Some cryptanalytic results on Stream ciphers with short internal states

Subhadeep Banik

EPF, Lausanne

Invited Talk to ASK 2019

14th December 2019

Outline

- Introduction
- Sprout (FSE15)
- Previous Work
- Attack by Esgin/Kara (SAC 2015)
- Distinguishing Attack
- State Recovery Attack
- After Sprout
- Attack on Plantlet

Sprout

- Biryukov, Shamir [Asiacrypt 2001] : State size must be 1.5 to 2 times size of Secret Key.
- Radical Departure: Sprout by Armknecht and Mikhalev in FSE 2015.

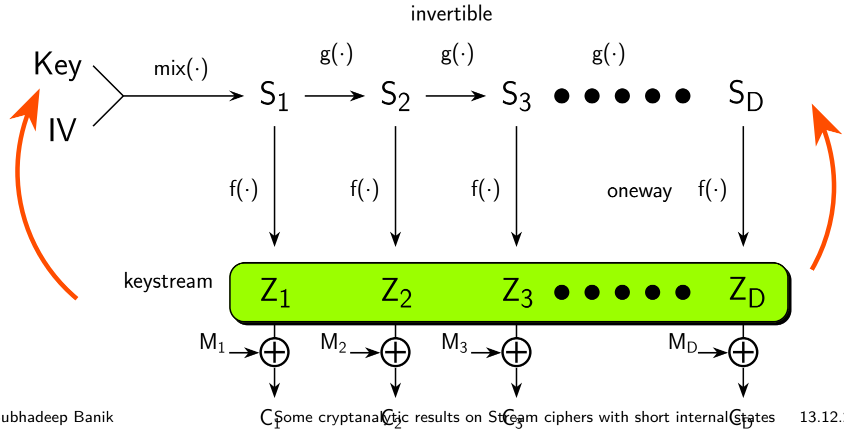
  $\rightarrow$ State Size equal to size of Secret Key.

  $\rightarrow$ Avoids Generic TMD Tradeoff Attacks due to Key mixing in state update.
- Grain like structure: LFSR and NFSR of size 40 bits each.
- Much smaller in area than any known stream cipher.

State twice the size of Secret Key

Biryukov, Shamir [Asiacrypt 2001]

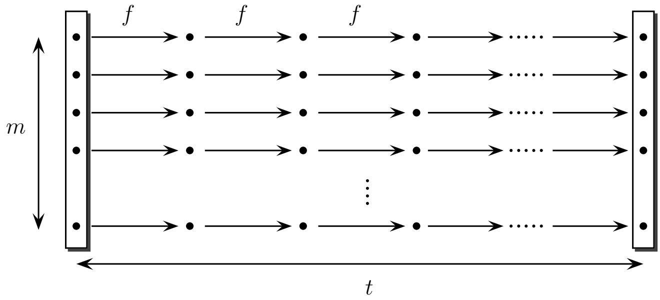
- Let N denote the size of the set of internal states.
- f denotes the function mapping state to keystream.



State twice the size of Secret Key

Biryukov, Shamir [Asiacrypt 2001]

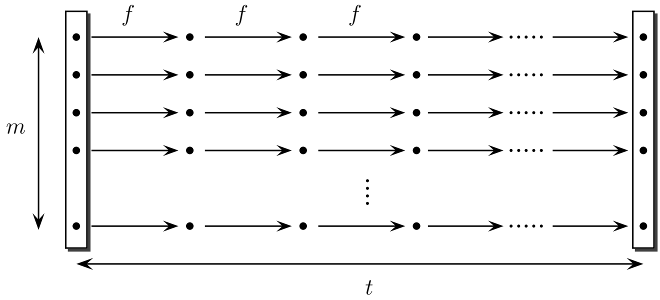
- Randomly choose m initial states and form a function chain.
- f is the function that maps state to keystream segment.



State twice the size of Secret Key

Biryukov, Shamir [Asiacrypt 2001]

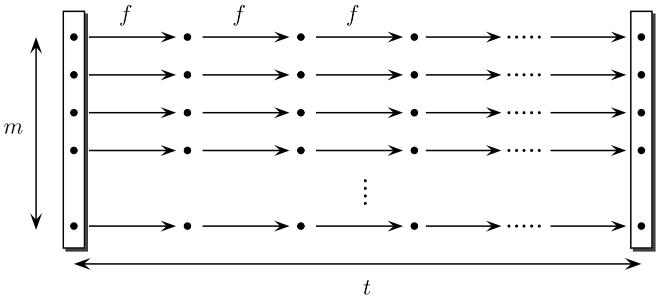
- Construct some tables to cover a fixed fraction of the state space.
- Online Stage: for every successive segment see if present in one of the tables.



