Algebraic cryptanalysis of round-reduced lightweight ciphers SIMON and SPECK

Aleksandr Kutsenko[1,2], Natalia Atutova[1,2], Darya Zyubina[1,2], Ekaterina Maro[3] and Stepan Filippov[4]

[1]Sobolev Institute of Mathematics, Novosibirsk, Russia
[2]Novosibirsk State University, Novosibirsk, Russia
[3]Southern Federal University, Taganrog, Russia
[4]Saint Petersburg State University, Saint Petersburg, Russia

CTCrypt 2021

Algebraic cryptanalysis

The main idea is to compose a complex system of Boolean equations describing the transformation of the cipher. The system is built on the basis of a fully known encryption algorithm.

$$\begin{cases} p_1 k_2 \oplus p_2 c_2 k_3 = k_4, \\ (p_1 \oplus k_1) \oplus (p_2 \oplus k_2) = c_5, \\ c_8 k_1 k_2 k_4 \oplus p_3 k_3 = 1, \\ \ldots. \end{cases}$$

Algebraic cryptanalysis

Encrypting a certain amount of plaintext in a key unknown to the cryptanalyst allows the system to be valued - substituting a system of bits from vectors into the equations P and C.

At the next stage, this system of Boolean equations is solved using various methods. The bits of the key are unknown - they correspond to the solution.

LRX- and ARX-structures

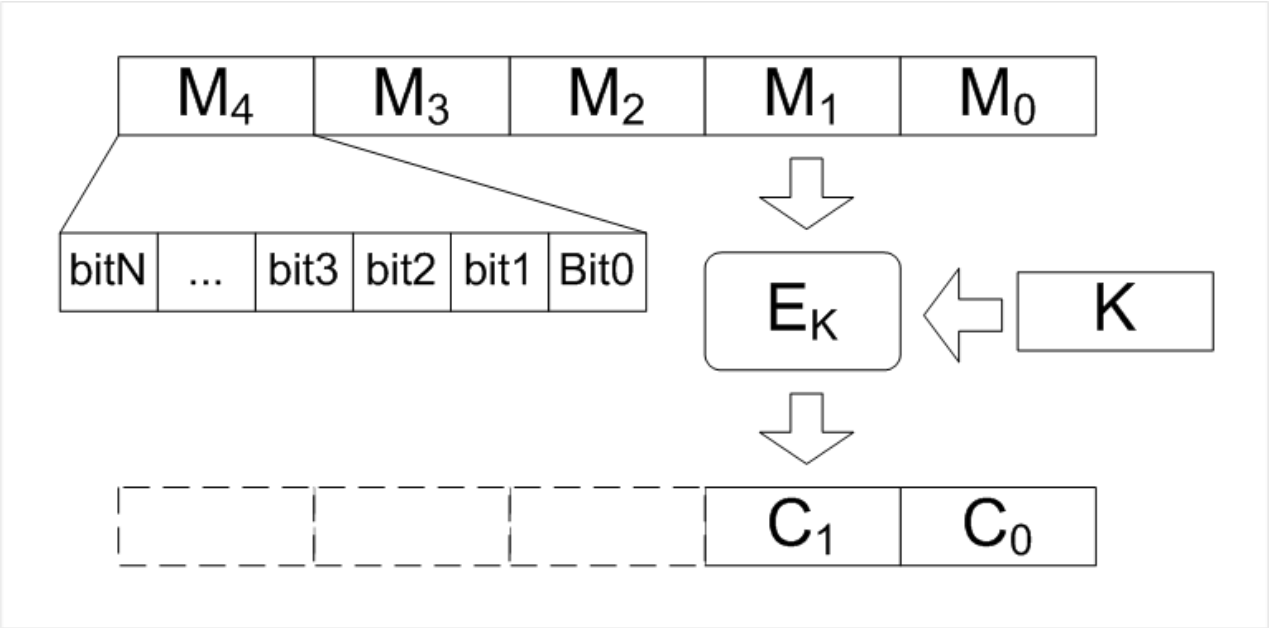
- LRX
  - logical operations
  - cyclic shift left (right)
  - addition mod 2
- ARX
  - addition mod 2
  - cyclic shift left (right)
  - addition mod 2^n

Objectives

- Investigate the resistance of LRX and ARX structures to algebraic cryptanalysis
- Consider the main methods of algebraic attacks for a specific pair of lightweight ciphers with a reduced number of encryption rounds
- Compare the effectiveness of various methods for the ciphers under consideration

Simon and Speck

Simon и Speck — family of block ciphers submitted by the US NSA in 2013.



