CONJURA · OPEN PROBLEM

The Censoring Conjecture for AES

Deleting linear mixing layers can only hurt approximate t-wise independence

Status. Statement: AI-written from Vishesh Jain, Tianren Liu, Clayton Mizgerd, Angelos Pelecanos, Stefano Tessaro, Vinod Vaikuntanathan, *How Fast Does the Inverse Walk Approximate a Random Permutation?*, Cryptology ePrint Archive 2025, not yet checked by a human. No instance of the implication is settled; what the source settles is the hypothesis side, namely the round count at which the censored construction becomes approximately t-wise independent.

Category. *Symmetric-Key Cryptography.*

Conjecture (informal). *A substitution-permutation network like AES alternates three kinds of operation: adding a round key, applying the S-box to each byte independently, and applying a linear layer that mixes the bytes together. A line of work on provable security for AES proves statistical guarantees not for AES itself but for a weakened relative called* censored *AES, in which many of the linear mixing layers are deleted and replaced by the identity. Censoring makes the analysis tractable, because long runs of key-addition and S-box applications with no mixing in between behave like a random walk that can be studied one byte at a time. The guarantee proved is approximate t-wise independence: on any t distinct inputs, the joint distribution of the t outputs, over the choice of round keys, is close in total variation distance to a uniformly random tuple of t distinct values. Everyone believes the mixing layers help rather than hurt, so the censored cipher should be the harder case and the bound should carry over to full AES with the same round count. Nobody has proved this. The conjecture is exactly that transfer, with a small allowance for the first few rounds, during which the cipher may still behave atypically.*

1 The setting

Proving unconditional security statements about AES [DR02] is out of reach, so a line of work initiated by Liu, Tessaro and Vaikuntanathan [LTV21] and continued by Liu, Pelecanos, Tessaro and Vaikuntanathan [LPTV23] instead proves a statistical property: *ε-approximate t-wise independence*. A cipher with this property resists every attack that looks at only t input-output pairs and distinguishes statistically, which for $t = 2$ covers differential and linear cryptanalysis and for larger t covers higher-order differential attacks.

The technical difficulty is the substitution layer. It is sidestepped in [LPTV23] by analysing SPN*, that paper's generalisation of the AES* construction of Baignères and Vaudenay [BV05], and proving that for $t < 2^{(0.499-1/(4k))b}$ a $\Theta(k)$-round SPN* with maximal-branch-number mixing is $2^{-\Theta(bk)}$-approximately t-wise independent for k blocks of b bits. In SPN* each S-box is a uniformly random secret permutation rather than the fixed AES S-box. To instantiate this with the concrete AES S-box, each random S-box is replaced by a run of consecutive AddRoundKey and INV operations — the "INV key-alternating cipher" — and the linear mixing layers that fall inside such a run are *censored*. The resulting object is what [LPTV23] call censored AES, and it is the object their theorems are about: for instance, 192-round censored AES is 2^{-128}-approximately pairwise independent.

The present source removes the remaining quantitative bottleneck on the censored side. It determines the mixing time of the inverse walk exactly: over a field of size n, $r = \Theta(n \log n + n \log(1/\varepsilon))$ rounds of the INV key-alternating cipher suffice and are necessary to come within total variation distance ε of a uniformly random permutation of the appropriate sign, and it proves matching $\Omega(n \log(1/\varepsilon))$ and $\Omega(n \log n)$ lower bounds already against 4-wise independence. Feeding this into the SPN* machinery upgrades the censored-AES statement from $t = 2$ to all t up to roughly the square root of the field size.

Progress to date. The censored side of the implication is established at essentially optimal round count, as just described, with matching lower bounds. Combined with the SPN* analysis of [LPTV23] this yields, for $t < 2^{(0.499-1/(4k))b}$, that a $\Theta(b2^b k)$-round censored SPN with k blocks of b bits and the AES S-box is

$2^{-\Theta(kb)}$-approximately t-wise independent. Towards the conjecture itself, nothing is proved: the only reported analogue is the censoring theorem of Peres and Winkler [PW11], showing that extra updates cannot speed up mixing, whose hypotheses — Glauber dynamics in a monotone spin system started from an extremal configuration — are far from anything a block cipher satisfies. How censoring affects mixing time in general is itself unresolved.