State twice the size of Secret Key

Biryukov, Shamir [Asiacrypt 2001]

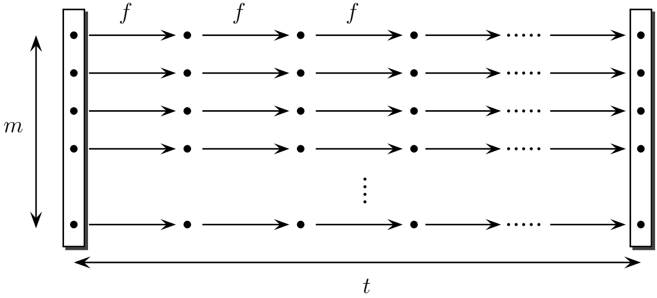
- Total complexity T, memory M, data D, state space N, offline complexity P.
- Get the tradeoff curve $TM^2D^2 = N^2$, with the limitation that $T \geq D^2$.



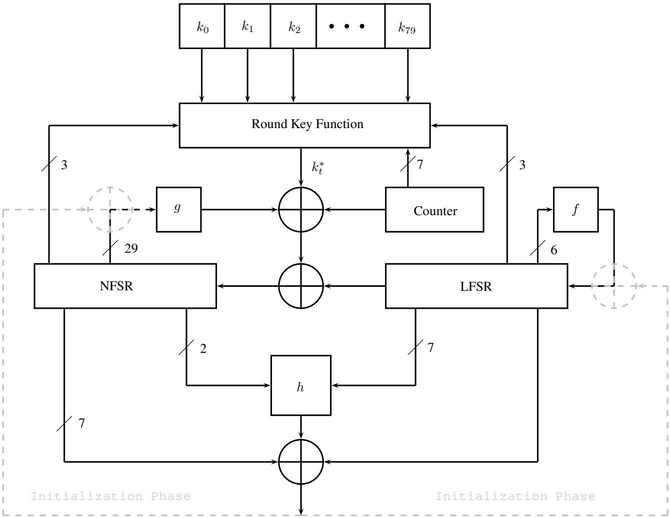
State twice the size of Secret Key

Biryukov, Shamir [Asiacrypt 2001]

- Typical point on curve is $T = N^{2/3}$, $M = N^{1/3}$, $D = N^{1/3}$, $P = N^{2/3}$.
- If $N = K$ this is a valid attack. Rule of the thumb is $N = K^2$.

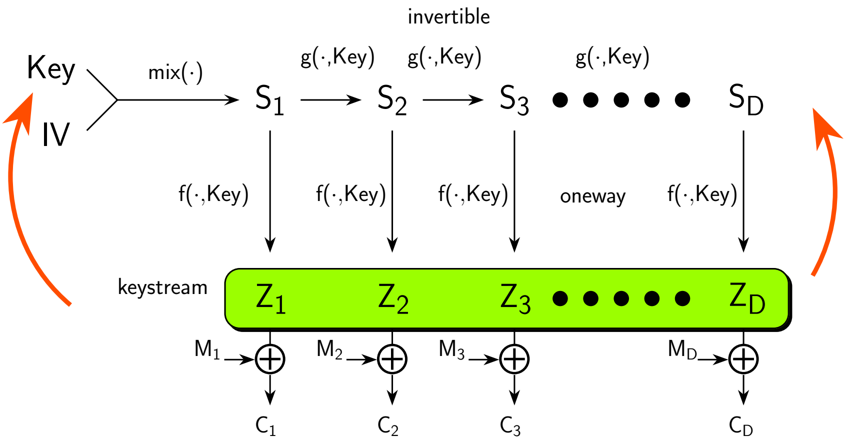


Structure



k_0
k_1
k_2
• • •
k_{79}
Round Key Function
3
k_t^*
7
3
g
Counter
f
29
6
NFSR
LFSR
2
7
h
7
Initialization Phase
Initialization Phase

One way inversion not possible without key



Algebraic Description

Description

- Uses an 80 bit Key and a 70 bit IV.
- Initialization: IV[0 to 39] $\rightarrow$ NFSR, IV[40 to 69]$||$`0x3fe` $\rightarrow$ LFSR
- Key-IV Mixing : Clock 320 cycles without producing Keystream.

  $\rightarrow$ Xor z_t to update functions of NFSR, LFSR.
- Keystream: After 320 cycles, discontinue feedback and produce keystream bit

Algebraic Description

Description

- Update of LFSR :

$$l_{t+40} = f(L_t) = l_t + l_{t+5} + l_{t+15} + l_{t+20} + l_{t+25} + l_{t+34}.$$

- Update of NFSR : $n_{t+40} = g(N_t) + c_t^4 + k_t^* + l_0^t$

$\rightarrow$ c_t^4 denotes the 4^{th} LSB of the modulo 80 up-counter.