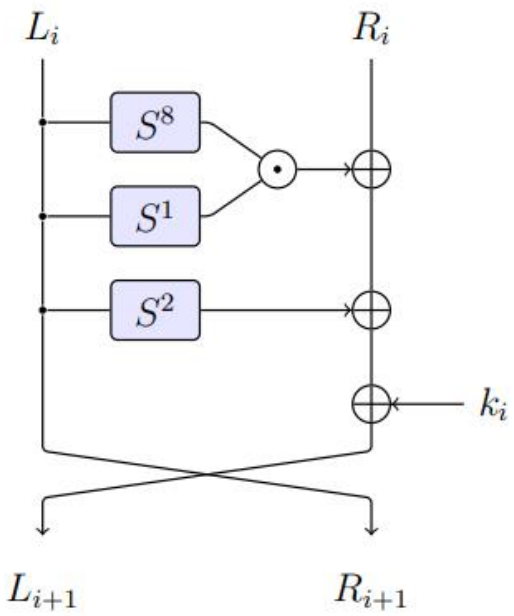
Simon and Speck

- SIMON
  - LRX-structure
  - International standard ISO/IEC 29167-21:2018
  - For hardware implementations
- SPECK
  - ARX-structure
  - International standard ISO/IEC 29167-22:2018
  - For software implementations

Cipher Simon

Simon2n/mn

$2n$ — block size, m — key size.
The block is divided into two parts of size n bits (word length).
Total T rounds of encryption. On the i -th round:



block size $2n$	key size mn	word size n	key words m	const seq	rounds T
32	**64**	16	4	z_0	32
48	**72**	24	3	z_0	36
	96		4	z_1	36
64	**96**	32	3	z_2	42
	128		4	z_3	44
96	**96**	48	2	z_2	52
	144		3	z_3	54
128	**128**	64	2	z_2	68
	192		3	z_3	69
	256		4	z_4	72

