Towards Optimally Secure Deterministic Authenticated Encryption Schemes

Yu Long Chen[1], Avijit Dutta[2,3], Ashwin Jha[4], and Mridul Nandi[5]

[1] COSIC, KU Leuven, Leuven, Belgium
yulong.chen@esat.kuleuven.be
[2] IAI TCG CREST, Kolkata, India
[3] AcSIR, Ghaziabad, India
avirocks.dutta13@gmail.com
[4] Ruhr-University of Bochum, Bochum, Germany
letterstoashwin@gmail.com
[5] Indian Statistical Institute, Kolkata, India
mridul.nandi@gmail.com

Abstract. The public comments received for the review process for NIST (SP) 800-38A pointed out two important issues that most companies face: (1) the limited security that AES can provide due to its 128-bit block size and (2) the problem of nonce-misuse in practice. In this paper, we provide an alternative solution to these problems by introducing two optimally secure deterministic authenticated encryption (DAE) schemes, denoted as DENC1 and DENC2 respectively. We show that our proposed constructions improve the state-of-the-art in terms of security and efficiency. Specifically, DENC1 achieves a robust security level of $O(r^2\sigma^2\ell/2^{2n})$, while DENC2 attains a near-optimal security level of $O(r\sigma/2^n)$, where σ is the total number of blocks, ℓ is maximum number of blocks in each query, and r is a user-defined parameter closely related to the rate of the construction. Our research centers on the development of two IV-based encryption schemes, referred to as IV1 and IV2, which respectively offer security levels of $O(r^2\sigma^2\ell/2^{2n})$ and $O(r\sigma/2^n)$. Notably, both of our DAE proposals are nearly rate 1/2 constructions. In terms of efficiency, our proposals compare favorably with state-of-the-art AE modes on contemporary microprocessors.

Keywords: IV-based encryption, Deterministic AE, SIV, GCM-SIV

1 Introduction

In May 2021, NIST initiated a review process for (SP) 800-38A: Recommendation for Block Cipher Modes of Operation: Methods and Techniques and its

An abridged version of this article appears in EUROCRYPT 2025. The authors would like to thank the program committees of ASIACRYPT 2022, EUROCRYPT 2024, CRYPTO 2024 and EUROCRYPT 2025 for their constructive comments. Ashwin Jha's work was supported by the German Research Foundation (DFG) within the framework of the Excellence Strategy of the Federal Government and the States – EXC 2092 CaSa – 39078197. Mridul Nandi's work was supported by MATRICS scheme MTR/2023/000302 of Science and Engineering Research Board, India.

addendum, Three Variants of Ciphertext Stealing for CBC Mode[16]. Two key concerns raised in the public comments[15,14] were the issue of nonce-misuse in practice and the limited security provided by AES due to its 128-bit block size.

In authenticated encryption (AE), using nonces is a standard practice to enhance security by ensuring non-repetition. However, even a single instance of nonce reuse can potentially compromise the entire scheme. For example, in GCM[29], reusing a nonce exposes the hash key. Ensuring nonce uniqueness requires a stateful implementation, which is not always practical, especially in cases where multiple devices share the same key but cannot reliably synchronize. Additionally, flawed implementations can inadvertently cause nonce reuse. A comprehensive Internet-wide scan conducted by Böck et al.[5] revealed several HTTPS servers that always use the same nonce, and many others that employ random 64-bit nonces. Due to the birthday problem, the latter approach drastically limits the number of messages that can be safely encrypted.

On the other hand, the GCM construction only provides security that approaches the $2^{n/2}$ birthday bound [29,26,20,43]. However, this security bound may be on the edge of acceptability due to the second problem we mentioned about: the underlying block cipher has a limited block size.

Grimm highlighted that, in practical implementations, certain designs generate a fresh AES key for each message or data item, while employing a constant nonce value for every encryption operation [15,14]. To enhance security beyond the limited 2^{64} birthday bound and to fortify resilience for customers, he recommends that NIST should consider the development of two new standards:

1. A novel block cipher with a substantial key and state size, while still maintaining software performance levels comparable to AES, particularly when utilizing AES-NI instructions.
2. Utilization of this newly devised cipher to construct an authenticated encryption scheme featuring authenticated data, with a crucial requirement that the mode should not necessitate a nonce generated by the caller.

Nonetheless, the first point seems difficult to achieve in the practice. On one hand, the task of designing a novel block cipher is far from straightforward. It required two decades of dedicated research to establish the level of trust and confidence that AES enjoys today. Conversely, demanding equivalent software performance while working with a primitive that possesses a substantially larger state size compared to AES presents its own set of challenges. On the other hand, as mentioned in the NSA presentation at the NIST workshop [46], it is extremely difficult to replace AES with another cipher due to its large economic impact [27] and the fact that it is used in millions of products. However, it is important to acknowledge that there are alternative solutions to each of these two problems.

Solution For The Unique-Nonce Problem: A comprehensive solution to address the unique-nonce challenge involves the adoption of a misuse-resistant AE (MRAE) scheme. This concept, initially introduced by Rogaway and Shrimpton [43], operates by employing a pseudo-random function (PRF) construction denoted as F to compute an authentication tag referred to as IV. Subsequently, it applies an IV-based encryption algorithm, utilizing IV as input, to encrypt the message.

This MRAE concept has been further refined by assessing the adversarial distinguishing advantage in terms of parameters such as the maximum number of multicollisions in nonce values [42] or the maximum number of reused nonce values [19]. In contrast, the MRAE notion transforms a deterministic AE (DAE) scheme into a nonce-based MRAE scheme, where the security of such AE schemes becomes independent of the number of nonce repetitions. While DAE schemes still accept an input nonce, they no longer adhere to the constraint of non-repetition. Therefore, the input nonce can be considered a part of the associated data A. In this scenario, the sole requirement is that the pair (A, M) is never duplicated.

Beyond Birthday Bound Secure AE Schemes: In recent years, there has been a surge in the introduction of block cipher-based AE schemes that offer security beyond the birthday bound. For example, the well-known AES-GCM-SIV, which has been demonstrated to offer $3n/4$-bit security when nonces are unique, and $n/2$-bit security in the event of nonce repetition, as established by Iwata and Seurin [24]. These security claims are contingent upon assuming the multi-user security of AES in the standard model. Conversely, Bose et al. [6] revealed that by employing a different key derivation function, the nonce-respecting security of AES-GCM-SIV in the ideal cipher model can be elevated to a robust n-bit level.

In 2006, Iwata [21] introduced an approach to reuse some of the block cipher evaluations, and proposed a block cipher based encryption scheme CENC and a nonce based AE scheme CHM. The security of these constructions can be reduced to the security of the sum of permutations [4] construction. Consequently, both CENC and CHM achieve nearly optimal n-bit security under the assumption that the input nonce is never repeated. A more recent addition, the SCM mode [10], adopts a similar approach as the CENC and CHM modes. Unlike its predecessors, SCM exhibits graceful security degradation in the faulty nonce model [19] when nonces are reused. Consequently, SCM maintains security levels beyond the birthday bound when the number of repeated nonces is limited. Unfortunately, as the authors have demonstrated, when instantiated with random nonces, the SCM mode can only attain birthday-bound security.

It is worth to mention that there are a number of AE schemes build on tweakable block ciphers that includes [42,28,23,35,32,33,34]. While there are several security advantages associated with designing an authenticated encryption (AE) scheme based on tweakable block ciphers, it is worth noting that block cipher-based schemes offer the convenience of instantiation with widely adopted block ciphers like AES. Consequently, this article's primary focus is on the following research question: Can we design a DAE scheme that achieves n-bit security while minimizing the use of block cipher calls?

1.1 Our Contributions

This paper makes a significant contribution by providing a positive response to the aforementioned question. Specifically, we introduce two optimally secure DAE schemes named $\mathsf{DENC1}$ and $\mathsf{DENC2}$, demonstrating that $\mathsf{DENC1}$ offers a security level of $O(n^2\sigma^2\ell/2^{2n})$, while $\mathsf{DENC2}$ achieves a security level of $O(n\sigma/2^n)$ when these constructions are instantiated with an n-bit block cipher, where ℓ denotes the maximum bit-length for a single message and σ denotes the total bit-length across all messages encrypted before rekeying. The most innovative part of our research centers on the development of two IV-based encryption (IVE) schemes, which we call $\mathsf{IV1}$ and $\mathsf{IV2}$. The structure of our proposal is reminiscent of Iwata's CENC construction [21]. However, the security of CENC is completely dependent on the unique input nonce, when the nonce is repeated, one can trivially break the construction. While in the case of DAE, this uniqueness is not available, instead we only have a uniform random tag to serve as input. Naive use of the tag will result in birthday bound security. To solve this problem, we double the tag length by introducing two block cipher calls during the MAC part of the DAE, and create input masks for our IVE schemes using this $2n$-bit random tag. Our scheme explicitly relies on the use of masks in the input, along with a CENC like transformation in the output, and the security of both $\mathsf{IV1}$ and $\mathsf{IV2}$ is dominated by the used mask.

$\mathsf{IV1}$: LENGTH DEPENDENT n-BIT SECURE IV-BASED ENCRYPTION. As our first contribution, we propose the $\mathsf{IV0}$ paradigm in Section 4, which takes a 2-wise independent sequence as the input and provides a security bound of $O(r\sigma/2^n)$. However, for a fixed parameter r, $r+1$ block cipher calls are needed to generate r blocks with n-bit output, leading to a rate of $r/(r+1)$. Therefore, the security of the $\mathsf{IV0}$ structure decreases dramatically once r becomes large. To avoid this problem and to create an IV-based encryption scheme with variable output length, we introduce $\mathsf{IV1}$ in Section 5. This construction can be seen as an iteration of chunks of r blocks. The output of each chunk is produced by evaluating the $\mathsf{IV0}$ construction, and therefore requires $r+1$ block cipher calls. We show that, given 2-wise independent sequences as input to each block of each chunk, $\mathsf{IV1}$ yields a security bound of $O(r^2\sigma^2\ell/2^{2n})$, where ℓ denotes the maximum number of message blocks in a message.

$\mathsf{IV2}$: LENGTH INDEPENDENT n-BIT SECURE IV-BASED ENCRYPTION. The security of the $\mathsf{IV1}$ construction can decrease dramatically if the message length ℓ becomes large. For example, if r is small and $\ell = O(2^{n/4})$, then $\mathsf{IV1}$ is secured upto $2^{7n/8}$ blocks[6]. To achieve truly message length-independent near-optimal n-bit security, we introduce $\mathsf{IV2}$ in Section 6 as our second contribution. The underlying design philosophy of $\mathsf{IV2}$ is very similar to that of $\mathsf{IV1}$. It is important to emphasize that the security of $\mathsf{IV2}$ depends on the property of the input string on which it is used. To create such an input string for the k-th block of the j-th chunk of i-th query, we use the HtmB paradigm outlined in [12] as follows:

$$(k_1, k_2) = \mathsf{mBenes}(u[1] \oplus \langle j \rangle_n, u[2] \oplus \langle j \rangle_n),$$

[6] In practice, the message length is usually smaller than $2^{n/4}$ blocks.

where $(u[1], u[2])$ is the initial $2n$-bit input random IV of the construction, $\langle j \rangle_n$ denotes the n-bit binary representation of the integer j, and mBenes is the modified Benes function [1]. We have shown that using this input sequence, IV2 achieves an optimal security bound of $O(r\sigma/2^n)$. However, it is important to note that IV1 achieves security bound roughly in the order of $O(r^2\sigma^2\ell/2^{2n})$ by invoking $r+1$ block cipher calls in each chunk. On the other hand, IV2 achieves an ℓ-free security bound, i.e., $O(r\sigma/2^n)$, at the cost of making six extra block cipher calls in each chunk and two additional block cipher keys to process the mBenes paradigm compared to the IV1 construction. However, the increasing key size does not seem to be a big problem for practical applications, since Microsoft is already requesting for a new block cipher using 512-bit secret key [15]. Hence, IV2 requires a total of $r+7$ block cipher calls for each chunk, where r represents the size of each chunk, leading to a rate of $r/(r+7)$.

DENC1 AND DENC2: TOWARDS BUILDING OPTIMALLY SECURE DAE. Our final contribution is to propose two block cipher based deterministic authenticated encryption schemes, called DENC1 and DENC2. Both DENC1 and DENC2 follow the SIV paradigm proposed by Rogaway and Shrimpton [43], where a variable-input length pseudorandom function is used to process the associated data A[7] and a message M to yield a $2n$-bit authentication tag IV, which in turn is used in an IV-based encryption scheme to encrypt the message M. For both DENC1 and DENC2, we use F^*, a variant of the 2k-mPMAC+-p2 [12] construction, as the underlying variable input-length PRF. However, we combine it with the IV1 construction as the underlying IV-based encryption scheme to yield our first DAE scheme DENC1. On the other hand, we combine F^* with IV2 as the underlying IV-based encryption scheme to yield another DAE scheme DENC2. We prove that DENC1 has a security bound of $O\left(r^2\sigma^2\ell/2^{2n}\right)$ (in Corollary 7.1), and DENC2 has a security bound of $O\left(r\sigma/2^n\right)$ (in Corollary 7.2), where σ is the total number of encrypted message blocks and ℓ the maximum number of blocks allowed in a message. Note that DENC1 requires a total of $2\ell + \ell/r$ block cipher calls, while DENC2 requires $2\ell + 7\ell/r$ calls. Table 1 compares DENC1 and DENC2 to well-known block-cipher based AE schemes. In the table, we include the most prominent nonce-based AE schemes for better comparison. As explained before, practical situations usually lead to many repeated nonces (for example, when random nonces are used). We see that DENC1 and DENC2 outperform all block cipher based AE schemes in the nonce-misuse setting, and it only requires $2\ell + \ell/r$ resp. $2\ell + 7\ell/r$ calls to the underlying block cipher to process a message of length ℓ.