$\rightarrow$ k_t^* is the output of the Round Key function defined as:

$$k_t^* = \begin{cases} K_{t \bmod 80}, & \text{if } t < 80, \\ K_{t \bmod 80} \cdot (l_{t+4} + l_{t+21} + l_{t+37} + n_{t+9} + n_{t+20} + n_{t+29}), & \text{otherwise.} \end{cases}$$

$\rightarrow$ The non-linear function g is given as:

$$\begin{aligned} g(N_t) = {} & n_{t+0} + n_{t+13} + n_{t+19} + n_{t+35} + n_{t+39} + n_{t+2}n_{t+25} + n_{t+3}n_{t+5} + \\ & n_{t+7}n_{t+8} + n_{t+14}n_{t+21} + n_{t+16}n_{t+18} + n_{t+22}n_{t+24} + n_{t+26}n_{t+32} + \\ & n_{t+33}n_{t+36}n_{t+37}n_{t+38} + n_{t+10}n_{t+11}n_{t+12} + n_{t+27}n_{t+30}n_{t+31}. \end{aligned}$$

Algebraic Description

Description

- Keystream bit is produced as

$$z_t = l_{t+30} + \sum_{i \in \mathcal{A}} n_{t+i} + h(N_t, L_t).$$

$\rightarrow \mathcal{A} = \{1, 6, 15, 17, 23, 28, 34\}$

$\rightarrow h(N_t, L_t) = n_{t+4}l_{t+6} + l_{t+8}l_{t+10} + l_{t+32}l_{t+17} + l_{t+19}l_{t+23} + n_{t+4}l_{t+32}n_{t+38}.$

Known Attacks

Known Attacks

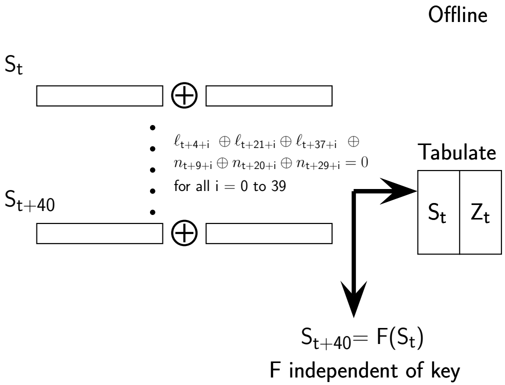
- Related Key Distinguisher : Yonglin Hao [eprint 2015/231]
- Partial State Exposure : Maitra et al [eprint 2015/236]

  → Guess 54 bits of the state.

  → Remaining bits of state and Key found by solving keystream equations in SAT solver.
- Guess and Determine: Lallemand and Naya-Plasencia [CRYPTO 2015]

  → Faster than Brute Force by 2^{10}, takes 2^{46} bits of memory.

Attack by Esgin/Kara (SAC 2015)



Offline Phase

- Note that the key mixing function is non linear.

$$k_t^* = K_{t \bmod 80} \cdot (l_{t+4} + l_{t+21} + l_{t+37} + n_{t+9} + n_{t+20} + n_{t+29})$$

- Enumerate class of states for which

$$l_{t+4} + l_{t+21} + l_{t+37} + n_{t+9} + n_{t+20} + n_{t+29} = 0 \text{ for } t = 0, 1, \ldots, 39$$

Online stage

Online stage

- For every keystream segment try to match in table.
  1 Does not exist in table
  2 Exists in table, but not produced by a weak state
  3 Exists in table, and produced by a weak state

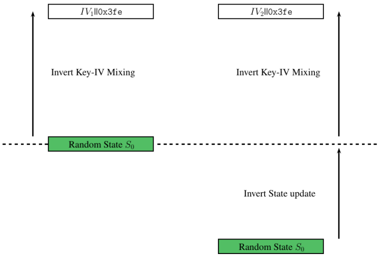
‘

- If match exists: from knowledge of keystream and state: find secret key.
- Use SAT method for this.
- The time complexity is practical 2^{33} encryptions

Sliding Key-IV pairs

Idea

- Fix Secret Key K and experiment with random states S_0



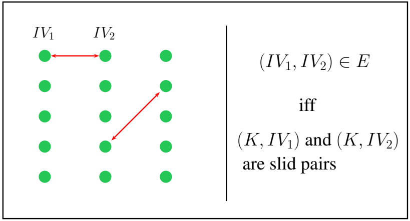
- 2^{20} trials to satisfy both requirements $\rightarrow (K, IV_1)$ and (K, IV_2) are slid pairs.