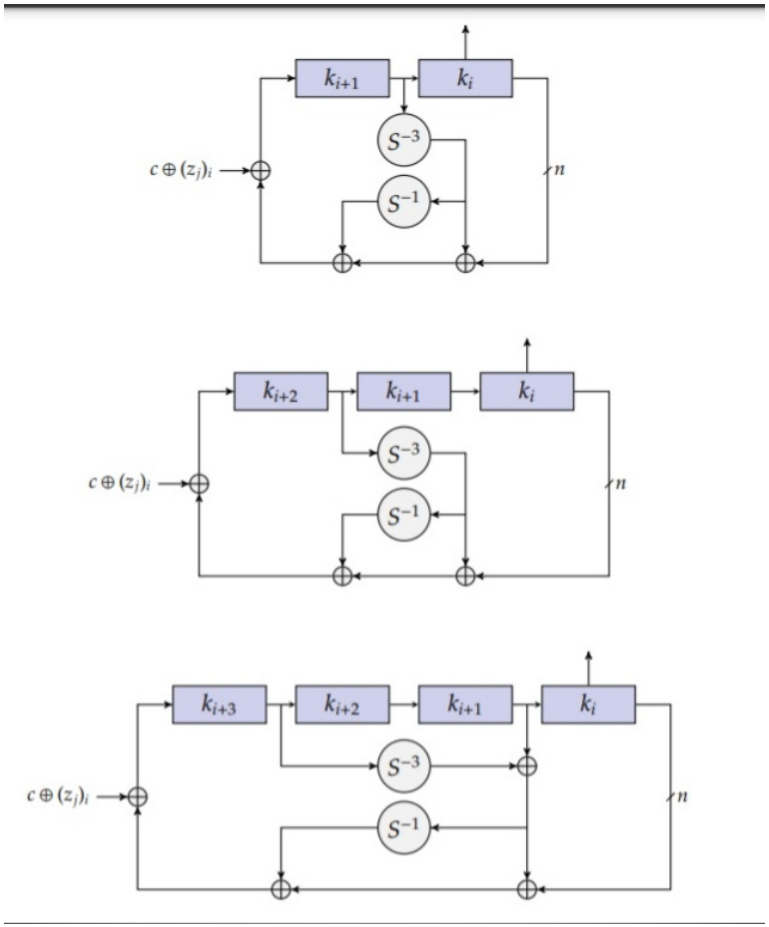
Key Simon Cipher Schedule

The first m keys are set, each of which consists of n bits. The sequence of keys is calculated recursively ($c = 2^n - 4$ is a constant, and z_j is a fixed periodic sequence. The value of m depends on the values of the block size $2n$ and the number of rounds T

$$k_{i+m} = \begin{cases} c \oplus (z_j)_i \oplus k_i \oplus \left(I \oplus S^{-1}\right) S^{-3} k_{i+1}, & \text{for } m = 2, \\ c \oplus (z_j)_i \oplus k_i \oplus \left(I \oplus S^{-1}\right) S^{-3} k_{i+2}, & \text{for } m = 3, \\ c \oplus (z_j)_i \oplus k_i \oplus \left(I \oplus S^{-1}\right) (S^{-3} k_{i+3} \oplus k_{i+1}), & \text{for } m = 4. \end{cases}$$

Cipher Simon

Simon2n/mn

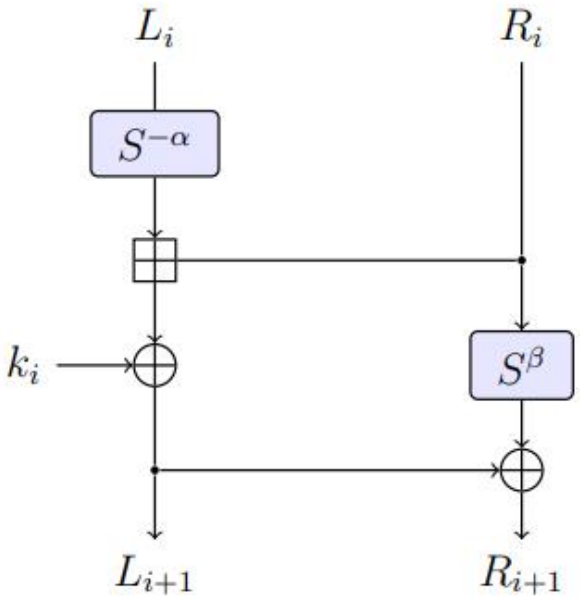


block size $2n$	key size mn	word size n	key words m	const seq	rounds T
32	**64**	16	4	z_0	32
48	**72**	24	3	z_0	36
	96		4	z_1	36
64	**96**	32	3	z_2	42
	128		4	z_3	44
96	**96**	48	2	z_2	52
	144		3	z_3	54
128	**128**	64	2	z_2	68
	192		3	z_3	69
	256		4	z_4	72

Cipher Speck

Speck2n/mn

$2n$ — block size, m — key size.
The block is divided into two parts of size n bits (word length).
Total T rounds of encryption. On the i-th round:



block size $2n$	key size mn	word size n	key words m	rot α	rot β	rounds T
32	**64**	16	4	7	2	22
48	**72**	24	3	8	3	22
	96		4			23
64	**96**	32	3	8	3	26
	128		4			27
96	**96**	48	2	8	3	28
	144		3			29
128	**128**	64	2	8	3	32
	192		3			33
	256		4			34

Speck: System of boolean equations

The round function of the Speck cipher is non-linear, i.e. it cannot be described solely by linear algebraic equations. This property in this cipher is provided by the addition operation modulo 2^n included in the encryption algorithm.