Figure 1.1 compares the influence of the maximum message length ℓ to the threshold number of the total length of the encryption queries σ for DENC1, DENC2, and AES-GCM-SIV in the nonce-respecting setting, where we distinguish two different models in which AES-GCM-SIV has been analyzed. We see that both DENC1 and DENC2 provide stronger bounds than AES-GCM-SIV both in

[7] The associated data A may contain a nonce N, therefore N does not need to be unique as long as the couple (A, M) is not repeated.

[7] Authenticity only. CWC$^+$ does not provide privacy in the nonce-misuse setting.

Table 1. Comparison of our AE modes, DENC1 and DENC2, with other prominent block cipher-based DAE schemes. Here, n is the block size and we set $r = n$. The security is measured when the message length ℓ is a constant number of blocks.

AE	Primitive	# Calls	Nonce-respecting Security	Nonce-misusing Security	References
			Nonce-based AE schemes		
GCM	PRF	2ℓ	$O(2^{n/2})$	—	[29]
CHM	PRP	2ℓ	$O(2^{2n/3})$	—	[21]
OCB3	PRP	ℓ	$O(2^{n/2})$	—	[26]
GCM-SIV	PRF	2ℓ	$O(2^{n/2})$	$O(2^{n/2})$	[20]
CWC$^+$	PRP	2ℓ	$O(2^{3n/4}/\ell^{1/4})$	$O(2^{n/2}/\ell)$	[19]
AES-GCM-SIV	muPRP	2ℓ	$O(2^{3n/4} + 2^n/\ell^2)$	$O(2^{n/2}/\ell)$	[24]
AES-GCM-SIV	ICM	2ℓ	$O(2^n/\ell)$	$O(2^n/\ell)$	[6]
SCM	PRP	2ℓ	$O(2^n/\ell + 2^{2n/3}/\ell^{1/3})$	$O(2^{n/2}/\ell)$	[10]
XOCB	PRP	ℓ	$O(2^{2n/3}/\ell)$	—	[3]
			DAE schemes		
SIV	PRP	2ℓ	$O(2^{n/2})$	$O(2^{n/2})$	[43]
SIV-d	PRP	$d\ell$	$O(2^{\frac{d}{d+1}n})$	$O(2^{\frac{d}{d+1}n})$	[22]
SUNDAE	PRP	$2\ell + 1$	$O(2^{n/2})$	$O(2^{n/2})$	[2]
ANYDAE	PRP	2ℓ	$O(2^{n/2})$	$O(2^{n/2})$	[8]
DENC1	PRP	$2\ell + \ell/r$	$O(2^n/r\ell^{1/2})$	$O(2^n/r\ell^{1/2})$	Sec. 7
DENC2	PRP	$2\ell + 7\ell/r$	$O(2^n/r)$	$O(2^n/r)$	Sec. 7

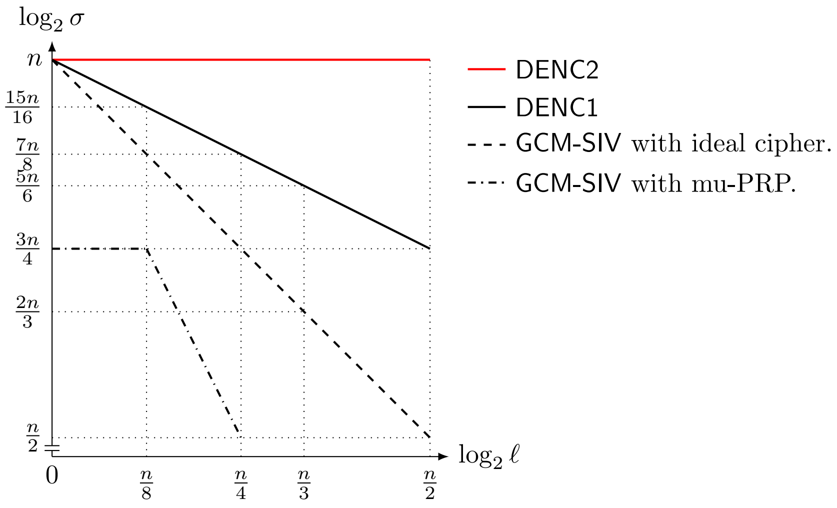

Fig. 1.1. The threshold number of the total length of the encryption queries σ as a function of ℓ, where we set $r = n$.

the standard model and the ideal cipher model for arbitrary message length ℓ. For practical applications where the messages to be encrypted have a limited length ℓ, we see that DENC1 already provides $15n/16$-bit security for $\ell = 2^{n/8}$, and $7n/8$-bit security for $\ell = 2^{n/4}$. In case long messages need to be encrypted and optimal n-bit security is required, we can choose for DENC2. Note that users can choose the rate parameter r in such a way that the best security/efficiency trade-off can be obtained according to the application and the used block cipher.

We have instantiated DENC1 and DENC2 with AES-128 and implemented them in software. The experimental results in Section 7.2 indicate that both DENC1 and DENC2 are highly competitive within the class of AE schemes (both nonce-based and deterministic). This is due to the fact that both schemes are highly parallelizable and can therefore benefit significantly from the pipelining support for AES in modern microprocessors.

2 Preliminaries

The set of non-negatives is denoted $\mathbb{N}$, and $\mathbb{N}^+$ denotes the set of positives. For any $n \in \mathbb{N}^+$, $[n]$ denotes the set $\{1, ..., n\}$ and $(n] = [n] \cup \{0\}$, respectively. The set of all n-bit strings is denoted by $\{0,1\}^n$, and $\{0,1\}^{\leq n} := \bigcup_{m=0}^{n} \{0,1\}^m$, where $\{0,1\}^0$ is the set of empty string $\perp$. We write $\langle i \rangle_n$ to denote the canonical unsigned n-bit binary representation of $i \in \mathbb{N}$. For any $x \in \{0,1\}^{\leq \infty}$, $|x|$ denotes the bit-length[8] of x. For any $x, y \in \{0,1\}^{\leq \infty}$, $x \parallel y$ denotes the concatenation of x and y. For any $x \in \{0,1\}^{\leq \infty}$ and $k \leq |x|$, $\lfloor x \rfloor_k$ (res. $\lceil x \rceil_k$) denotes the rightmost (res. leftmost) k bits of x. We define two padding schemes:

$$\mathtt{zs}_n(x) := x \parallel 0^{n-(|x| \bmod n)} \qquad \mathtt{ozs}_n(x) := \mathtt{zs}_n(x \parallel 1),$$

for any $n \in \mathbb{N}^+$ and $x \in \{0,1\}^{\leq \infty}$. We write $(x[1], \ldots, x[\ell]) \leftarrow_n x$ to denote the n-bit parsing of x, i.e., $x[1] \parallel \ldots \parallel x[\ell] = \mathtt{zs}_n(x)$ and $|x[i]| = n$ for all $i \in [\ell]$.

We often identify $\{0,1\}^n$ as the Galois field $\mathbb{F}_{2^n}$ with some implicitly fixed irreducible polynomial $p(x)$. In this context, we distinguish an arbitrary root of $p(x)$ by 2, as the primitive element of $\mathbb{F}_{2^n}$. For any $x, y \in \mathbb{F}_{2^n}$, $x \oplus y$ and $x \cdot y$ correspond to the field addition and multiplication modulo $p(x)$, respectively.

Without loss of generality, we assume[9] that any finite set is a subset of $\{0,1\}^{\leq \infty}$. For any finite sets $\mathcal{X}$ and $\mathcal{Y}$, we write $\mathcal{F}(\mathcal{X}, \mathcal{Y})$ to denote the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$. Similarly, we write $\mathcal{P}(\mathcal{X})$ to denote the set of all permutations of $\mathcal{X}$, respectively. Two finite sequences $(X_i)_{i \in \mathcal{I}}$ and $(Y_i)_{i \in \mathcal{I}}$ are said to be *bijectively-consistent* if, $X_i = X_j \iff Y_i = Y_j$, for all $i, j \in \mathcal{I}$.

For any finite sets $\mathcal{K}$, $\mathcal{X}$ and $\mathcal{Y}$, a $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function f is a family of functions $\{f_K : \mathcal{X} \to \mathcal{Y}\}_{K \in \mathcal{K}}$. One can similarly define a $(\mathcal{K}, \mathcal{X})$-keyed permutation π. For a finite set $\mathcal{X}$, we write $X \leftarrow_{\$} \mathcal{X}$ to denote the uniform at random sampling of X from $\mathcal{X}$.

[8] The number of bits in x.

[9] This is without loss of generality as any finite set $\mathcal{S}$ can be bijectively mapped to a subset of binary strings of length $\lceil \log_2 |\mathcal{S}| \rceil$.

Definition 2.1 (d-wise Independence). *A $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function f is said to be a d-wise independent hash function, if for distinct $x_1, \ldots, x_d \in \mathcal{X}$ and any $y_1, \ldots, y_d \in \mathcal{Y}$, we have $\Pr_{K \leftarrow\!\$\, \mathcal{K}}(f_K(x_1) = y_1, \ldots, f_K(x_d) = y_d) = |\mathcal{Y}|^{-d}$.*

2.1 Security Definitions

A *distinguisher* $\mathcal{A}$ is simply a binary-output oracle-algorithm, denoted $\mathcal{A}^{\mathcal{O}}$ when run with oracle $\mathcal{O}$. The *computational distance* between oracle $\mathcal{O}_1$ and oracle $\mathcal{O}_0$ with respect to a distinguisher $\mathcal{A}$ is defined as:

$$\mathbf{CD}(\mathcal{O}_1 - \mathcal{O}_0 \,|\, \mathcal{A}) \coloneqq \left| \Pr\left(\mathcal{A}^{\mathcal{O}_1} = 1\right) - \Pr\left(\mathcal{A}^{\mathcal{O}_0} = 1\right) \right|.$$

We say that $\mathcal{A}$ is a (q, ℓ, σ, τ)-distinguisher if it runs in time at most τ and makes at most q queries to its oracle, each of length at most ℓ bits and a total length of at most σ bits across all queries. For oracles operating over fixed-length inputs and/or outputs, we simplify this to a (q, τ)-distinguisher, and further, drop the time parameter for all computationally unbounded distinguishers. Without loss of generality, we assume that $\mathcal{A}$ never makes a pointless[10] query.

H-coefficients Technique: Suppose $\mathcal{A}$ is computationally unbounded and deterministic[11] distinguisher. Let Θ_1 (res. Θ_0) denote the transcript generated by $\mathcal{A}$'s interaction with $\mathcal{O}_1$ (res. $\mathcal{O}_0$). A transcript ω is said to be *attainable* if $\Pr(\Theta_0 = \omega) > 0$. The following result due to Patarin is an ubiquitous tool in information-theoretic security proofs. A proof of this result is available in multiple papers including [37,9,25].

Theorem 2.1 (H-coefficient Technique [37]). *Let Ω be the set of all attainable transcripts. For some $\epsilon_1, \epsilon_2 \geq 0$, suppose there is a set $\Omega_{\text{bad}} \subseteq \Omega$ such that:*

- $\Pr(\Theta_0 \in \Omega_{\text{bad}}) \leq \epsilon_1$*;*
- *for any $\omega \notin \Omega_{\text{bad}}$, $\Pr(\Theta_1 = \omega) \geq (1 - \epsilon_2)\Pr(\Theta_0 = \omega)$.*

Then, $\mathbf{CD}(\mathcal{O}_1 - \mathcal{O}_0 \,|\, \mathcal{A}) \leq \epsilon_1 + \epsilon_2$.

PSEUDORANDOM FUNCTION (PRF): A $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function F is said to be a $(q, \ell, \sigma, \tau, \epsilon)$-PRF if for all (q, ℓ, σ, τ)-distinguisher $\mathcal{A}$

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{F}}(\mathcal{A}) \coloneqq \mathbf{CD}(\mathsf{F}_K - \$_{\mathsf{f}} \,|\, \mathcal{A}) \leq \epsilon, \tag{1}$$

for $K \leftarrow\!\$\, \mathcal{K}$ and $\$_{\mathsf{f}} \leftarrow\!\$\, \mathcal{F}(\mathcal{X}, \mathcal{Y})$.