Sliding Key-IV pairs

Idea

- 2^{80} possible choices of S_0 $\rightarrow$ for every K we have 2^{60} such IV pairs.
- Define a graph $G = (V, E)$ such that

Secret Key K

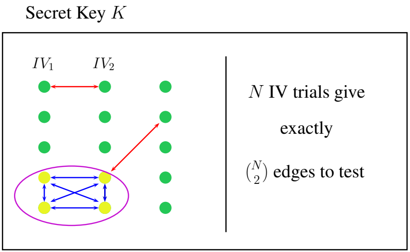


- So we have $|E| = 2^{60}$.

Distinguisher

Attack

- For any K get keystream from random IVs until we get one pair that slide.
- How many random trials necessary ?



- By Birthday rule $\binom{N}{2} \cdot 2^{60} = \binom{2^{70}}{2} \Rightarrow N \approx 2^{40}$ and 2^{48} bits memory.

Distinguisher

Attack

- In general for n bit LFSR and NFSR, Δ bit pad.
- $\binom{N}{2} * 2^{2n-2\Delta} = \binom{2^{2n-\Delta}}{2} \Rightarrow N \approx 2^n$

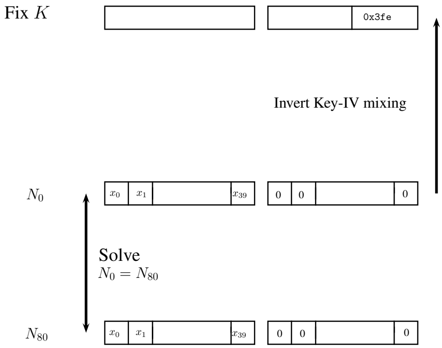
#	n	N (Experimental)	N (Theoretical)
1	8	222.4	256
2	9	446.9	512
3	10	911.7	1024
4	11	1865.7	2048

Table: Experimental values of N for smaller versions of Sprout

Keystream with Period 80

Idea

- If LFSR $=$ All zero vector after Key-IV mixing: it remains all zero forever.
- Key-IV pairs with period 80 Keystream.



- Because of pad, one in 2^{10} random trials will produce success.

Keystream with Period 80

Results

#	K	V
1	`2819 5612 323c 2357 3518`	`2 fbfc75bfcb4396485`
2	`7047 18a0 f88a aff7 7df5`	`1 4d57f42712b395015`

Table: Key-IV pairs that produce keystream sequence with period 80. (Note that the first hex character in V encodes the first 2 IV bits, the remaining 17 hex characters encode bits 3 to 70)

Key Recovery

Attack

- For any K, there exist around 2^{30} IVs that land LFSR to all zero after mixing.

- Algebraic Structure of the cipher is weakened:

  $\rightarrow n_{t+40} = g(N_t) + c_t^4 + k_t^*$

  $\rightarrow k_t^* = K_{t \bmod 80} \cdot (n_{t+9} + n_{t+20} + n_{t+29})$

  $\rightarrow z_t = n_{t+1} + n_{t+6} + n_{t+15} + n_{t+17} + n_{t+23} + n_{t+28} + n_{t+34}.$

- Efficient Guess and Determine possible.

Key Recovery

Attack

- Define $x_i = n_{i+1}$, for all $i \geq 0$.
- For z_0 to z_6 we have the following equations

$$
\begin{aligned}
z_0 &= x_0 + x_5 + x_{14} + x_{16} + x_{22} + x_{27} + x_{33} \\
z_1 &= x_1 + x_6 + x_{15} + x_{17} + x_{23} + x_{28} + x_{34} \\
&\vdots \\
z_6 &= x_6 + x_{11} + x_{20} + x_{22} + x_{28} + x_{33} + x_{39}
\end{aligned}
$$

- Guess x_0 to x_{32} (2^{33} guesses). x_{33} to x_{39} can be determined easily.

$$x_{i+33} = z_i + x_i + x_{i+5} + x_{i+14} + x_{i+16} + x_{i+22} + x_{i+27}$$

Key Recovery

Attack

1. Assign $K_i = \phi, \ \forall i \ \in [0, 79]$
2. For Each of the 2^{33} candidates do the following
   - $\rightarrow$ Assign $i \leftarrow 0$
   - $\rightarrow$ Calculate $x_{i+40} = z_{i+7} + x_{i+7} + x_{i+12} + x_{i+21} + x_{i+23} + x_{i+24} + x_{i+31}$
   - $\rightarrow$ Calculate $k_i^* = x_{i+40} + c_i^4 + g(N_{i+1})$
   - $\rightarrow$ Calculate $m_i = x_{i+8} + x_{i+19} + x_{i+28}$ (note $k_i^* = K_{i \bmod 80} * m_i$)

$$\text{Next Step} = \begin{cases} \text{No Deduction}, & \text{if } k_i^* = 0 \ \wedge \ m_i = 0, \\ \text{Assign } K_{i \bmod 80} = 0, & \text{if } k_i^* = 0 \ \wedge \ m_i = 1 \ \wedge \ K_{i \bmod 80} = \phi, \\ \text{Contradiction}, & \text{if } k_i^* = 0 \ \wedge \ m_i = 1 \ \wedge \ K_{i \bmod 80} = 1, \\ \text{Assign } K_{i \bmod 80} = 1, & \text{if } k_i^* = 1 \ \wedge \ m_i = 1 \ \wedge \ K_{i \bmod 80} = \phi, \\ \text{Contradiction}, & \text{if } k_i^* = 1 \ \wedge \ m_i = 1 \ \wedge \ K_{i \bmod 80} = 0, \\ \text{Contradiction}, & \text{if } k_i^* = 1 \ \wedge \ m_i = 0 \end{cases}$$

