & Arshad, M. J. (2019). An innovative design of substitution-boxes using cubic polynomial mapping. Symmetry, 11(3), 437. https://doi.org/10.3390/sym11030437

Zahid, A. H., Arshad, M. J., & Ahmad, M. (2019). A novel construction of efficient substitution-boxes using cubic fractional transformation. Entropy, 21(3), 245. https://doi.org/10.3390/e21030245

Zahid, A. H., Al-Solami, E., & Ahmad, M. (2020). A Novel Modular Approach based substitution-box design for image encryption. IEEE Access, 8, 150326–150340. https://doi.org/10.1109/ACCESS.2020.3016401

Zenzin, O. S., & Ivanov, M. A. (2002). Cryptographic protection standard – AES. Finite fields (pp. 176).

Zhang, X.-M. (1995). GAC – The criterion for global avalanche characteristics of cryptographic functions. Journal of Universal Computer Science, 1(5), 316–333.

Zhang, Y. (2018). The unified image encryption algorithm based on chaos and cubic S-Box. Information Sciences, 450, 361–377. [CrossRef]. https://doi.org/10.1016/j.ins.2018.03.055