PSEUDORANDOM PERMUTATION (PRP): A $(\mathcal{K}, \mathcal{X})$-keyed permutation E is said to be a $(q, \ell, \sigma, \tau, \epsilon)$-PRP if for all (q, ℓ, σ, τ)-distinguisher $\mathcal{A}$

$$\mathbf{Adv}^{\mathsf{prp}}_{\mathsf{E}}(\mathcal{A}) \coloneqq \mathbf{CD}(\mathsf{E}_K - \boldsymbol{\pi} \,|\, \mathcal{A}) \leq \epsilon, \tag{2}$$

[10] A query is pointless if it is either a duplicate query or if the corresponding response is deducible from the previous queries.

[11] This is without loss of generality for computationally unbounded distinguishers.

for $K \leftarrow_{\$} \mathcal{K}$ and $\boldsymbol{\pi} \leftarrow_{\$} \mathcal{P}(\mathcal{X})$.

IV-based Encryption: A $(\mathcal{K}, \mathcal{T}, \mathcal{M})$-IV-based encryption scheme (abbreviated as IVE), denoted $\mathcal{E}$, is a tuple of $(\mathcal{K}, \mathcal{T} \times \mathcal{M}, \mathcal{M})$-keyed functions $(\mathcal{E}.\mathsf{Enc}, \mathcal{E}.\mathsf{Dec})$ that satisfies:

1. the *correctness* condition: $\mathcal{E}.\mathsf{Dec}_K(T, \mathcal{E}.\mathsf{Enc}_K(T, M)) = M$, and
2. the *length-preserving* property: $|\mathcal{E}.\mathsf{Enc}_K(T, M)| = |M|$.

for all $(K, T, M) \in \mathcal{K} \times \mathcal{T} \times \mathcal{M}$. It is customary to refer to $\mathcal{E}.\mathsf{Enc}$ as the encryption algorithm and $\mathcal{E}.\mathsf{Dec}$ as the decryption algorithm, with $T \in \mathcal{T}$ being the *initialization vector* (IV). The outputs of $\mathcal{E}.\mathsf{Enc}$ and $\mathcal{E}.\mathsf{Dec}$ are referred as ciphertext and plaintext, respectively.

$\mathcal{E}$ is said to achieve $(q, \ell, \sigma, \tau, \epsilon)$-*random-IV privacy* (Priv\$) security if for all (q, ℓ, σ, τ)-distinguisher $\mathcal{A}$ that is restricted to sample mutually independent and uniform at random IVs across queries

$$\mathbf{Adv}^{\mathsf{priv\$}}_{\mathcal{E}}(\mathcal{A}) := \mathbf{CD}(\mathcal{E}.\mathsf{Enc}_K - \$_{\mathsf{e}} \,|\, \mathcal{A}) \leq \epsilon, \tag{3}$$

for $K \leftarrow_{\$} \mathcal{K}$ and the oracle $\$_{\mathsf{e}}$ takes $(T, M) \in \mathcal{T} \times \mathcal{M}$ as input and outputs a uniform random string of length $|M|$ bits.

Convention: Throughout we fix some $n \in \mathbb{N}^+$ as the *block size*. Let $\mathcal{B} := \{0,1\}^n$ and set $\mathcal{T} := \mathcal{B}^2$. By extension, $\mathcal{B}^* := \cup_{i=1}^{\infty} \mathcal{B}^i$, denotes the set of all block-strings. We may refer to any element in $\mathcal{B}$ and $\mathcal{T}$ as a *block* and a *diblock*, respectively. In this context, the block length of any $x \in \{0,1\}^{\leq\infty}$ is defined as, $\|x\| := |\mathsf{zs}_n(x)|/n = \lceil |x|/n \rceil$. Similarly, $\|\ell\| := \lceil \ell/n \rceil$ for all $\ell \in \mathbb{N}$.

Pseudorandom Blocks Generator (PRBG) and PRBG-based IVE: Let G be a $(\mathcal{K}, \mathcal{T} \times \mathbb{N}, \mathcal{B}^*)$-keyed function satisfying $\|\mathsf{G}_K(T, m)\| = m$ for all $(K, T, m) \in \mathcal{K} \times \mathcal{T} \times \mathbb{N}$. G is said to be a $(q, \ell, \sigma, \tau, \epsilon)$-PRBG if for all (q, ℓ, σ, τ)-distinguishers $\mathcal{A}$ that is restricted to sample mutually independent and uniform at random values in $\mathcal{T}$ across all queries

$$\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{G}}(\mathcal{A}) := \mathbf{CD}(\mathsf{G}_K - \$_{\mathsf{g}} \,|\, \mathcal{A}) \leq \epsilon, \tag{4}$$

for $K \leftarrow_{\$} \mathcal{K}$ and the oracle $\$_{\mathsf{g}}$ takes $(T, m) \in \mathcal{T} \times \mathbb{N}$ as input and outputs a uniform random string of length m blocks.

One can define a natural $(\mathcal{K}, \mathcal{T} \times \mathcal{M}, \mathcal{M})$-IVE $\mathcal{E}$ based on G as follows:

$$\begin{aligned} \mathcal{E}.\mathsf{Enc}_K(T, M) &:= \lfloor \mathsf{G}_K(T, \|M\|) \rfloor_{|M|} \oplus M, \\ \mathcal{E}.\mathsf{Dec}_K(T, C) &:= \lfloor \mathsf{G}_K(T, \|C\|) \rfloor_{|C|} \oplus C, \end{aligned} \tag{5}$$

for all $K \in \mathcal{K}$, $T \in \mathcal{T}$, and $M, C \in \mathcal{M}$. Here, G is referred as the *keystream generator* of $\mathcal{E}$. Indeed, most of the existing IVE follow this approach, and all the IVE schemes in this paper also employ this technique. The following security reduction follows directly by definitions.

Proposition 2.1. *For all (q,ℓ,σ,τ)-Priv\$ distinguisher $\mathcal{A}$, there exists a $(q, n\|\ell\|, n\|\sigma\| + nq, O(\tau))$-PRBG distinguisher $\mathcal{B}$ such that*

$$\mathbf{Adv}_{\mathcal{E}}^{\mathsf{priv\$}}(\mathcal{A}) \leq \mathbf{Adv}_{\mathsf{G}}^{\mathsf{prbg}}(\mathcal{B}).$$

Thus, in context of the security of such IVEs, it is sufficient to show that the underlying keystream generator is a secure PRBG.

Deterministic Authenticated Encryption: A $(\mathcal{K},\mathcal{H},\mathcal{M},\mathcal{T})$-deterministic authenticated encryption scheme (abbreviated as DAE), denoted $\mathsf{\Pi}$, is a tuple $(\mathsf{\Pi.Enc}, \mathsf{\Pi.Dec})$ where:

- $\mathsf{\Pi.Enc}$ is a $(\mathcal{K}, \mathcal{H} \times \mathcal{M}, \mathcal{M} \times \mathcal{T})$-keyed function,
- $\mathsf{\Pi.Dec}$ is a $(\mathcal{K}, \mathcal{H} \times \mathcal{M} \times \mathcal{T}, \mathcal{M} \cup \{\perp\})$-keyed function,

that satisfies:

1. the *correctness* condition: $\mathsf{\Pi.Dec}_K(A, \mathsf{\Pi.Enc}_K(A, M)) = M$, and
2. the *fixed-stretch* property: $|\mathsf{\Pi.Enc}_K(A, M)| = |M| + 2n$,

for all $(K, A, M) \in \mathcal{K} \times \mathcal{H} \times \mathcal{M}$. It is customary to refer to $\mathsf{\Pi.Enc}$ as the encryption algorithm and $\mathsf{\Pi.Dec}$ as the decryption algorithm, with $A \in \mathcal{H}$ being the *header* or *associated data*. The outputs of $\mathsf{\Pi.Enc}$ and $\mathsf{\Pi.Dec}$ are referred to as ciphertext-tag and plaintext, respectively. While DAE does not inherently include the concept of a *nonce*, we will treat it as part of the associated data for interoperability purposes.

$\mathsf{\Pi}$ is said to achieve $(q,\ell,\sigma,\tau,\epsilon)$-*deterministic authenticated encryption* (DAE) security if for all (q,ℓ,σ,τ)-distinguisher $\mathcal{A}$:

$$\mathbf{Adv}_{\mathsf{\Pi}}^{\mathsf{dae}}(\mathcal{A}) := \mathbf{CD}((\mathsf{\Pi.Enc}_K, \mathsf{\Pi.Dec}_K) - (\$_{\mathsf{a}}, \perp) \,|\, \mathcal{A}) \leq \epsilon, \tag{6}$$

for $K \leftarrow_{\$} \mathcal{K}$, the oracle $\$_{\mathsf{a}}$ takes $(A, M) \in \mathcal{H} \times \mathcal{M}$ as input and outputs a uniform random string of length $|M| + 2n$ bits, and the oracle $\perp$ denotes the constant function $\perp : \mathcal{H} \times \mathcal{M} \times \mathcal{T} \rightarrow \{\perp\}$.

3 Towards Optimally Secure DAE

Our objective in this section is to lay the foundation towards a DAE scheme. Rogaway and Shrimpton's synthetic IV or SIV [43] is a generic technique to construct a DAE scheme given any PRF and any IVE $\mathcal{E}$. First, we describe the SIV paradigm along with the well-known SIV security reduction in Section 3.1. Next, in Section 3.2, we employ the Hash-then-modified-Benes approach due to Cogliati et al. [12] to construct an optimally secure $2n$-bit pseudorandom function F^*.

3.1 The SIV Paradigm

Definition 3.1. *Given a $(\mathcal{K}_1, \mathcal{H} \times \mathcal{M}, \mathcal{T})$-keyed function* F *and a $(\mathcal{K}_2, \mathcal{T} \times \mathcal{M}, \mathcal{M})$-IVE $\mathcal{E}$, we define a $(\mathcal{K}_1 \times \mathcal{K}_2, \mathcal{H} \times \mathcal{M}, \mathcal{T})$-DAE, called the* $\mathsf{SIV}[\mathsf{F}, \mathcal{E}]$ *construction as follows: for all $(K_1, K_2) \in \mathcal{K}_1 \times \mathcal{K}_2$, $(A, M) \in \mathcal{H} \times \mathcal{M}$, and $(C, T) \in \mathcal{M} \times \mathcal{T}$*

$$\mathsf{SIV.Enc}_{K_1,K_2}(A, M) \coloneqq (\mathcal{E}.\mathsf{Enc}_{K_2}(\mathsf{F}_{K_1}(A, M), M), \mathsf{F}_{K_1}(A, M))$$

$$\mathsf{SIV.Dec}_{K_1,K_2}(A, C, T) \coloneqq \begin{cases} \mathcal{E}.\mathsf{Dec}_{K_2}(C, T) & \textit{if } \mathsf{F}_{K_1}(A, \mathcal{E}.\mathsf{Dec}_{K_2}(C, T)) = T \\ \perp & \textit{otherwise.} \end{cases}$$

The following result is the Iwata-Minematsu adaptation [22] of a celebrated generic composition result shown in multiple papers, including [43, Theorem 2], [20, Corollary 2.3], and [36, Theorem 1].

Lemma 3.1 (SIV Security). *For any (q, ℓ, σ, τ)-DAE distinguisher $\mathcal{A}$, there exists a $(q, \ell, \sigma, O(\tau))$-PRF distinguisher $\mathcal{B}$ and a $(q, \ell, \sigma, O(\tau))$-Priv\$ distinguisher $\mathcal{C}$, such that*

$$\mathbf{Adv}^{\mathsf{dae}}_{\mathsf{SIV}[\mathsf{F},\mathcal{E}]}(\mathcal{A}) \leq \mathbf{Adv}^{\mathsf{prf}}_{\mathsf{F}}(\mathcal{B}) + \mathbf{Adv}^{\mathsf{priv\$}}_{\mathcal{E}}(\mathcal{C}) + \frac{q}{2^{2n}}.$$

The SIV security lemma effectively states that we can construct an optimally secure DAE by combining an optimally secure PRF F and an optimally secure IVE $\mathcal{E}$.

3.2 F^*: Optimally Secure Variable-Length Input PRF

In [12], Cogliati et al. proposed a generic paradigm to construct optimally secure PRFs, called Hash-then-modified Benes (or HtmB).

The high level idea is as follows: first, the variable-length input is hashed into a $2n$-bit value using a hash function with certain collision-freeness properties. This $2n$-bit hash value is then fed through the modified Benes (or mBenes) function [1] to generate the output. Although mBenes can produce a $2n$-bit output, the authors truncate the output to the first n-bits. We emphasize that this has no bearing on the security of the resulting construction, and was done solely for practical reasons. See [12, Remark 4.1] for more details.