   - $\rightarrow$ If Contradiction then Abort and try new guess,
   - $\rightarrow$ Else $i \leftarrow i + 1$ and continue from start.

Complexity

- Abort in 1 out of 4 cases $\leftarrow$ probability $\frac{1}{4}$ of 1st round abort.
- Abort after 2 rounds $\leftarrow$ $(1 - \frac{1}{4}) * \frac{1}{4}$.
- Abort after i rounds $\leftarrow$ $(1 - \frac{1}{4})^{i-1} * \frac{1}{4}$.
- Average number of rounds before elimination:

$$\theta = \sum_{i=1}^{\infty} \frac{i}{4} * \left(1 - \frac{1}{4}\right)^{i-1} = 4.$$

- Try 2^{40} IVs before we get a weak state, so total guesses $= 2^{40} \cdot 2^{33} \cdot 4 = 2^{75}$.
- Equivalent to $2^{66.7}$ encryptions and takes surprisingly little memory.

The stream cipher Plantlet

Changes

- Plantlet proposed in IACR TOSC 2017 by same authors as Sprout.
- Increase state size to 101 bits (40+61).

  $\rightarrow$ Defeats guess and determine attacks
- Key mixing changed to linear i.e. $k_t^* = K[t \bmod 80]$
- To counteract weak states which result from all zero LFSR:

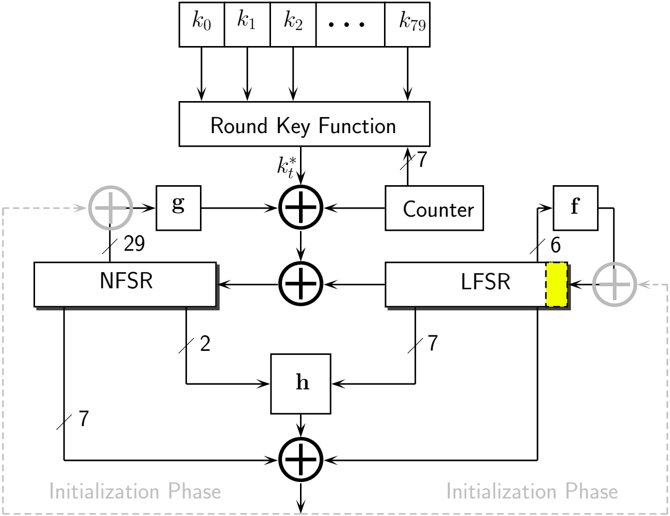
  $\rightarrow$ An interesting solution is provided: 61 bit LFSR used in 2 phases

  $\rightarrow$ During Key-IV mixing only the first 60 bits are updated: 61st bit held at 1.

  $\rightarrow$ Full 61 bits are updated only during keystream phase.

  $\rightarrow$ LFSR never becomes all zero.

Structure



k_0
k_1
k_2
...
k_{79}
Round Key Function
k_t^*
7
g
Counter
f
29
6
NFSR
LFSR
2
7
h
7
Initialization Phase
Initialization Phase

The stream cipher Plantlet

Changes

- LFSR update : During Key IV mixing

$$
\begin{aligned}
l_{60}^{t+1} &= 1\\
l_{59}^{t+1} &= l_{54}^{t} + l_{43}^{t} + l_{34}^{t} + l_{20}^{t} + l_{14}^{t} + z^{t}\\
l_{i}^{t+1} &= l_{i+1}^{t}, \text{ for } 0 \le i \le 58
\end{aligned}
$$

- LFSR update : During keystream phase

$$
\begin{aligned}
l_{60}^{t+1} &= l_{54}^{t} + l_{43}^{t} + l_{34}^{t} + l_{20}^{t} + l_{14}^{t} + z^{t}\\
l_{i}^{t+1} &= l_{i+1}^{t}, \text{ for } 0 \le i \le 59
\end{aligned}
$$