What remains between these theorems and AES itself is precisely the censoring step. The conjecture below says that it is free: that adding the deleted mixing layers back can only improve approximate t-wise independence, so that a bound for r-round censored AES transfers to r-round AES, modulo a threshold number of initial rounds.

2 Notation and parameters

- $\mathbb{F} := \mathbb{F}_{2^8}$, the AES byte field; the block cipher acts on $\mathbb{F}^{16}$ (that is, on 128-bit blocks viewed as 16 bytes).
- $\mathrm{INV} : \mathbb{F} \to \mathbb{F}$, $\mathrm{INV}(x) = x^{2^8-2}$, the patched inverse permutation: $\mathrm{INV}(x) = x^{-1}$ for $x \neq 0$ and $\mathrm{INV}(0) = 0$. This is the AES S-box up to a fixed affine map, which plays no role below.
- $S : \mathbb{F}^{16} \to \mathbb{F}^{16}$, $S(x_1, \ldots, x_{16}) = (\mathrm{INV}(x_1), \ldots, \mathrm{INV}(x_{16}))$, the blockwise substitution layer.
- $L : \mathbb{F}^{16} \to \mathbb{F}^{16}$, the linear mixing layer, an $\mathbb{F}_2$-linear bijection with maximal branch number (the source does not decompose it further).
- $\mathrm{ARK}_K : \mathbb{F}^{16} \to \mathbb{F}^{16}$, $\mathrm{ARK}_K(x) = x + K$, the AddRoundKey permutation for $K \in \mathbb{F}^{16}$.
- $r \in \mathbb{N}$, the number of rounds; $\vec{K} = (K_0, \ldots, K_r) \in (\mathbb{F}^{16})^{r+1}$, the round keys, drawn independently and uniformly (the round keys are modelled as independent, not as the output of the AES key schedule).
- $t \in \mathbb{N}$ with $t \geq 2$, the order of independence; $\varepsilon \in (0,1)$, the statistical slack.

- $\|\mu - \nu\|_{\mathrm{TV}} := \frac{1}{2}\sum_{\omega} |\mu(\omega) - \nu(\omega)|$, the total variation distance between distributions on a finite set.
- $C \subseteq \{1, \ldots, r\}$, the set of rounds whose linear mixing layer is censored; $C^*(r)$, the censoring set of r-round censored AES (Definition 2).
- r_{thresh}, the round threshold below which no claim is made; it depends on t alone.

3 The conjecture

Definition 1 (C-censored r-round AES). Let $r \geq 1$ and let $C \subseteq \{1, \ldots, r\}$. For $i \in \{1, \ldots, r\}$ put

$$L_i^C := \begin{cases} \mathrm{id}_{\mathbb{F}^{16}}, & i \in C, \\ L, & \text{otherwise.} \end{cases}$$

For a key vector $\vec{K} = (K_0, K_1, \ldots, K_r) \in (\mathbb{F}^{16})^{r+1}$ define the permutation

$$\begin{aligned} E_{C,\vec{K}}^{(r)} := \mathrm{ARK}_{K_r} &\circ \left(L_r^C \circ S \circ \mathrm{ARK}_{K_{r-1}}\right) \\ &\circ \left(L_{r-1}^C \circ S \circ \mathrm{ARK}_{K_{r-2}}\right) \circ \cdots \\ &\circ \left(L_1^C \circ S \circ \mathrm{ARK}_{K_0}\right) \end{aligned}$$

of $\mathbb{F}^{16}$, and let $\mathcal{F}_C^{(r)}$ denote the distribution of $E_{C,\vec{K}}^{(r)}$ when $\vec{K}$ is drawn uniformly from $(\mathbb{F}^{16})^{r+1}$. The case $C = \emptyset$ is (round-key-independent) r-round AES, which applies L in every one of its r rounds and performs $r + 1$ AddRoundKey operations; a nonempty C records the rounds whose linear mixing layer has been censored, i.e. replaced by the identity.

Definition 2 (the censoring set of censored AES). Fix the instantiation of [LPTV23] in which each SPN* S-box layer is replaced by a run of consecutive (ARK, INV) operations. Then $C^*(r) \subseteq \{1, \ldots, r\}$, the censoring set of r-round censored AES,

is the set of rounds whose linear mixing layer falls strictly inside one of these runs.