Speck: Addition mod 2^n

$$
\begin{cases}
w_0 x_{i+\alpha} = x_\alpha x_{i+\alpha} \oplus y_0 x_{i+\alpha} \ , \ i = \overline{1, n-1} \\
w_0 y_i = x_\alpha y_i \oplus y_0 y_i \ , \ i = \overline{0, n-1} \\
w_0 w_i = x_\alpha w_i \oplus y_0 w_i \ , \ i = \overline{0, n-1} \\
w_1 x_\alpha = x_{1+\alpha} x_\alpha \oplus y_1 x_\alpha \oplus x_\alpha y_0 \\
w_1 y_0 = x_{1+\alpha} y_0 \oplus y_1 y_0 \oplus x_\alpha y_0 \\
w_i = x_{i+\alpha} \oplus y_i \oplus x_{i-1+\alpha} \oplus y_{i-1} \oplus x_{i-1+\alpha} y_{i-1} \oplus x_{i-1+\alpha} w_{i-1} \oplus y_{i-1} w_{i-1}, \\
\quad i = \overline{2, n-1} \\
w_i (x_{i-1+\alpha} \oplus y_{i-1}) = x_{i-1+\alpha} x_{i+\alpha} \oplus x_{i-1+\alpha} y_i \oplus x_{i-1+\alpha} \oplus x_{i-1+\alpha} w_{i-1} \\
\quad \oplus x_{i+\alpha} y_{i-1} \oplus y_{i-1} y_i \oplus y_{i-1} \oplus y_{i-1} w_{i-1} \ , \ i = \overline{2, n-1} \\
w_i (x_{i-1+\alpha} + w_{i-1}) = x_{i-1+\alpha} x_{i+\alpha} \oplus x_{i-1+\alpha} y_i \oplus x_{i-1+\alpha} \oplus x_{i+\alpha} w_{i-1} \oplus y_i w_{i-1} \\
\quad \oplus x_{i-1+\alpha} w_{i-1} \ , \ i = \overline{2, n-1}
\end{cases}
$$

Algebraic cryptanalysis

- Abed F., List E., Lucks S., Wenzel J., Derential Cryptanalysis of Round-Reduced Simon and Speck, (2015);
- Courtois N, Mourouzis T, Song G, Sepehrdad P, Susil P., Combined Algebraic and Truncated Dierential Cryptanalysis on Reduced-round Simon, (2014);
- Raddum H., Algebraic Analysis of the Simon Block Cipher Family, (2015);
- Andrzejczak M., Dudzic W., SAT Attacks on ARX Ciphers with Automated Equations Generation, (2019); S.L., Le D.-P., Khoo K., Improved algebraic attacks on lightweight block ciphers, (2021).

Linearization-based methods

The idea behind this method is to assign a new variable to each monomial in the original system. The system becomes linear after assignment.
The efficiency of linearization depends on the rank r of the system, while the number of different monomials in the original system determines the number of variables n' in the system of linear equations. The solution set is not empty, so it is equal to $2^{n'-r} > 0$, so to evaluate performance, it is necessary to analyze the bounds for the values of n' and r.

Estimating the number of monomials Simon

Provided that new variables are introduced at each round for the bitwise output of the nonlinear operation, the upper bound for the number of different monomials has the form:

$$M \leqslant 6nT$$

Without introducing new variables, the following formula is valid:

$$P(T) = \begin{cases} 4 \cdot n, & T = 1 \\ 7 \cdot n, & T = 2 \\ n\big(P^2(T-2) + P(T-2) + 1\big), & \text{if } T \vdots 4 \\ n\big(P^2(T-2) + P(T-1) + P(T-2) + 1\big), & \text{otherwise.} \end{cases}$$