In this paper, we will consider an HtmB instance with full $2n$-bit output. In particular, we reuse the HtmB instance $\mathsf{2k\text{-}HtmB\text{-}p2}$ from [12], albeit with the full $2n$-bit output, and call it $\mathsf{2k\text{-}HtmB\text{-}p2}^*$. We instantiate this construction with the hash layer from $\mathsf{PMAC+}$ [47], and call it PHASH^*. The resulting $(\mathcal{P}(\mathcal{B})^3, \mathcal{H} \times \mathcal{M}, \mathcal{T})$-keyed function F^* is described in Algorithm 1.

Theorem 3.1 (F^* Security). *Let $q, \ell, \sigma \in \mathbb{N}^+$ such that $n^3 \leq 2^{\frac{n}{2}-1}$, $n^2 q \leq 2^{n-4}$ and $\|\ell\| \leq 2^{\frac{n}{2}-1}$. Then, for any (q, ℓ, σ)-PRF distinguisher $\mathcal{A}$ we have*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{F}^*}(\mathcal{A}) \leq \frac{134(\|\sigma\| + q)}{2^n} + \frac{392(\|\sigma\| + q)^2}{2^{2n}} + \frac{128q^2}{2^{3n}}.$$

Algorithm 1 The F^* function.

```
function $\mathsf{F}^*_{\pi_1,\pi_2,\pi_3}(A, M)$
=== $(U, V) \leftarrow \mathsf{PHASH}^*_{\pi_1}(A, M)$ ===
    $(D[1], \ldots, D[\ell]) \leftarrow \mathtt{ozs}_n(A) \parallel \mathtt{ozs}_n(M)$
    $U \leftarrow 0^n$
    $V \leftarrow 0^n$
    for $i = 1 \ldots \ell$ do
        $\Delta_i \leftarrow 2^i \cdot \pi_1(0^n) \oplus 2^{2i} \cdot \pi_1(10^{n-1})$
        $U \leftarrow U \oplus \pi_1(D[i] \oplus \Delta_i)$
        $V \leftarrow 2 \cdot V \oplus \pi_1(D_i \oplus \Delta_i)$
    $U \leftarrow 0 \parallel \lfloor U \rfloor_{n-1}$
    $V \leftarrow 1 \parallel \lfloor V \rfloor_{n-1}$
======
    $X \leftarrow \pi_2(U) \oplus V$
    $Y \leftarrow \pi_2(V) \oplus U$
    $T[1] \leftarrow \pi_3(00 \parallel \lfloor X \rfloor_{n-2}) \oplus \pi_3(01 \parallel \lfloor Y \rfloor_{n-2})$
    $T[2] \leftarrow \pi_3(10 \parallel \lfloor X \rfloor_{n-2}) \oplus \pi_3(11 \parallel \lfloor Y \rfloor_{n-2})$
    $T \leftarrow T[1] \parallel T[2]$
    return $T$
```

A formal proof of this result follows from the proof of [12, Theorem 7.3], where the underlying hash function is instantiated with PHASH^*. The analysis of this hash function is similar to the analysis presented in [12, Section 6.2]. See Appendix A for further details.

With F^*, we have been able to exploit the HtmB paradigm to construct an optimally secure variable-length input PRF. Now, all that is required is to instantiate SIV with an optimally secure IVE that works with $2n$-bit IV values. Unfortunately, barring SIV-r construction [22], there are not many options in this direction. While SIV-r can theoretically achieve close to optimal security, it does so at a very high cost in terms of efficiency.[12] In particular, to achieve security up to $2^{\frac{rn}{r+1}}$ queries, it requires r independent permutations and makes $r\ell$ calls to process ℓ-block messages. Clearly, this becomes practically infeasible even for a modest value of r, say 4.

We devote the rest of this paper to gradually build towards an efficient solution for this problem, and as a side-effect, obtain an optimally secure DAE.

4 IV0: Optimally Secure IVE for Short Inputs

In our quest for an optimally secure IVE, we first propose the $\mathsf{IV0}$ paradigm for handling short-length inputs. Fix:

- a *rate parameter* r, representing the maximum number of input blocks,

[12] Here, efficiency is measured in terms of the rate and key size.

- a $(\mathcal{T}, \mathbb{N}, \mathcal{B})$-keyed function H,

where H serves as an implicit parameter for $\mathsf{IV0}$. At its core, $\mathsf{IV0}$ utilizes a $(\mathcal{P}(\mathcal{B}), \mathcal{T} \times \mathbb{N} \times [r], \mathcal{B}^{\le r})$-keyed function, called Star, as the underlying keystream generator. Algorithm 2 provides a complete description of this function.

Given Star, we now define $\mathsf{IV0}$ as a $(\mathcal{P}(\mathcal{B}), \mathcal{T}, \mathcal{B}^{\le r})$-IVE as follows: for all $\pi \in \mathcal{P}(\mathcal{B})$, $T \in \mathcal{T}$ and $M, C \in \mathcal{B}^{\le r}$:

$$\mathsf{IV0.Enc}_\pi(T, M) \coloneqq \mathsf{Star}_\pi(T, 1, \|M\|) \oplus M \tag{7}$$

$$\mathsf{IV0.Dec}_\pi(T, C) \coloneqq \mathsf{Star}_\pi(T, 1, \|C\|) \oplus C \tag{8}$$

One might question whether the second argument in the definition of Star serves any purpose. Although it is fixed in $\mathsf{IV0}$, we include it in the argumentation to ensure notational simplicity in latter constructions. Star makes $r'+1$ permutation calls to process r'-block input, where $r' \le r$. Thus, the *rate* for $\mathsf{IV0}$ is $r'/(r'+1)$ for any r'-block input. Theorem 4.1 shows that $\mathsf{IV0}$ achieves near-optimal security for small r.

Algorithm 2 The Star function.

```
function $\mathsf{Star}_\pi(T,\, j,\, r')$
    $\widehat{j} \leftarrow (j-1)(r+1)$
    $X_j[0] \leftarrow \mathsf{H}_T(\widehat{j})$
    $Y_j[0] \leftarrow \pi(X_j[0])$
    for $k = 1 \ldots r'$ do
        $X_j[k] \leftarrow \mathsf{H}_T(\widehat{j} + k)$
        $Y_j[k] \leftarrow \pi(X_j[k])$
        $Z_j[k] \leftarrow Y_j[0] \oplus Y_j[k]$
    $Z_j \leftarrow Z_j[1] \,\|\, \ldots \,\|\, Z_j[r']$
    return $Z_j$
```

Theorem 4.1 ($\mathsf{IV0}$ Security). *Fix some $n, r, q, \sigma \in \mathbb{N}^+$ such that $n^3 r^2 \le 2^{\frac{n}{2}-5}$, $\sigma \le qrn$ and $n^2 r^2(\|\sigma\| + q) \le 2^n/48$. Suppose H is a 2-wise independent hash function. Then, for all (q, rn, σ)-Priv\$ distinguisher $\mathcal{A}$, we have*

$$\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV0}}(\mathcal{A}) \le \frac{17r(\|\sigma\| + q + r)}{2^n}.$$

Proof Overview: A detailed proof of this theorem is postponed to Section 9. Here, we briefly outline the approach. For simplicity in the discussion, we assume that all the queries are r-block long, i.e., $\sigma = nrq$. First, it is sufficient to upper bound the statistical distance between Z and $U \leftarrow\!\!\$\; \mathcal{B}^{rq}$, where Z is the output across all queries. To bound this, we need a strong lower bound on $\Pr(Z = z)$ for most $z \in \mathcal{B}^{rq}$, which leads to analyzing the system of equations:

$$S \coloneqq \{Y_1^i[0] \oplus Y_1^i[k] = z_1^i[k] \,:\, i \in [q],\, k \in [r]\}$$

A valid solution to these q bivariate equations must satisfy the so-called *bijectively-consistent* condition (see Section 2). We apply Patarin's mirror theory [11] to lower bound the number of solutions to S satisfying this condition.

Applying mirror theory typically requires specific restrictions on S. To express and analyse these, we use a graph-theoretic approach, associating a *dependency graph* (see Section 8 for details) with (X, Z) and reformulating the aforementioned combinatorial problem as an enumeration of certain vertex-labelings of this graph. We show that if the graph is acyclic and composed of *small* components, mirror theory gives the desired lower bound on valid labelings, leading to the security bound. Finally, we prove that the said dependency graph for random (X, Z) meets these conditions with overwhelming probability.

4.1 A 2-wise Independent Hash Function

Theorem 4.1 specifies that $\mathsf{IV0}$ must be instantiated with a 2-wise independent hash function. While one can use any one of the off-the-shelf algebraic hash functions [44,45], they usually involve multiplications by arbitrary field elements in $\mathbb{F}_{2^n}$, which can be computationally expensive for large n. Instead, we propose the following 2-wise independent hash function.

Definition 4.1 (Galois-Wegman-Carter Hash). *Define the* $(\mathbb{F}_{2^n}^2, \mathbb{N}, \mathbb{F}_{2^n})$*-hash function* gwc *as: for all* $K = (K_0, K_1) \in \mathbb{F}_{2^n}^2$ *and* $i \in \mathbb{N}$,

$$\mathsf{gwc}_K(i) \coloneqq K_0 \oplus 2^{(i \bmod 2^n)+1} \cdot K_1.$$

Since the rank of

$$\begin{bmatrix} 1 & 2^{(i \bmod 2^n)+1} \\ 1 & 2^{(j \bmod 2^n)+1} \end{bmatrix}$$

is 2 for any $0 \le i \ne j \le 2^{n-1} - 1$, ensuring that the equations $\mathsf{gwc}_K(i) = x$ and $\mathsf{gwc}_K(j) = y$ are linearly independent, yielding a unique solution given any x and y. Moreover, for random $K \leftarrow_{\$} \mathbb{F}_{2^n}^2$, each of these solutions hold with $1/2^{2n}$ probability, making gwc a 2-wise independent hash function. We use gwc whenever a 2-wise independent hash function is required. Note that we do not claim any novelty with respect to this hash function; similar constructions have been implicitly used in prior works [47,31,7].

5 IV1: Extending IV0 to Variable-Length Inputs

As discussed in the preceding section, $\mathsf{IV0}$ is an efficient and highly secure IVE, provided the input size remains small, typically up to a small multiple of n. For longer, arbitrarily sized inputs, this straightforward approach encounters a technical limitation: the $r = O(\sqrt{2^n/n^2q})$ restriction in Theorem 4.1. However, it is possible to extend $\mathsf{IV0}$ to accommodate general scenarios. We refer to this updated construction as $\mathsf{IV1}$. At a high level, this can be achieved by dividing the input into *chunks* of size at most r blocks and applying $\mathsf{IV0}$ to each chunk individually. In this context, for any $x \in \{0,1\}^*$, $\|x\|_r \coloneqq \lceil \|x\|/r \rceil$ is referred as the *chunk-length* of x. Similarly, $\|\ell\|_r \coloneqq \lceil \|\ell\|/r \rceil$ for all $\ell \in \mathbb{N}$.

We define an efficient extension of Star (see Algorithm 2), called $\mathsf{GiantStar}$, which is a $(\mathcal{P}(\mathcal{B}), \mathcal{T} \times \mathbb{N}, \mathcal{B}^*)$-keyed function and acts as the underlying keystream generator in $\mathsf{IV1}$. Algorithm 3 provides a complete description of this function. Given $\mathsf{GiantStar}$, we now define $\mathsf{IV1}$ as a $(\mathcal{P}(\mathcal{B}), \mathcal{T}, \{0,1\}^*)$-IVE as follows: for all $\pi \in \mathcal{P}(\mathcal{B})$, $T \in \mathcal{T}$ and $M, C \in \{0,1\}^*$:

$$\mathsf{IV1.Enc}_\pi(T, M) \coloneqq \lceil \mathsf{GiantStar}_\pi(T, \|M\|) \rceil_{|M|} \oplus M \tag{9}$$

$$\mathsf{IV1.Dec}_\pi(T, C) \coloneqq \lceil \mathsf{GiantStar}_\pi(T, \|C\|) \rceil_{|C|} \oplus C \tag{10}$$

To produce an m-block output, $\mathsf{GiantStar}$ requires exactly $m + \lceil m/r \rceil$ permutation calls. Thus, the *rate* for $\mathsf{IV1}$ is approx. $r/(r+1)$. Theorem 5.1 below states the security bound for $\mathsf{IV1}$.