Definition 3 (ε-approximate t-wise independence). Let D be a finite set, let $2 \le t \le |D|$, and let μ be a probability distribution on the group of permutations of D. Say that μ is *ε-approximately t-wise independent* if for every tuple $(x_1, \dots, x_t)$ of distinct elements of D,

$$\left\| \mathcal{L}\big((f(x_1), \dots, f(x_t))\big)_{f \sim \mu} \; - \; \mathcal{U}_t \right\|_{\mathrm{TV}} \le \varepsilon,$$

where $\mathcal{L}(\cdot)$ denotes the law of a random variable and $\mathcal{U}_t$ is the uniform distribution on ordered t-tuples of distinct elements of D.

Conjecture 1 (censoring can only hurt). *For every integer $t \ge 2$ there exists a constant $r_{\mathrm{thresh}} = r_{\mathrm{thresh}}(t) > 0$ such that the following holds. For every integer $r > r_{\mathrm{thresh}}$ and every $\varepsilon \in (0,1)$,*

$$\mathcal{F}^{(r)}_{C^*(r)} \text{ is } \varepsilon\text{-approximately } t\text{-wise independent} \quad \Longrightarrow \quad \mathcal{F}^{(r)}_{\emptyset} \text{ is } \varepsilon\text{-approximately } t\text{-wise independent,}$$

where $\mathcal{F}^{(r)}_{C}$ is the distribution of the C-censored r-round AES cipher of Definition 1 under independent uniform round keys, $C^*(r)$ is the censoring set of Definition 2, $\mathcal{F}^{(r)}_{\emptyset}$ is the corresponding distribution for uncensored r-round AES, and ε-approximate t-wise independence is as in Definition 3 with domain $D = \mathbb{F}^{16}$.

Remark 1 (which censoring pattern). The hypothesis concerns the one distinguished set $C^*(r)$ of Definition 2, not an arbitrary one. The variant that quantifies universally over all $C \subseteq \{1, \dots, r\}$ is a strengthening the source does not pose, so a counterexample at some exotic C would not by itself refute Conjecture 1; the pattern of Definition 2 is the target. The informal justification offered — that removing a large fraction of the mixing layers should only hurt convergence to t-wise independence — is pattern-agnostic, which is what makes the stronger reading tempting.

Remark 2 (quantifier order and the threshold). The slack ε is read here as universally quantified, and r_{thresh} depends on t alone. The threshold exists only to let the cipher enter a generic state and so to excuse any non-typical behaviour during the first few rounds. If the intended reading lets r_{thresh} depend on ε as well, the conjecture is weaker than the one stated here.

Remark 3 (independent round keys). Both sides of the implication use round keys drawn independently and uniformly, as throughout this line of work. The AES key schedule is out of scope, and Conjecture 1 says nothing about it.

Remark 4 (what settling it would give). Everything this programme currently proves about the concrete AES S-box is proved about a cipher nobody uses: censored AES, with a large fraction of its diffusion deliberately removed. The conjecture is the single missing step that would turn those statements into statements about AES with independent round keys — immediately, that 192-round AES is 2^{-128}-approximately pairwise independent, and, with the mixing-time theorem quoted in Section 1, approximate t-wise independence of AES for t up to roughly the square root of the S-box domain size. A refutation would be at least as interesting: it would exhibit a diffusion layer that actively destroys statistical independence over a fixed round budget, contradicting the design intuition the whole SPN paradigm rests on.

4 Bibliography

[BV05] T. Baignères, S. Vaudenay. *Proving the Security of AES Substitution-Permutation Network.* Selected Areas in Cryptography, 12th International Workshop, SAC 2005, LNCS 3897, pages 65–81, Springer, 2005.

[DR02] J. Daemen, V. Rijmen. *The Design of Rijndael: AES — The Advanced Encryption Standard.* Information Security and Cryptography, Springer, 2002.

[LPTV23] T. Liu, A. Pelecanos, S. Tessaro, V. Vaikuntanathan. *Layout Graphs, Random Walks and the t-wise Independence of SPN Block Ciphers.* Advances in Cryptology — CRYPTO 2023, LNCS 14083, pages 694–726, Springer, 2023.

[LTV21] T. Liu, S. Tessaro, V. Vaikuntanathan. *The t-wise Independence of Substitution-Permutation Networks.* CRYPTO 2021.

[PW11] Y. Peres, P. Winkler. *Can Extra Updates Delay Mixing?* 2011. No venue given in the source bibliography.