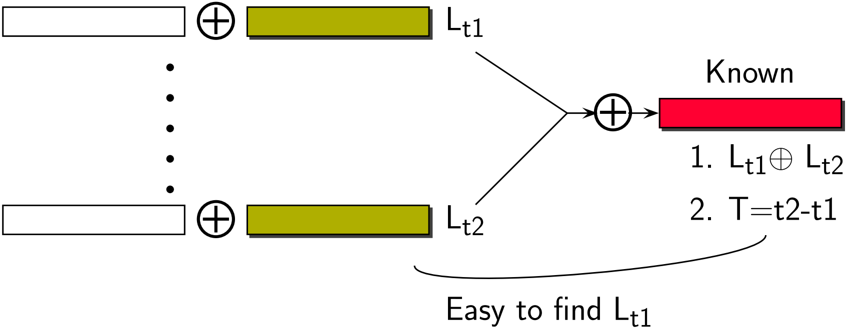
- Both LFSR functions have maximum period.

Changes

- This does not solve the problem of distingusing attacks using slid keystream
- The authors have admitted as much in the paper.
- But it is difficult to convert the distinguisher into a key recovery attack.
- Also only 2^{30} keystream bits are allowed per key-IV pair.

Plantlet: Observation 1 [IACR ePrint 2019/702]

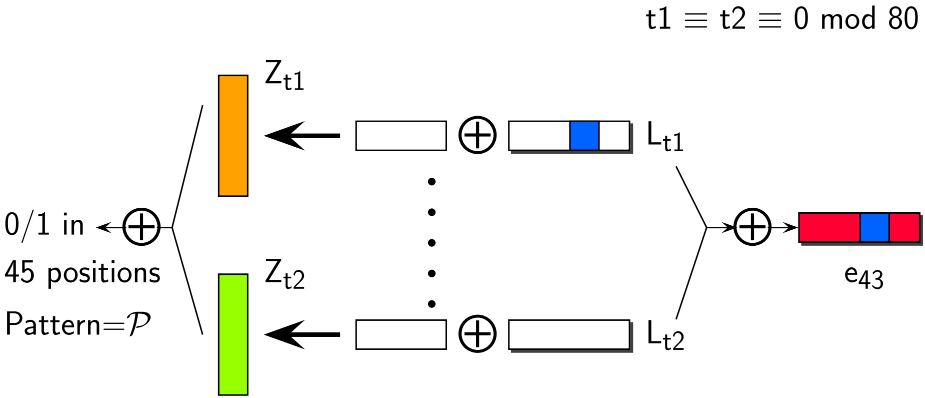
EPFL



How

- $L_{t2} = M^{t2-t1} \cdot L_{t1} \Rightarrow L_{t2} \oplus L_{t1} = (I \oplus M^T) \cdot L_{t1}$
- System of linear equations, $(I \oplus M^T)$ is always invertible.

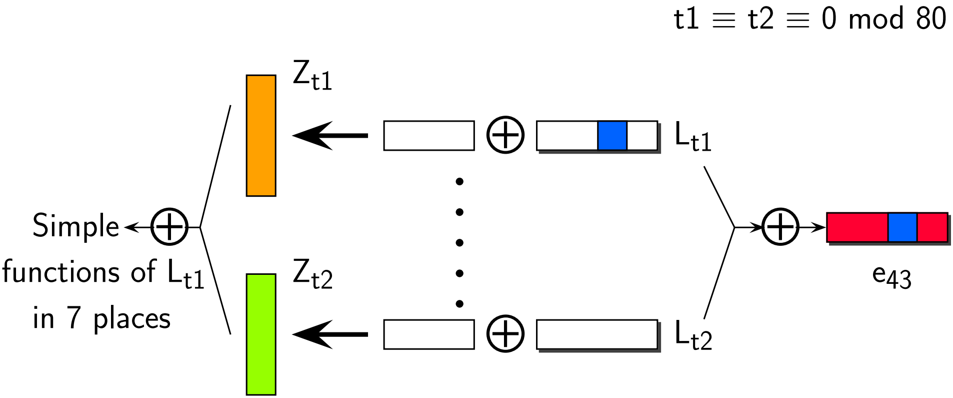
Plantlet: Observation 2



How

- This gives us an interesting filter.
- However the opposite direction is **NOT TRUE**.

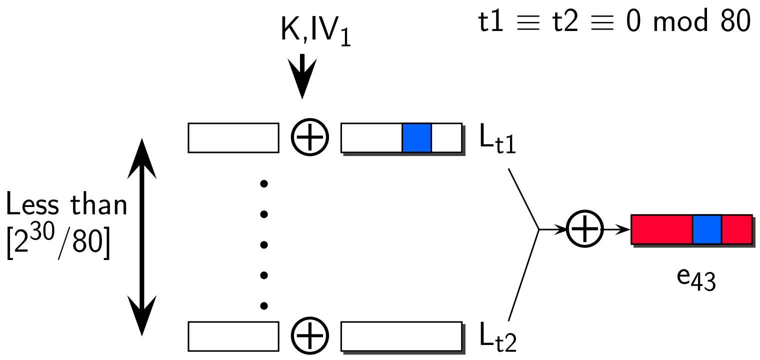
Plantlet: Observation 3



How

- Helps reduce complexity more (we will see how).
- Also $z_{t1+46} + z_{t2+46} = n_{t1+50} \cdot l_{t1+78}$.