Algorithm 3 The $\mathsf{GiantStar}$ function.

```
function GiantStar$_\pi(T, \|\ell\|)$
    if $\|\ell\| = 0$ then
        return $\perp$
    $r' \leftarrow \|\ell\| \bmod r$
    for $j = 1 \ldots \|\ell\|_r - 1$ do
        $Z_j \leftarrow \mathsf{Star}_\pi(T, j, r)$
    if $r' = 0$ then
        $r' \leftarrow r$
    $Z_{\|\ell\|_r} \leftarrow \mathsf{Star}_\pi(T, \|\ell\|_r, r')$
    $Z \leftarrow Z_1 \| \ldots \| Z_{\|\ell\|_r}$
    return $Z$
```

Theorem 5.1 (IV1 Security). *Fix some $n, r, q, \ell, \sigma \in \mathbb{N}^+$ such that $n^3 r^2 \leq 2^{\frac{n}{2}-5}$, $\sigma \leq q\ell$ and $n^2 r^2(\|\sigma\| + q) \leq 2^n/48$. Suppose H is instantiated with the gwc hash function. Then, for all (q, ℓ, σ)-Priv\$ distinguisher $\mathcal{A}$, we have*

$$\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV1}}(\mathcal{A}) \leq \frac{16r^2\|\ell\|(\|\sigma\| + q + r)^2}{2^{2n}} + \frac{18r(\|\sigma\| + q + r)}{2^n}.$$

Proof Overview: A detailed proof of this theorem is provided in Section 10. The proof is inherently similar to the proof of Theorem 4.1, except for an additional bad event analysis and some notational extensions. Again, we have to analyze system of equations of the form:

$$S \coloneqq \{Y^i_j[0] \oplus Y^i_j[k] = z^i_j[k] \,:\, i \in [q],\, j \in [\|\ell_i\|_r],\, k \in [r]\}$$

where ℓ_i denotes the length of the i-th query. As in the case of Theorem 4.1, a valid solution to this system must also satisfy the bijectively-consistent condition. We again apply mirror theory to provide a lower bound on the number of valid solutions to S, leveraging our graph-theoretic formulation as before.

A key difference between this proof and that of Theorem 4.1 lies in the probability bound of a large component in the random dependency graph. In this case, due to the limited randomness and the possibility of multiple chunks per query, an additional term of $O(r^2\|\ell\|(\|\sigma\| + q + r)^2/2^{2n})$ appears in the security bound.

Remark 5.1. Note the subtle change in the theorem statement of $\mathsf{IV1}$ compared to $\mathsf{IV0}$: the hash function H is specifically instantiated with gwc. This choice is necessary to obtain a bound of $O(r^2\|\ell\|(\|\sigma\|+q+r)^2/2^{2n})$. While other instantiations may yield similar bounds, our proof approach would only guarantee a bound of $O(\|\ell\|(\|\sigma\|+q)/2^n)$ for a general 2-wise independent hash function, imposing a stricter limit on ℓ under the same query limit.

Remark 5.2. In [30], Minematsu and Tsunoo proposed a *weak PRF*[13] construction called Extended PRT (ERT), an extension of the Damgård-Neilsen weak PRF construction PRT [17]. Their construction uses a tree-based structure built on a weak PRF primitive. While ERT improves upon PRT by reducing the key size by about 63%, it still requires a key size proportional to the tree depth. In contrast, $\mathsf{IV1}$ requires only a single key and is inherently parallel.

6 IV2: Optimally Secure IVE for Variable-Length Inputs

While $\mathsf{IV1}$ handles arbitrary-length inputs, its security degrades linearly with input length, which is undesirable for applications with long inputs. To address this, we introduce $\mathsf{IV2}$, an IVE that achieves optimal security independent of the maximum message length, providing an ℓ-free security bound. Essentially, it is a variant of $\mathsf{IV1}$, where each chunk is processed with an independent and uniform IV. This makes it equivalent to $\mathsf{IV0}$ but with at most $\|\sigma\|_r+2q$ invocations of Star, instead of q as in $\mathsf{IV0}$.

Let G denote a $(\mathcal{K},\mathcal{T}\times\mathbb{N},\mathcal{B}^*)$-keyed function. Given G, we first define a $(\mathcal{K}\times\mathcal{P}(\mathcal{B}),\mathcal{T}\times\mathbb{N},\mathcal{B}^*)$-keyed function, called $\mathsf{Snowflake}$, which acts as the underlying keystream generator in $\mathsf{IV2}$. Algorithm 4 provides a complete description of this function. Given $\mathsf{Snowflake}$, we now define $\mathsf{IV2}$ as a $(\mathcal{K}\times\mathcal{P}(\mathcal{B}),\mathcal{T},\{0,1\}^*)$-IVE as follows: for all $K\in\mathcal{K}$, $\pi\in\mathcal{P}(\mathcal{B})$, $T\in\mathcal{T}$ and $M,C\in\{0,1\}^*$:

$$\mathsf{IV2.Enc}_{K,\pi}(T,M) := \lceil \mathsf{Snowflake}_{K,\pi}(T,\|M\|)\rceil_{|M|}\oplus M \tag{11}$$

$$\mathsf{IV2.Dec}_{K,\pi}(T,C) := \lceil \mathsf{Snowflake}_{K,\pi}(T,\|C\|)\rceil_{|C|}\oplus C \tag{12}$$

To produce an m-block output, $\mathsf{Snowflake}$ requires exactly $m+\lceil m/r\rceil$ permutation calls and $\lceil m/r\rceil$ calls to G. Thus, the *rate* of $\mathsf{IV2}$ is contingent upon the efficiency of G and the relative amortization of G calls for long messages. Theorem 5.1 below states the security bound for $\mathsf{IV2}$.

Theorem 6.1 ($\mathsf{IV2}$ Security). *Fix some $n,r,q,\ell,\sigma\in\mathbb{N}^+$ such that $n^3r^2\le 2^{\frac{n}{2}-5}$, $\sigma\le q\ell$ and $n^2r^2(\|\sigma\|+q)\le 2^n/48$. Suppose H is a 2-wise independent hash function. Then, for all (q,ℓ,σ)-Priv\$ distinguisher $\mathcal{A}$, there exists a $(q,2\ell+2n,2\sigma+6nq)$-PRBG distinguisher $\mathcal{B}$ and a $(\|\sigma\|_r+2q,rn,\sigma+4rnq)$-Priv\$ distinguisher $\mathcal{C}$ such that*

$$\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV2}}(\mathcal{A})\le\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{G}}(\mathcal{B})+\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV0}}(\mathcal{C})+\frac{4(\|\sigma\|+3q)^2}{2^{2n}}.$$

[13] In context of this paper, a weak PRF is equivalent to the notion of PRBG defined in Section 2.1.

Algorithm 4 The $\mathsf{Snowflake}$ function.

```
function $\mathsf{Snowflake}_{K,\pi}(T,\ \|\ell\|)$
    if $\|\ell\| = 0$ then
        return $\bot$
    $r' \leftarrow \|\ell\| \bmod r$
    $W_1 \| \dots \| W_{\|\ell\|_r} \leftarrow_{2n} \mathsf{G}_K(T, 2\|\ell\|_r)$
    for $j = 1 \dots \|\ell\|_r - 1$ do
        $Z_j \leftarrow \mathsf{Star}_\pi(W_j, 0, r)$
    if $r' = 0$ then
        $r' \leftarrow r$
    $Z_{\|\ell\|_r} \leftarrow \mathsf{Star}_\pi(W_{\|\ell\|_r}, 0, r')$
    $Z \leftarrow Z_1 \| \dots \| Z_{\|\ell\|_r}$
    return $Z$
```

Proof Overview: From the security bound, the security of $\mathsf{IV2}$ reduces to $\mathsf{IV0}$ if G is replaced by $\$_{\mathsf{g}}$, at the cost of $\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{G}}(\mathcal{B})$. With no IV (i.e. T) collisions, each chunk has independent, uniform IVs. A cursory glance at Algorithm 4 then shows that one can construct $\mathcal{C}$ that perfectly simulates the oracle access to this modified $\mathsf{IV2}$ using oracle access to $\mathsf{IV0}$, as long as there are no collisions among the IV inputs of $\mathsf{IV0}$. Note that the additional term of $O(\|\sigma\|+3q)^2/2^{2n})$ accounts for the IV collisions: $O(q^2/2^{2n})$ for an IV collision in G, and $O((\|\sigma\|+2q)^2/2^{2n})$ for an IV collision in Star. See Appendix B for a detailed proof.

6.1 Instantiating the G Function

Theorem 6.1 dictates that we instantiate G with a secure PRBG. Recall that G is invoked at most $\|\ell\|_r$ times for any ℓ-bit input. So, the effect of a slightly heavier G will be somewhat milder on the rate of $\mathsf{IV2}$. In Algorithm 5 we propose a $(\mathcal{P}(\mathcal{B})^2, \mathcal{T} \times \mathbb{N}, \mathcal{B}^*)$-keyed function, called G^*, which is based on the HtmB paradigm [12]. It generates diblocks of output in parallel (see also Fig. 6.1), where each diblock index will correspond to a chunk index in $\mathsf{Snowflake}$. For the j'-th diblock, it generates a $2n$-bit string $(U_{j'}[1], U_{j'}[2])$ as follows:

$$U_{j'}[1] = 0 \,\|\, (\lfloor T[1] \rfloor_{n-1} \oplus \langle j' \rangle_{n-1}), \tag{13}$$

$$U_{j'}[2] = 1 \,\|\, (\lfloor T[2] \rfloor_{n-1} \oplus \langle j' \rangle_{n-1}), \tag{14}$$

where $T = (T[1], T[2])$ denotes the $2n$-bit IV input of G^*. This initial $2n$-bit is then fed to the modified Benes [1,41] function that generates the $2n$-bit output keystream $W_{j'}$. In the following lemma, we show that G^* is a secure PRBG.

Lemma 6.1 (G^* Security). *Fix some $n, r, q, \ell, \sigma \in \mathbb{N}^+$ such that $n^3 \leq 2^{\frac{n}{2}-5}$, $\sigma \leq q\ell$ and $n^2(\|\sigma\| + q) \leq 2^n/48$. Then, for all (q, ℓ, σ)-PRBG distinguisher $\mathcal{A}$ we have*

$$\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{G}^*}(\mathcal{A}) \leq \frac{11(\|\sigma\| + q)}{2^n} + \frac{49(\|\sigma\| + q)^2}{2^{2n}}.$$

Algorithm 5 The G^* function.

```
function $\mathsf{G}^*_{\pi_1,\pi_2}(T, \ell)$
    if $\ell = 0$ then
        return $\perp$
    $m \leftarrow \lceil \ell/2 \rceil$
    for $j' = 1 \ldots m$ do
        $j \leftarrow 2j' - 1$
        $U_{j'}[1] \leftarrow 0 \parallel (\lfloor T[1] \rfloor_{n-1} \oplus \langle j' \rangle_{n-1})$
        $U_{j'}[2] \leftarrow 1 \parallel (\lfloor T[2] \rfloor_{n-1} \oplus \langle j' \rangle_{n-1})$
        $V_{j'}[1] \leftarrow \pi_1(U_{j'}[1]) \oplus U_{j'}[2]$
        $V_{j'}[2] \leftarrow \pi_1(U_{j'}[2]) \oplus U_{j'}[1]$
        $X_j[0] \leftarrow 00 \parallel \lfloor V_{j'}[1] \rfloor_{n-2}$
        $X_j[1] \leftarrow 01 \parallel \lfloor V_{j'}[2] \rfloor_{n-2}$
        $X_{j+1}[0] \leftarrow 10 \parallel \lfloor V_{j'}[1] \rfloor_{n-2}$
        $X_{j+1}[1] \leftarrow 11 \parallel \lfloor V_{j'}[2] \rfloor_{n-2}$
        $W_j[1] \leftarrow \pi_2(X_j[0]) \oplus \pi_2(X_j[1])$
        $W_{j+1}[1] \leftarrow \pi_2(X_{j+1}[0]) \oplus \pi_2(X_{j+1}[1])$
        $W_{j'} \leftarrow W_j[1] \parallel W_{j+1}[1]$
    $W \leftarrow W_1 \parallel \ldots \parallel W_m$
    return $\lceil W \rceil_{n\ell}$
```

Proof Overview: The proof ideas for this lemma are similar to those of the PRBG constructions discussed. We need to show that the mirror theory conditions hold for the system of equations:

$$\{\pi_2(X_{j'}[0]) \oplus \pi_2(X_{j'}[1]) = W_{j'}[1] \;:\; j \in [\ell],\, b \in \{0,1\}\}.$$

This is achieved using techniques from [12], with mirror theory completing the proof. See Appendix F for further details.

We instantiate $\mathsf{IV2}$ with G^*, and refer the resulting instance as $\mathsf{IV2}^*$. Using Theorem 6.1 and 4.1, and Lemma 6.1, we have the following corollary:

Corollary 6.1 ($\mathsf{IV2}^*$ Security). *Fix some $n, r, q, \ell, \sigma \in \mathbb{N}^+$ such that $n^3r^2 \leq 2^{\frac{n}{2}-5}$, $\sigma \leq q\ell$ and $n^2r^2(\|\sigma\| + q) \leq 2^n/48$. Suppose H is a 2-wise independent hash function. Then, for all (q, ℓ, σ)-Priv\$ distinguisher $\mathcal{A}$ we have*

$$\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV2}^*}(\mathcal{A}) \leq \frac{28r(\|\sigma\| + 2q + r)}{2^n} + \frac{54(\|\sigma\| + 4q)^2}{2^{2n}}.$$

Note that, $\mathsf{IV2}^*$ achieves an ℓ-independent n-bit security bound, whereas $\mathsf{IV1}$ contains an ℓ factor in its security bound. However, $\mathsf{IV1}$ requires $r + 1$ block cipher calls per chunk, where the chunk size is r, leading to its rate $r/(r + 1)$, whereas $\mathsf{IV2}^*$ takes a few extra permutation calls per chunk (i.e. $r+7$ permutation calls, leading to a rate of $r/(r + 7)$) compared to the $\mathsf{IV1}$ construction.