Estimating the number of monomials Simon

Block size $2n$	Word size n	Rounds T	Num. of monomials	Num. of equations	Num. of unknowns with key schedule
32	16	32	$\approx 2^{11.585}$	$\approx 2^{8.9069}$	$\approx 2^{8.9542}$
48	24	36	$\approx 2^{12.34}$	$\approx 2^{9.6724}$	$\approx 2^{9.7142}$
64	32	42	$\approx 2^{12.977}$	$\approx 2^{10.322}$	$\approx 2^{10.358}$
		44	$\approx 2^{13.044}$	$\approx 2^{10.392}$	$\approx 2^{10.426}$
96	48	52	$\approx 2^{13.87}$	$\approx 2^{11.229}$	$\approx 2^{11.257}$
		54	$\approx 2^{13.925}$	$\approx 2^{11.285}$	$\approx 2^{11.313}$
128	64	68	$\approx 2^{14.672}$	$\approx 2^{12.044}$	$\approx 2^{12.066}$
		69	$\approx 2^{14.693}$	$\approx 2^{12.066}$	$\approx 2^{12.087}$
		72	$\approx 2^{14.755}$	$\approx 2^{12.129}$	$\approx 2^{12.15}$

Estimating the number of Speck monomials

In the system of equations describing the addition modulo 2^n there are only $5(7n-8)$ monomials. The formula for estimating the number of monomials, excluding those that come from the key equations of the schedule (all equations are linear), is

$$M \leqslant (28n-18)T$$

Estimating the number of monomials Speck

Block size $2n$	Rounds T	Num. of monomials	Num. of eqs. without key sch.	Num. of eqs. with key schedule	Num. of unknowns without key sch.	Num. of unknowns with key schedule
32	22	$\approx 2^{13.2}$	$\approx 2^{11.4}$	$\approx 2^{12.4}$	$\approx 2^{9.95}$	$\approx 2^{11}$
48	22	$\approx 2^{13.81}$	$\approx 2^{12}$	$\approx 2^{13.03}$	$\approx 2^{10.5}$	$\approx 2^{11.59}$
	23	$\approx 2^{13.88}$	$\approx 2^{12.1}$	$\approx 2^{13.09}$	$\approx 2^{10.6}$	$\approx 2^{11.65}$
64	26	$\approx 2^{14.47}$	$\approx 2^{12.7}$	$\approx 2^{13.7}$	$\approx 2^{11.2}$	$\approx 2^{12.25}$
	27	$\approx 2^{14.5}$	$\approx 2^{12.75}$	$\approx 2^{13.74}$	$\approx 2^{11.27}$	$\approx 2^{12.3}$
96	28	$\approx 2^{15.2}$	$\approx 2^{13.4}$	$\approx 2^{14.4}$	$\approx 2^{11.9}$	$\approx 2^{12.94}$
	29	$\approx 2^{15.23}$	$\approx 2^{13.44}$	$\approx 2^{14.44}$	$\approx 2^{11.96}$	$\approx 2^{13}$
128	32	$\approx 2^{15.79}$	$\approx 2^{14}$	$\approx 2^{15}$	$\approx 2^{12.52}$	$\approx 2^{13.56}$
	33	$\approx 2^{15.84}$	$\approx 2^{14.04}$	$\approx 2^{15.04}$	$\approx 2^{12.57}$	$\approx 2^{13.6}$
	34	$\approx 2^{15.88}$	$\approx 2^{14.08}$	$\approx 2^{15.08}$	$\approx 2^{12.62}$	$\approx 2^{13.64}$

XL-attack

The algorithm was proposed (Courtois, Shamir, Patarin, Klimov, 2000) The input is a system of m polynomial equations in n unknowns of degree d and outputs its solution or solutions if the equations have sufficient rank.