Plantlet: Attack

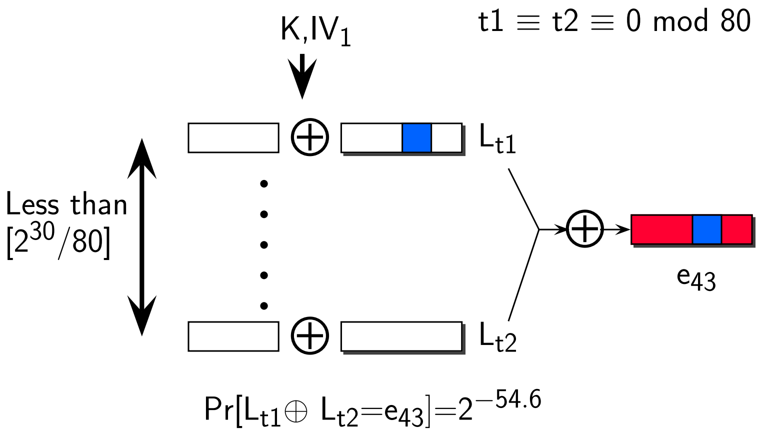


$$\Pr[L_{t1} \oplus L_{t2} = e_{43}] = 2^{-54.6}$$

How

- The probability that for a single IV this happens is $\approx 2^{-55}$.
- Note that not more than 2^{30} keystream bits are allowed for one IV.

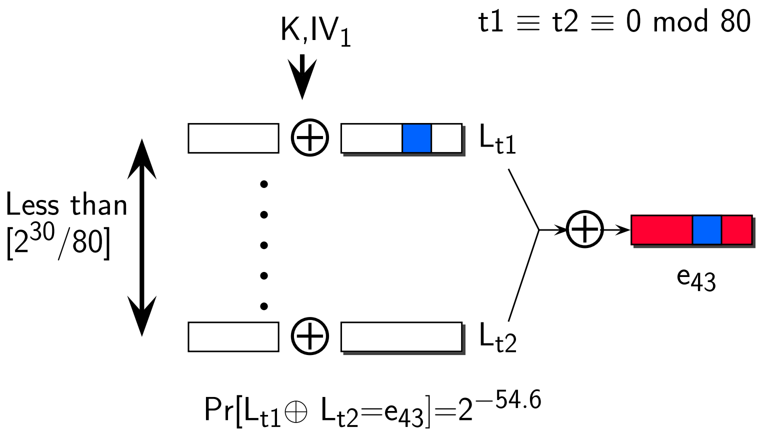
Plantlet: Attack



How

- The probability that for a single IV this happens is $\approx 2^{-55}$.
- For 2^{55} IVs we get one hit on average !!!!

Plantlet: Attack



How

- When you get a hit: first recover L_{t1} (e_{43} and $T = t2 - t1$ known).
- From polynomial eqn of z_{t1+i} solve for NFSR+Secret key !!!!

Plantlet: Attack

Remaining paper is how to make it happen

A: Generate 2^{30} keystream bits key and random IV.

B: For all $t = 80 \cdot i$ where $i \in [1, N-1]$, store in a hash table t, Z_t as defined.

C: Find, if it exists, t_1, t_2 so that $\mathcal{P} = Z_{t_1} \oplus Z_{t_2}$

D: If exists assume that the state differential is $0^{40}||e_{43}$.

E: Try to solve for the remaining system of equations to find the key.

F: If a contradiction is reached, try other values of t_1, t_2 or another IV.

Part A: Precomputation

Pre solve linear system

A: All linear systems of form $e_{43} = (I + M^T) \cdot L_t$

B: T is less than $[2^{30}/80] \approx 2^{24}$.

C: Use Gaussian elimination to solve all such systems

D: Solutions can be stored as T, L_T in the memory

E: Less than 2^{42} steps and less than 2^{30} bits of memory

Part B: Collect keystream bits

Look for pattern $\mathcal{P}$

A: For each IV collect keystream bits

B: The idea is to find $t1$ and $t2$ so that $Z_{t1} + Z_{t2} = \mathcal{P}$.