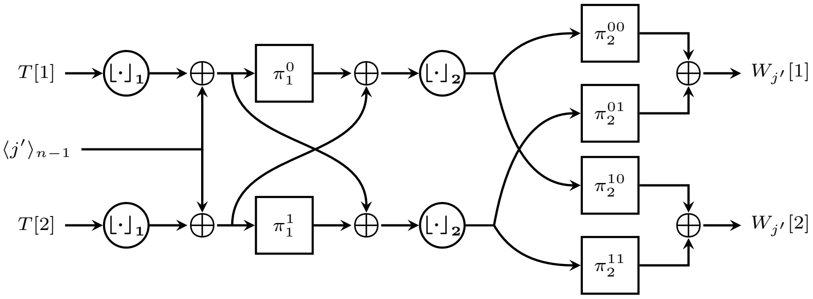

Fig. 6.1. The j'-th diblock of the keystream generated by G^* with IV T, where $\lfloor\cdot\rfloor_d$ denotes the chopped version of input to the rightmost $n-d$ bits, and π_k^d denotes the injective function $\pi_k(d \parallel \cdot)$.

7 The DENC Family of DAE

We propose two SIV-based DAE constructions, namely DENC1 and DENC2 using F^*, IV1 and IV2*. Formally, we define:

$$\mathsf{DENC1} := \mathsf{SIV}\left[\mathsf{F}^*, \mathsf{IV1}\right] \qquad \text{and} \qquad \mathsf{DENC2} := \mathsf{SIV}\left[\mathsf{F}^*, \mathsf{IV2}^*\right], \tag{15}$$

where the underlying hash function H in both IV1 and IV2* is instantiated with the gwc hash. Thus, DENC1 is a $(\mathcal{P}(\mathcal{B})^4, \mathcal{H} \times \mathcal{M}, \mathcal{T})$-DAE and DENC2 is a $(\mathcal{P}(\mathcal{B})^6, \mathcal{H} \times \mathcal{M}, \mathcal{T})$-DAE.

7.1 DAE Security of the DENC Family

Using Lemma 3.1, Theorem 3.1, Theorem 5.1 and Corollary 6.1 we immediately get the following corollaries:

Corollary 7.1 (Security Theorem of DENC1). *Fix some $n, r, q, \ell, \sigma \in \mathbb{N}^+$ such that $n^3r^2 \le 2^{\frac{n}{2}-5}$, $\sigma \le q\ell$, $\|\ell\| \le 2^{\frac{n}{2}-1}$, and $n^2r^2(\|\sigma\| + q) \le 2^n/48$. Then, for all (q, ℓ, σ)-DAE distinguisher $\mathcal{A}$, we have*

$$\mathbf{Adv}^{\mathsf{dae}}_{\mathsf{DENC1}}(\mathcal{A}) \le \frac{152r(\|\sigma\| + q + r)}{2^n} + \frac{409r^2\|\ell\|(\|\sigma\| + q + r)^2}{2^{2n}} + \frac{128q^2}{2^{3n}}.$$

Corollary 7.2 (Security Theorem of DENC2). *Fix some $n, r, q, \ell, \sigma \in \mathbb{N}^+$ such that $n^3r^2 \le 2^{\frac{n}{2}-5}$, $\sigma \le q\ell$, $\|\ell\| \le 2^{\frac{n}{2}-1}$, and $n^2r^2(\|\sigma\| + q) \le 2^n/48$. Then, for all (q, ℓ, σ)-DAE distinguisher $\mathcal{A}$, we have*

$$\mathbf{Adv}^{\mathsf{dae}}_{\mathsf{DENC2}}(\mathcal{A}) \le \frac{162r(\|\sigma\| + 2q + r)}{2^n} + \frac{446(\|\sigma\| + 4q)^2}{2^{2n}} + \frac{128q^2}{2^{3n}}$$

7.2 On Practical Instantiations of DENC1 and DENC2.

In practice the independent instance of random permutations in DENC1 and DENC2 can be easily replaced with independently keyed instances of any efficient keyed permutation, most notably AES. In terms of security, this replacement costs at most a small[14] constant multiple of the PRP advantage against the chosen keyed permutation.

It is worth mentioning that both constructions offer nearly n-bit security, although the security bound of DENC1 has an additional factor of "ℓ", while the security bound of DENC2 is ℓ-free. However, the enhancement in the security bound comes at the expense of a slight increase in the number of block cipher calls and two additional block cipher keys. Depending on the application, if the message sizes are small (e.g., $\ell = O(\mathsf{poly}(n))$), it is advisable to use DENC1. On the other hand, if $\ell = O(2^{n/4})$, using DENC1 provides only $7n/8$-bit security, whereas DENC2 always offers nearly n-bit security. We consider the chunk size r to be roughly $O(n)$, the block size of the underlying block ciphers of the constructions.

We implemented[15] DENC1 and DENC2 for chunk size of $r = 64$ with AES-128 as the underlying block cipher. Table 2 compares the software performance of DENC1 and DENC2 with existing authenticated encryption (AE) schemes. We focus primarily on schemes that are either deterministic or offer nonce-misuse resistance. For this comparison, we reuse performance data from [10]. This is a fair comparison, as DENC1 and DENC2 were benchmarked under identical conditions: our measurements were conducted on an Intel Skylake processor (i7-6700 CPU @ 4.20 GHz) with compiler optimization level -O2. In [10], authors note that the performance of ZAE is estimated based on the speed of Deoxys-BC-256 in counter mode. The results clearly demonstrate that both DENC1 and DENC2 are highly competitive within the category of DAE schemes, including AES-GCM-SIV. This is not surprising, as both schemes are highly parallelizable and significantly benefit from the pipelining support for AES in modern microprocessors. Specifically, on the Skylake processor, it is possible to pipeline four AES calls concurrently, which is especially advantageous in DENC2, where it helps amortize the overhead of the additional AES calls per chunk.

Table 2. Software performance comparison of our modes DENC1 and DENC2 with several prominent deterministic and/or nonce-misuse resistant AE schemes. The performance figures presented are throughputs, in units of cycles-per-byte (cpb).

AE	Primitive	Length			Reference
		1kB	4kB	64kB	
ChaCha20-Poly1305	–	2.17	1.55	1.47	[10]
AES-GCM	AES-128	1.23	0.63	0.56	[10]
AES-GCM-SIV	AES-128	1.57	0.89	0.81	[10]
Deoxys-I	Deoxys-BC-256	1.38	0.91	0.77	[10]
Deoxys-II	Deoxys-BC-256	2.19	1.68	1.52	[10]
ZAE	Deoxys-BC-256	≥ 1.94	≥ 1.41	≥ 1.25	[10]
SCM	AES-128	0.94	0.86	0.83	[10]
DENC1	AES-128	0.96	0.89	0.84	This work
DENC2	AES-128	1.04	0.95	0.91	This work

[14] For DENC1 it is 4 and for DENC2 it is 6.

[15] A reference implementation in C is included as accompanying code.

8 Dependency Graphs and Vertex-Labelings

The security proofs for IV0 and IV1 employ a common graph-based analysis. We abstract out this graph structure and present some important results on it. The proofs of these results are postponed to Appendix C.

For some fixed $q, c, r \geq 1$, let $(c_1, \ldots, c_q)$ be a sequence over $[c]$ indexed by $[q]$, and $(r_{1,1}, \ldots, r_{q,c_q})$ be a sequence over $[r]$ indexed by $\{(i,j) \in [q] \times [c_i]\}$. Let $\mathcal{V} = \{(i,j,k) \in [q] \times [c_i] \times (r_{i,j}]\}$ and $\mathcal{V}_{|0} = \mathcal{V} \setminus \{(i,j,0) : (i,j) \in [q] \times [c_i]\}$.

Definition 8.1 (Dependency Graph). *To any sequences X and Z over $\mathcal{B}$ and $\mathcal{B} \setminus \{0^n\}$, respectively, indexed by $\mathcal{V}$ and $\mathcal{V}_{|0}$, we associate an edge-labeled bichromatic graph, denoted $\mathcal{G}$, with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$ consisting of two types of edges:*

- *for all $(i,j,k) \in \mathcal{V}_{|0}$, $(i,j,0)$ and (i,j,k) are connected by a blue solid edge labeled $Z^i_j[k]$, denoted $(i,j,0) \xrightarrow{Z^i_j[k]} (i,j,k)$.*
- *$(i,j,k) \neq (i',j',k') \in \mathcal{V}$ are connected by a red dotted edge labeled 0, denoted $(i,j,0) \xrightarrow{0} (i',j',k')$, if $X^i_j[k] = X^{i'}_{j'}[k']$.*

Whenever convenient, we will drop X and Z to lighten the notation. Fig. 8.1 illustrates a dependency graph. Let $\mathcal{E}_b$ and $\mathcal{E}_r$ denote the set of blue and red edges, respectively. We often write $\lambda(e)$ to denote the label of edge e, and similarly write $\chi(e)$ to denote the color of e. One can also view $\lambda : \mathcal{E} \to Z \cup \{0\}$ and $\chi : \mathcal{E} \to \{\text{blue}, \text{red}\}$ as the labeling and coloring functions, respectively.

By extension, to any trail $p = (e_1, \ldots, e_k)$, we associate the label $\lambda(p) \coloneqq \lambda(e_1) \oplus \cdots \oplus \lambda(e_k)$. In a similar fashion, the trail p is said to be (colored) blue (res. red) if and only if $\chi(e_i) = \text{blue}$ (res. $\chi(e_i) = \text{red}$) for all $i \in [k]$, and otherwise the trail is said to be bichromatic. The following proposition characterizes two simple properties of the dependency graph. A proof of this result follows directly from the definition of dependency graphs.

Proposition 8.1. *In any dependency graph $\mathcal{G}$,*

1. *All blue paths must contain at most two edges.*
2. *All cycles must either be red or bichromatic.*

Definition 8.2 (Star). *For $(i,j) \in [q] \times [c_i]$, the (i,j)-th star of $\mathcal{G}$, denoted $\mathcal{S}^i_j$, is the subgraph induced by the edge set $\{\{(i,j,0),(i,j,k)\} : k \in (r_{i,j}]\} \subseteq \mathcal{E}_b$.*

In other words, a star is a tree of distance two and contains only blue edges. Note that, $\mathcal{V}(\mathcal{S}^i_j) \cap \mathcal{V}(\mathcal{S}^{i'}_{j'}) = \emptyset$ for $(i,j) \neq (i',j')$, and $\mathcal{V} = \bigsqcup_{(i,j) \in [q] \times [c_i]} \mathcal{V}(\mathcal{S}^i_j)$. Thus, the stars partition the vertex set of $\mathcal{G}$, and edges, if any, between two distinct stars must always be red.

Definition 8.3 (Maximally blue). *A maximally blue subgraph $\mathcal{T}$ of $\mathcal{G}$ is a connected subgraph to which no more blue edges can be added without disconnecting it.*

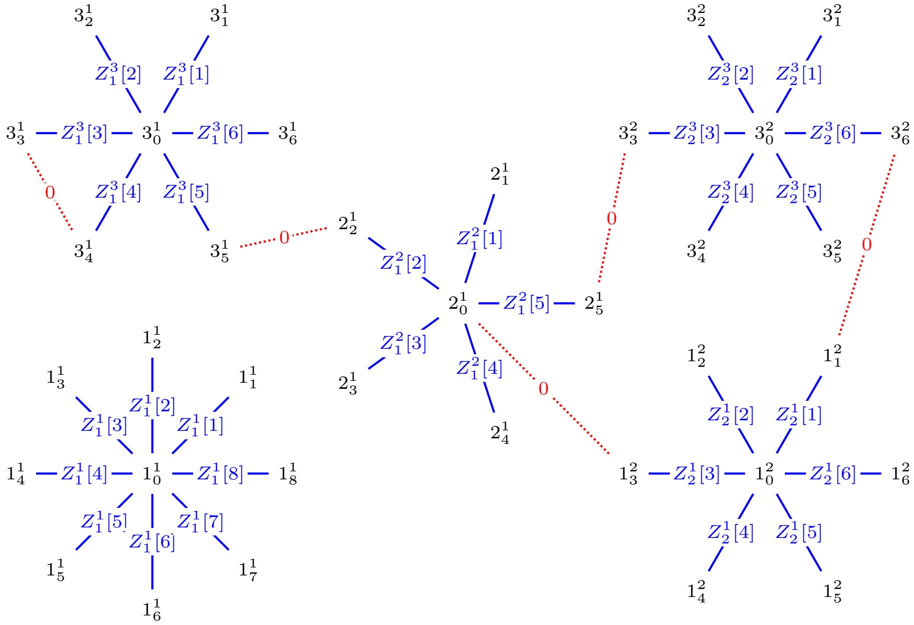

Fig. 8.1. A dependency graph. The vertex labels $i^j_k = (i, j, k)$ are purely informational.