- 1: Grade $D > d$ is selected. Usually $D = d + 1$
- 2: Multiply the equations of the original system by monomials of degree at most $D - d$
- 3: Linearize the System
- 4: Solve the resulting system using linear algebra methods

Linearization-based methods

	Simon parameters	Number of equations	Number of variables	Number of monomials	Number of solutions
Pure linearization	$T = 3,\ m = 1$	48	32	48	4, only one corresponds to the key
XL-method	$T = 3,\ m = 1$	1584	32	992	1
Pure linearization	$T = 4,\ m = 1$	64	48	80	65536
XL-method	$T = 4,\ m = 1$	3136	48	2616	256, only one corresponds to the key
Pure linearization	$T = 5,\ m = 1$	80	64	112	2^{32}
XL-method	$T = 5,\ m = 1$	5200	64	5008	2^{336}
	Speck parameters				
Pure linearization	$T = 3,\ m = 1$	500	176	1236	—
XL-method	$T = 3,\ m = 1$	88500	176	185216	—

ElimLin

The main steps of ElimLin (Courtois, Bard, 2015):

- all linear equations in the linear envelope of the equations of the system are searched;
- the variables of linear equations are expressed in turn through the remaining terms until there are no linear equations left in the system.

	Parameters	**(Equations, Linear equations)**	**(Equations, Linear equations after ElimLin applied)**
Simon	$T = 3,\ m = 1$	(48, 32)	(48, 32)
Simon	$T = 5,\ m = 1$	(80, 32)	(80, 48)
Speck	$T = 3,\ m = 1$	(500, 132)	(307, 137)
Speck	$T = 5,\ m = 2$	(1032, 296)	(654, 297)

SAT-solver attack

The Boolean satisfiability problem (SAT) is a recognition problem whose solution consists in finding an answer to the question whether the Boolean equation in question has a solution.
Original system in the form of
ANF (Zhegalkin polynomial) -> CNF -> SAT solver processing.

SAT-solver attack

Simon parameters	Num. of equations	Num. of unknowns	Num. of unknowns	SAT	Time (RAM)
$T = 3$, $m = 1$ (with round key)	80	80	96 lit., 432 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	0.17 sec. 0.01 sec., 0.1 MB 1.1 sec., 0.7 MB 0.50 sec., 0.05 MB
$T = 5$, $m = 2$ (with round key)	128	128	192 lit., 1136 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	8.43 sec. 0.9 sec., 2.0 MB 2.9 sec., 21.0 MB 2.36 sec., 10 MB
$T = 5$, $m = 2$ (key schedule)	128	128	176 lit., 1710 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	15.79 sec. 1.4 sec., 2.0 MB 2.2 sec., 15.4 MB 0.86 sec., 3 MB
$T = 7$, $m = 2$ (with round key)	192	192	320 lit., 2064 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	287.31 sec. 3687.9 sec., 45.9 MB 212.7 sec., 103.3 MB 681.14 sec., 77 MB
$T = 7$, $m = 2$ (key schedule)	192	192	320 lit., 3632 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	101.23 sec. 1867.2 sec., 38.0 MB 229.5 sec., 99.2 MB 389.84 sec., 62 MB

SAT-solver attack

Simon parameters	Num. of equations	Num. of unknowns	Num. of unknowns	SAT	Time (RAM)
$T = 8$, $m = 2$ (with round key)	224	224	384 lit., 2528 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	- 69811.9 sec., 120.5 MB 4775.5 sec., 260.3 MB 12702.81 sec., 182 MB
$T = 8$, $m = 2$ (key schedule)	128	128	368 lit., 4448 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	- 845.4 sec., 26.6 MB 1188.8 sec., 169.2 MB 4426.12 sec., 95 MB
$T = 9$, $m = 2$ (key schedule)	144	144	480 lit., 6448 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	- - 47799.2 sec., 620.3 MB -
$T = 10$, $m = 2$ (key schedule)	160	160	560 lit., 8096 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	- - 17554.9 sec., 458.8 MB -
$T = 11$, $m = 2$ (key schedule)	176	176	640 lit., 9648 clause	CryptoMiniSat (SageMath) Lingeling Plingeling Treengeling	- - - -