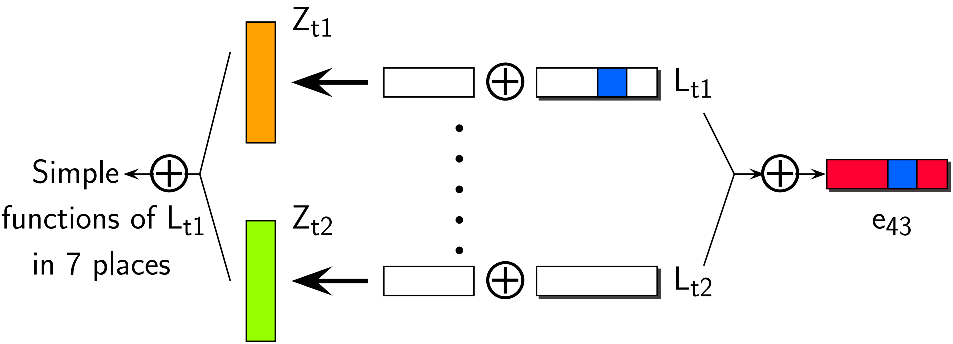
C: Use a good data structure to store keystream

D: If $Z_{t1} + Z_{t2} = \mathcal{P} \Rightarrow L_{t1} + L_{t2} = e_{43}$ (Not always true)

E: Pick up L_{t1} from precomputed table.

Part C: Filter further

$t1 \equiv t2 \equiv 0 \bmod 80$



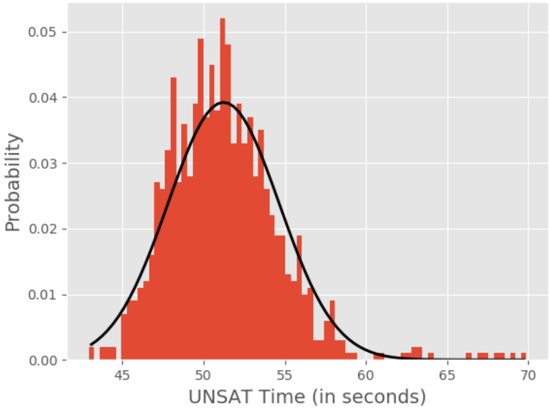
Look for further filtering

A: For 7 values of i, $z_{t1+i} + z_{t2+i} =$ simple function of L_{t1}

B: If the above does not hold for L_{t1} from offline table $\Rightarrow$ Reject

C: If not use SAT solver for next stage

Part D: Use Solver



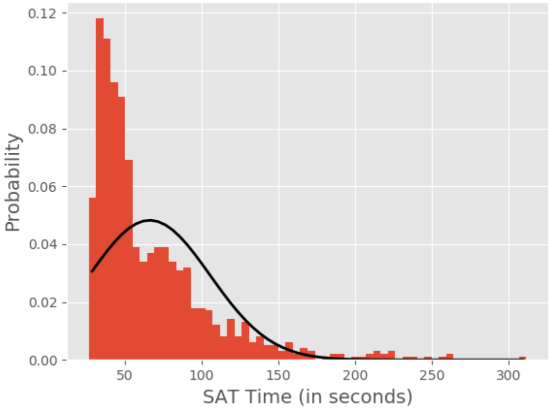
Solver stats

A: Form polynomial equations for all z_{t1+i} in NFSR, Key variables

B: Ask a solver to solve them

C: If assumption was incorrect solver returns UNSAT

Part D: Use Solver



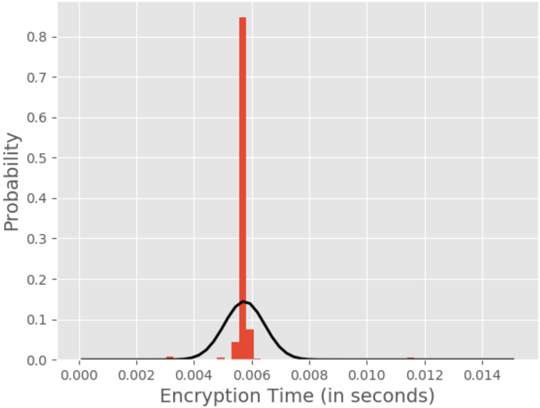
Solver stats

A: Form polynomial equations for all z_{t1+i} in NFSR, Key variables

B: Ask a solver to solve them

C: If assumption was correct solver returns key/NFSR state

Part D: Use Solver



Solver stats

A: We can only estimate this complexity in terms of Plantlet encryption.

B: Compute average time on seconds to compute Plantlet enc.

C: Take the ratio between the two as an estimate.

Conclusion

Conclusion

A: We have one more optimization stage.

B: We find key in around 2^{70} Plantlet encryptions

C: Please read the paper for analysis of complexity.

EPFL

What now ?

A: Small state stream ciphers.

B: Sprout, Plantlet, Fruit cryptanalyzed.

C: Lizard has a distinguisher and some other undesirable results.

D: Maybe a research direction is to put together another design.

THANK YOU