Definition 8.4 (Snowflake). *For any $k \in \mathbb{N}^+$, a k-snowflake $\mathcal{T}$ of $\mathcal{G}$ is a maximally blue subtree of that contains exactly k red edges.*

Example 8.1. Consider the dependency graph illustrated in Fig. 8.1.

1. The graph consists of 5 stars, namely $\mathcal{S}^1_1$, $\mathcal{S}^1_2$, $\mathcal{S}^2_1$, $\mathcal{S}^3_1$, $\mathcal{S}^3_2$.
2. The graph consists of several maximally blue subgraphs, including the star $\mathcal{S}^1_1$ and the subgraph induced by the remaining stars.
3. The graph has several snowflakes, including the 0-snowflake $\mathcal{S}^1_1$, the 3-snowflake constructed by deleting the edges $3^1_3 \cdots 0 \cdots 3^1_4$ and $1^2_1 \cdots 0 \cdots 3^2_6$.

Proposition 8.2. *For some $k \geq 0$, let $\mathcal{T}$ be a k-snowflake of $\mathcal{G}$. Then $\mathcal{V}(\mathcal{T}) = \bigsqcup_{(i,j)\in\mathcal{I}} \mathcal{V}(\mathcal{S}^i_j)$, where $\mathcal{I} \subseteq \{(i, j) \in [q] \times [c_i]\}$ such that $|\mathcal{I}| = k + 1$.*

Definition 8.5 (Interesting Trail). *For $k = 0$, a 0-interesting trail is simply a blue path. For some $k \geq 1$, a k-interesting trail is a bichromatic trail that starts with a blue edge, and has exactly k red edges, with no two red edges adjacent to each other.*

Definition 8.6 (Circle-Graph). *For some $k \geq 1$, a k-circle is a k-interesting cycle that ends with a red edge. $\mathcal{G}$ is said to be a circle-graph if it contains a k-circle, for some $k \geq 1$, and a circle-free graph otherwise.*

Definition 8.7 (Line-Graph). *For some $k \geq 0$, a k-line p is a k-interesting path that ends with a blue edge, and has $\lambda(p) = 0$. $\mathcal{G}$ is said to be a line-graph if it contains a k-line for some $k \geq 0$, and a line-free graph otherwise.*

Example 8.2. Consider the dependency graph illustrated in Fig. 8.1.

1. The graph consists of several interesting trails including the 0-interesting trail $(1_1^1,\, 1_0^1,\, 1_6^1)$, the 4-interesting trail $(2_1^1,\, 2_0^1,\, 1_3^2,\, 1_0^2,\, 1_1^2,\, 3_6^2,\, 3_0^2,\, 3_3^2,\, 2_5^1,\, 2_0^1,\, 2_2^1,\, 3_5^1,\, 3_0^1,\, 3_1^1)$ etc.
2. The graph has two circles: the 1-circle $(3_4^1,\, 3_0^1,\, 3_3^1,\, 3_4^1)$ and the 3-circle $(1_3^2,\, 1_0^2,\, 1_1^2,\, 3_6^2,\, 3_0^2,\, 3_3^2,\, 2_5^1,\, 2_0^1,\, 1_3^2)$.
3. Suppose $Z_1^3[2] \oplus Z_1^3[5] \oplus Z_1^2[2] \oplus Z_2^1[3] \oplus Z_2^1[2] = 0^n$. Then, the graph contains a 2-line $(3_2^1,\, 3_0^1,\, 3_5^1,\, 2_2^1,\, 2_0^1,\, 1_3^2,\, 1_0^2,\, 1_2^2)$.

Additional Notations: Let $\mu = |\{(i,j) \in [q] \times [c_i]\}| \leq qc$ and $\nu = |\mathcal{V}_{|0}| = \sum_{(i,j)\in[q]\times[c_i]} r_{i,j} \leq r\mu$. Then, $|\mathcal{V}| = \mu + \nu \leq \mu(r+1) \leq 2\mu r \leq 2qcr$.

Proposition 8.3. *For $k \geq 0$:*

1. *The number of k-snowflakes is at most $(4r^2\mu)^{k+1}$.*
2. *The number of k-lines is at most $(2r^2\mu)^{k+1}$.*
3. *The number of $(k+1)$-circles is at most $(2r^2\mu)^{k+1}$.*

8.1 Vertex-Labeling of Dependency Graph

Definition 8.8 (Valid Vertex-Labeling). *A vertex-labeling $Y : \mathcal{V} \to \mathcal{B}$ of $\mathcal{G}$ is said to be valid if:*

1. *for all $(u,v) \notin \mathcal{E}$, $Y(u) \neq Y(v)$, and*
2. *for all $(u,v) \in \mathcal{E}$, $Y(u) \oplus Y(v) = \lambda(u,v)$.*

We write $h(\mathcal{G})$ to denote the number of valid vertex-labelings for $\mathcal{G}$.

One can also view a valid vertex-labeling Y as a sequence over $\mathcal{B}$, indexed by $\mathcal{V}$, by writing $Y_j^i[k] = Y(i,j,k)$ for all $(i,j,k) \in \mathcal{V}$. This view gives the following interesting property for any valid vertex-labeling of $\mathcal{G}$.

Proposition 8.4. *If Y is a valid vertex-labeling of $\mathcal{G}[X,Z]$ then X and Y are bijectively-consistent.*

Note that, the above result also hints at another interesting property. The number of vertices with distinct labels is independent of the valid vertex-labeling itself. In fact, using Proposition 8.4, this number is exactly

$$|X| \coloneqq |\{X_j^i[k] : (i,j,k) \in \mathcal{V}\}|.$$

Starting with Patarin's foundational works [38,39,40], a series of papers [38,39,40,13,18,11] developed an elegant combinatorial technique, the so-called *mirror theory*, to bound $h(\mathcal{G})$ under the assumption that $\mathcal{G}$ satisfies:

1. $\mathtt{CF}[X, Z]$: $\mathcal{G}$ is circle-free,
2. $\mathtt{LF}[X, Z]$: $\mathcal{G}$ is line-free, and
3. $\mathtt{GSF}_\xi[X, Z]$: every component in $\mathcal{G}$ has less than ξ vertices,

for some fixed $\xi \geq 2$. Borrowing nomenclature from Cogliati et al. [12], we say that $\mathcal{G}$ is *mirror theory compatible up to* ξ if the following predicate holds:

$$\mathtt{MTC}_\xi[X, Z] = \mathtt{CF}[X, Z] \wedge \mathtt{LF}[X, Z] \wedge \mathtt{GSF}_\xi[X, Z] \tag{16}$$

The following proposition reduces $\mathtt{GSF}_\xi[X, Z]$ to the existence of a $(k-1)$-snowflake in $\mathcal{G}$, where $k \geq \lceil \xi/(r+1) \rceil$.

Proposition 8.5. *If $\mathcal{G}$ does not contain a $(k-1)$-snowflake for all $k \geq \lceil \xi/(r+1) \rceil$ then $\mathtt{GSF}_\xi$ holds.*

The following result due to Cogliati et al. [11] is the fundamental theorem on valid vertex-labelings for mirror theory compatible dependency graphs.

Theorem 8.1 (Theorem 1 in [11]). *Suppose $|X| \leq \sqrt{2^n}$ or $\sqrt{2^n} \geq \xi(n\xi + 1)$, and $1 \leq |X| \leq 2^n/12\xi^2$ for some $\xi \geq 2$. If $\mathtt{MTC}_\xi[X, Z]$ holds then*

$$h(\mathcal{G}[X, Z]) \geq \frac{(2^n)_{|X|}}{2^{n\nu}}.$$

A rigorous and complete proof of this theorem is available in Cogliati et al. [11].

9 Proof of IV0 Security Theorem

First, by Proposition 2.1, we have

$$\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV0}}(\mathcal{A}) \leq \mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{Star}}(\mathcal{B}), \tag{17}$$

for some $(q, nr, \sigma + nq)$-PRBG distinguisher $\mathcal{B}$. Thus, it is sufficient to prove that Star is a secure PRBG. We will use the H-coefficient terminology and technique for this. Let $\mathcal{O}_1$ and $\mathcal{O}_0$ denote the oracles corresponding to Star (the real world) and $\$_{\mathsf{g}}$ (the ideal oracle), respectively. We adopt notations similar to those in Algorithm 2 to describe the transcripts. Specifically, for the i-th query:

- the input is given by $(T^i, r_{i,1}) \in \mathcal{T} \times [r]$,
- the keystream output is given by $Z^i = (Z^i_1[1], \ldots, Z^i_1[r_{i,1}])$.

At this point, $\mathcal{B}$ knows H_{T^i}, allowing it to immediately deduce the inputs to the underlying permutation (in both worlds). These inputs are denoted as

$$X^i = (X^i_1[0], \ldots, X^i_1[r_{i,1}]),$$

and satisfy $X^i_1[k] = \mathsf{H}_{T^i}(k)$ for all $k \in (r_{i,1}]$. Let $T := (T^i)_{i \in [q]}$, $Z := (Z^i)_{i \in [q]}$ and $X := (X^i)_{i \in [q]}$. Let $\mathcal{V}$ and $\mathcal{V}_{|0}$ denote the index sets corresponding to the sequences X and Z, respectively.

In addition, we modify the oracles to release the outputs of the underlying permutation after the query-response phase is over, but before $\mathcal{B}$ outputs its decision. This modification obviously does not decrease $\mathcal{B}$'s advantage. These outputs are denoted by the sequence $Y \coloneqq (Y^i)_{i\in[q]}$, where $Y^i = (Y_1^i[0], \ldots, Y_1^i[r_{i,1}])$.

In the real world, we store Y during the query-response phase and release it afterward.

In the ideal world, we sample Y uniformly at random from the valid vertex labelings of the dependency graph $\mathcal{G}[X, Z]$. This requires Z to be a sequence over $\mathcal{B} \setminus \{0^n\}$; otherwise, the graph is not defined. Consider the event

$$\texttt{ZRO}: \ \exists\,(i,k) \in [q] \times [r_{i,1}], \text{ such that } Z_1^i[k] = 0^n$$

Suppose $\texttt{ZRO}$ does not occur. Then $\mathcal{G}[X, Z]$ is well-defined. Let $\mathcal{L}$ denote the set of all valid vertex labelings of $\mathcal{G}$. At this stage, the reader might foresee our intention to use Theorem 8.1 to lower bound $h(\mathcal{G})$. Let $\xi \coloneqq (n+1)(r+1)$. Recall the predicate $\texttt{MTC}_\xi : \mathcal{G}$ is mirror theory compatible up to ξ.

Suppose $\texttt{MTC}_\xi$ holds. Then, we set $Y \leftarrow_{\$} \mathcal{L}$, and with a slight abuse of notation, we write $Y_1^i[k] = Y(i, 1, k)$ for all $(i,k) \in [q] \times (r_{i,1}]$.

Finally, we define $\texttt{Bad} \coloneqq (\texttt{ZRO} \vee \neg\texttt{MTC}_\xi)$, to capture the event that one of the aforementioned assumptions does not hold. If $\texttt{Bad}$ holds, then we set $Y_1^i[k] = 0^n$ for all $(i,k) \in [q] \times (r_{i,1}]$.

Transcript Analysis: Given the aforementioned sampling mechanism in the ideal world, the set of all attainable transcripts Ω can be deduced as the set of all tuples (t, x, y, z), where t, x, y, and z are sequences over $\mathcal{B}$ indexed analogously as T, X, Y, and Z, respectively, satisfying the following relations for all (i,k):

- $x_1^i[k] = \mathsf{H}_{t^i}(k)$.
- $\neg\texttt{ZRO} \wedge \texttt{MTC}_\xi$ implies that:
  - x and y are bijectively-consistent.
  - $z_1^i[k] = y_1^i[0] \oplus y_1^i[k]$, whenever $k \neq 0$.
- $\texttt{ZRO} \vee \neg\texttt{MTC}_\xi$ implies that $y_j^i[k] = 0^n$.

We define $\Omega_{\text{bad}} \coloneqq \{(t, x, y, z) \in \Omega : \texttt{Bad} \text{ holds}\}$. Any $\omega \in \Omega_{\text{bad}}$ is referred as *bad*, and the remaining transcripts are all *good*.