SAT-solver attack

Speck parameters	Num. of equations	Num. of unknowns	Num. of unknowns	SAT	Time (RAM)
$T = 3,\ m = 1$	500	176	1460 lit., 11020 clause	CryptoMiniSat (SageMath) Plingeling Treengeling Lingeling	0.56 sec. 0.9 sec., 9.6 MB 0.97 sec., 4 MB 0.2 sec., 1.9 MB
$T = 4,\ m = 2$	782	320	2492 lit., 17380 clause	CryptoMiniSat (SageMath) Plingeling Treengeling Lingeling	21.4 sec. 3.0 sec., 17.3 MB 8.25 sec., 15 MB 61.4 sec., 14.8 MB
$T = 5,\ m = 2$	1032	416	3312 lit., 23184 clause	CryptoMiniSat (SageMath) Plingeling Treengeling Lingeling	- - 14448.17 sec., 278 MB -
$T = 6,\ m = 2$	1282	512	4132 lit., 28988 clause	CryptoMiniSat (SageMath) Plingeling Treengeling Lingeling	- - 123353.82 sec., 546 MB -

Raddum-Semaev algorithm

The method can be used to solve sparse systems of equations.

- A graph is constructed, each vertex of which corresponds to an equation of the original system. The vertex has two attributes (X_i, L_i)
- The graph is supplemented with vertices with attributes $(X_{ij} = X_i \cap X_j, \mathbb{F}_2^{|X_i \cap X_j|})$
- Vertices X_i,X_j are connected by edges with a vertex X_{ij}
- The procedure "Vertex matching"is started
  - If the size of the lists L_i allows you to find a solution to the system, then « Solution Search »
  - Otherwise, the procedure « Merge vertices »

Raddum-Semaev algorithm

For pairs (X_1, L_1) and (X_2, L_2) , sets of variables $Z = X_1 \cup X_2$ and $Y = X_1 \cap X_2$ are defined by the rule $U = \{a_1, b, a_2\}$ with $(a_1, b) \in L_1$, $(b, a_2) \in L_2$, $a_i = X_i \setminus Y$ and b belongs to Y. Then the vector (a_1, b, a_2) is the gluing of (a_1, b) and $\{b, a_2\}$.

Raddum-Semaev algorithm

$m = 2$	**Num. of rounds (T)**	5	6	7	8	9	10	11	12
	Max. num. of variables	9	11	12	12	14	17	18	18
	Num. of rounds (T)	13	14	15	16	17	18	19	20
	Max. num. of variables	18	18	18	18	22	25	26	26
$m = 4$	**Num. of rounds (T)**	5	6	7	8	9	10	11	12
	Max. num. of variables	6	10	18	21	23	25	28	31

Raddum-Semaev algorithm: Cipher Simon

The Raddum-Semaev method made it possible to find solutions to the system of equations for the number of rounds 7,8,9

Simon parameters	**Num. of equations**	**Num. of unknowns**	**Upper set Lower set**
$T = 7,\ m = 2$	112	112	112 800
$T = 8,\ m = 2$	128	128	128 1072
$T = 9,\ m = 2$	144	144	144 1600

Raddum-Semaev algorithm: Cipher Speck

Number of variables	Number of equations
6	$6n - 12$
5	6
4	$8n - 2$
3	4
2	n

Raddum-Semaev algorithm: Cipher Speck

The Raddum-Semaev method made it possible to find solutions to the system of equations for the number of rounds 3,5,6

Speck parameters	**Num. of equations**	**Num. of unknowns**	**Upper set Lower set**
$T = 3,\ m = 1$	500	176	500 558
$T = 4,\ m = 2$	782	320	782 749
$T = 5,\ m = 2$	1032	416	1032 1005
$T = 6,\ m = 2$	1282	512	1282 1229

Conclusion

- Analyzed the resistance of LRX and ARX structures to algebraic attacks on the example of ISO standards Simon and Speck
- With regard to these ciphers, a conclusion was made about the low efficiency of methods based on linearization
- Shows the efficiency of using the SAT-solver, as well as methods using the sparsity of the systems of equations

Thank you for your attention!