Lemma 9.1. *For $r(\|\sigma\| + q + r) \leq 2^{n-3}$ and 2-wise independent hash function H, we have*

$$\Pr(\Theta_0 \in \Omega_{\text{bad}}) \leq \frac{17r(\|\sigma\| + q + r)}{2^n}.$$

Proof. By definition, we have

$$\begin{aligned}\Pr(\Theta_0 \in \Omega_{\text{bad}}) = \Pr(\texttt{Bad}) &\leq \Pr(\texttt{ZRO}) + \Pr(\neg\texttt{MTC}_\xi \mid \neg\texttt{ZRO}) \\ &\leq \frac{\|\sigma\| + q}{2^n} + \Pr(\neg\texttt{MTC}_\xi \mid \neg\texttt{ZRO}), \qquad (18)\end{aligned}$$

where the last inequality follows from the uniformity of $Z_1^i[k]$ for all $(i,k) \in [q] \times [r_{i,1}]$, and the fact that $\sum_i r_{i,1} \leq \|\sigma\| + q$. Let $\mathtt{E} \coloneqq \neg\mathtt{ZRO} \wedge \mathtt{GSF}_\xi$. For the remaining terms on the r.h.s., Eq. (16) gives

$$\Pr\left(\neg\mathtt{MTC}_\xi \,|\, \neg\mathtt{ZRO}\right) \leq \Pr\left(\neg\mathtt{GSF}_\xi \,|\, \neg\mathtt{ZRO}\right) + \Pr\left(\neg\mathtt{CF} \,|\, \mathtt{E}\right) + \Pr\left(\neg\mathtt{LF} \,|\, \mathtt{E}\right). \tag{19}$$

We make the following claim with regards to the three terms on the r.h.s.:

Claim 9.1 *For $r(\|\sigma\| + q + r) \leq 2^{n-3}$, we have*

$$\Pr\left(\neg\mathtt{GSF}_\xi \,|\, \neg\mathtt{ZRO}\right) \leq \epsilon, \quad \Pr\left(\neg\mathtt{CF} \,|\, \mathtt{E}\right) \leq \epsilon, \quad \Pr\left(\neg\mathtt{LF} \,|\, \mathtt{E}\right) \leq 2\epsilon,$$

where $\epsilon = 4r(\|\sigma\| + q + r)/2^n$.

A proof of this claim is deferred to the Appendix D. The result follows by combining Eq. (18) and (19) with Claim 9.1. □

Coming back to the main proof, fix a good transcript $(t,x,y,z) \in \Omega \setminus \Omega_{\text{bad}}$. Let $\nu = |\mathcal{V}_{|0}|$. By hypothesis x and y is bijectively-consistent. Using $n^3 r^2 \leq 2^{\frac{n}{2}-5}$ and $n^2 r^2 (\|\sigma\| + q) \leq 2^n/48$ and Theorem 8.1, we have

$$\Pr\left(\Theta_0 = (t,x,y,z)\right) \leq \frac{1}{2^{n\nu}} \times \frac{2^{n\nu}}{(2^n)_{|x|}} \leq \frac{1}{(2^n)_{|x|}} = \Pr\left(\Theta_1 = (t,x,y,z)\right) \tag{20}$$

The result now follows from Lemma 9.1, Eq. (20) and (17), and Theorem 2.1.

10 Proof of $\mathsf{IV1}$ Security Theorem

Using Proposition 2.1, it is sufficient to bound $\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{GiantStar}}(\mathcal{B})$ for any $(q, \|\ell\|, \|\sigma\| + q)$-PRBG distinguisher $\mathcal{B}$. The proof idea is mostly the same as in the proof of Theorem 4.1 given in the previous section. So, we borrow (and extend) notations from Section 9 whenever convenient. For the i-th query:

- the input is given by $(T^i, \|\ell_i\|) \in \mathcal{T} \times [\|\ell\|]$,
- the keystream output is given by $Z^i = (Z_1^i, \ldots, Z_{\|\ell_i\|_r}^i)$, where for all $(i,j) \in [q] \times [\|\ell_i\|_r]$, $Z_j^i = (Z_j^i[1], \ldots, Z_j^i[r_{i,j}])$, and $r_{i,j} = r$ for all $(i,j) \in [q] \times [\|\ell_i\|_r - 1]$, while $r_{i,\|\ell_i\|_r} \in [r]$.

At this point, $\mathcal{B}$ can deduce the inputs to the underlying permutation (in both worlds). These inputs are denoted as $X^i = (X_1^i, \ldots, X_{\|\ell_i\|_r}^i)$, where $X_j^i = (X_j^i[0], \ldots, X_j^i[r_{i,j}])$ for all $(i,j) \in [q] \times [\|\ell_i\|_r]$ such that $X_j^i[k] = \mathsf{H}_{T^i}(\widehat{j}+k)$ for all $k \in (r_{i,1}]$ and $\widehat{j} = (j-1)(r+1)$. Let $T = (T^i)_{i \in [q]}$, $Z = (Z^i)_{i \in [q]}$ and $X = (X^i)_{i \in [q]}$. Let $\mathcal{V}$ and $\mathcal{V}_{|0}$ denote the index sets corresponding to the sequences X and Z, respectively.

In addition, the oracles release the outputs of the underlying permutation after the query-response phase is over, but before $\mathcal{B}$ outputs its decision. These outputs are denoted by the sequence $Y \coloneqq (Y^i)_{i \in [q]}$, where $Y^i = (Y_1^i, \ldots, Y_{\|\ell_i\|_r}^i)$

and $Y^i_j = (Y^i_j[0], \dots, Y^i_j[r_{i,j}])$ for all $(i,j) \in [q] \times [\|\ell_i\|_r]$. We skip the description of the sampling strategy since it is identical to the one used in Section 9, with

$$\texttt{ZRO}: \ \exists\,(i,j,k) \in [q] \times [\|\ell_i\|_r] \times [r_{i,j}], \text{ such that } Z^i_j[k] = 0^n,$$

while $\texttt{MTC}_\xi$ and $\texttt{Bad}$ are defined as before in Section 9.

Transcript Analysis: Given the aforementioned sampling mechanism in the ideal world, the set of all attainable transcripts Ω can be deduced as the set of all tuples (t,x,y,z), where t, x, y, and z are sequences over $\mathcal{B}$ indexed analogously as T, X, Y, and Z, respectively, satisfying the following relations for all (i,j,k):

- $x^i_j[k] = \mathsf{H}_{t^i}(\widehat{j} + k)$.
- $\neg\texttt{ZRO} \wedge \texttt{MTC}_\xi$ implies that:
  - x and y are bijectively-consistent.
  - $z^i_j[k] = y^i_j[0] \oplus y^i_j[k]$, whenever $k \neq 0$.
- $\texttt{ZRO} \vee \neg\texttt{MTC}_\xi$ implies that $y^i_j[k] = 0^n$.

We define $\Omega_{\text{bad}} \coloneqq \{(t,x,y,z) \in \Omega : \texttt{Bad} \text{ holds}\}$. Any $\omega \in \Omega_{\text{bad}}$ is referred as *bad*, and the remaining transcripts are all *good*.

Lemma 10.1. *For $r(\|\sigma\| + q + r) \leq 2^{n-3}$, and H instantiated with gwc hash function, we have*

$$\Pr\left(\Theta_0 \in \Omega_{\text{bad}}\right) \leq \frac{16r^2\|\ell\|(\|\sigma\| + q + r)^2}{2^{2n}} + \frac{18r(\|\sigma\| + q + r)}{2^n}.$$

Proof. Following the proof of Lemma 9.1, first we get

$$\Pr\left(\Theta_0 \in \Omega_{\text{bad}}\right) \leq \frac{\|\sigma\| + q}{2^n} + \Pr\left(\neg\texttt{MTC}_\xi \mid \neg\texttt{ZRO}\right). \tag{21}$$

Now, we say that two vertices $u, v \in \mathcal{V}$ are *query-related*, denoted $u \sim v$, if $u = (i,*,*)$ and $v = (i,*,*)$ for some $i \in [q]$, where $*$ denotes some *appropriate* value. Observe that, for any sequence of query-related indices $(u_1, \dots, u_k)$, the corresponding subsequence $(X_{u_1}, \dots, X_{u_k})$ of X can be at most 2-wise independent. Note that, k can be as large as $\|\ell\|$. This differs from the proof of Theorem 4.1, where this limitation only applies within chunks of size at most r. As a consequence, we need the following auxiliary event to avoid a large degradation in the bound on $\Pr\left(\neg\texttt{MTC}_\xi \mid \neg\texttt{ZRO}\right)$:

$\texttt{AUX}$: for some $d \in [n]$, there exists a d-interesting path $(u_1, u_2, \dots, u_{d+1})$ such that $u_1 \sim u_{d+1}$, and $u_i \not\sim u_{i+1}$ otherwise.

Let $\texttt{E} \coloneqq \neg\texttt{ZRO} \wedge \neg\texttt{AUX} \wedge \texttt{GSF}_\xi$. Coming back to Eq. 21, we now have

$$\begin{aligned}\Pr\left(\neg\texttt{MTC}_\xi \mid \neg\texttt{ZRO}\right) &\leq \Pr\left(\texttt{AUX} \mid \neg\texttt{ZRO}\right) + \Pr\left(\neg\texttt{GSF}_\xi \mid \neg\texttt{AUX} \wedge \neg\texttt{ZRO}\right) \\ &\qquad + \Pr\left(\neg\texttt{CF} \mid \texttt{E}\right) + \Pr\left(\neg\texttt{LF} \mid \texttt{E}\right) \\ &\leq \Pr\left(\texttt{AUX} \mid \neg\texttt{ZRO}\right) + \frac{16r(\|\sigma\| + q + r)}{2^n}.\end{aligned} \tag{22}$$

where the last inequality follows from the observation that once we condition on $\neg\mathtt{AUX}$ the analysis of $\neg\mathtt{GSF}_{\xi}$, $\neg\mathtt{CF}$ and $\neg\mathtt{LF}$ is analogous to the ones given in Claim 9.1. We make the following claim with regards to $\Pr\left(\mathtt{AUX} \mid \neg\mathtt{ZRO}\right)$:

Claim 10.1 *For $r(\|\sigma\| + q + r) \leq 2^{n-3}$:*

$$\Pr\left(\mathtt{AUX} \mid \neg\mathtt{ZRO}\right) \leq \frac{16r^2\|\ell\|(\|\sigma\| + q + r)^2}{2^{2n}} + \frac{q}{2^n}.$$

A proof of this claim is deferred to Appendix E. The result follows by combining Eq. (21) and (22) with Claim 10.1. □

By observing that the good transcript analysis is again identical with the one in Section 9, we establish the result using Lemma 10.1 and Theorem 2.1.

11 Conclusion

We presented DENC1 and DENC2, two highly secure DAE schemes based on a block cipher. We provided a complete security analysis of both constructions in the standard prp model. Our analysis showed that both DENC1 and DENC2 compare favorably to well-known AE schemes in terms of security and efficiency. A possible future research direction for this work is to reduce the number of needed keys, while another promising avenue is to analyze the multi-user security of DENC1 and DENC2.

References

1. Aiello, W., Venkatesan, R.: Foiling birthday attacks in length-doubling transformations - Benes: A non-reversible alternative to Feistel. In: EUROCRYPT'96. LNCS, vol. 1070, pp. 307–320
2. Banik, S., Bogdanov, A., Luykx, A., Tischhauser, E.: SUNDAE: Small universal deterministic authenticated encryption for the internet of things. IACR Trans. Symm. Cryptol. **2018**(3), 1–35
3. Bao, Z., Hwang, S., Inoue, A., Lee, B., Lee, J., Minematsu, K.: XOCB: beyond-birthday-bound secure authenticated encryption mode with rate-one computation. In: EUROCRYPT 2023, Part IV. LNCS, vol. 14007, pp. 532–561
4. Bellare, M., Krovetz, T., Rogaway, P.: Luby-Rackoff backwards: Increasing security by making block ciphers non-invertible. In: EUROCRYPT'98. LNCS, vol. 1403, pp. 266–280
5. Böck, H., Zauner, A., Devlin, S., Somorovsky, J., Jovanovic, P.: Nonce-disrespecting adversaries: Practical forgery attacks on GCM in TLS. In: WOOT 16
6. Bose, P., Hoang, V.T., Tessaro, S.: Revisiting AES-GCM-SIV: Multi-user security, faster key derivation, and better bounds. In: EUROCRYPT 2018, Part I. LNCS, vol. 10820, pp. 468–499
7. Chakraborty, B., Chattopadhyay, S., Jha, A., Nandi, M.: On length independent security bounds for the PMAC family. IACR Cryptol. ePrint Arch. **2020**, 656
8. Chang, D., Datta, N., Dutta, A., Mennink, B., Nandi, M., Sanadhya, S., Sibleyras, F.: Release of unverified plaintext: Tight unified model and application to ANYDAE. IACR Trans. Symm. Cryptol. **2019**(4), 119–146

9. Chen, S., Steinberger, J.P.: Tight security bounds for key-alternating ciphers. In: Advances in Cryptology - EUROCRYPT '14. Proceedings. pp. 327–350
10. Choi, W., Lee, B., Lee, J., Lee, Y.: Toward a fully secure authenticated encryption scheme from a pseudorandom permutation. In: ASIACRYPT 2021, Part III. LNCS, vol. 13092, pp. 407–434
11. Cogliati, B., Dutta, A., Nandi, M., Patarin, J., Saha, A.: Proof of mirror theory for any $\xi_{\max}$. IACR Cryptol. ePrint Arch. p. 686
12. Cogliati, B., Jha, A., Nandi, M.: How to build optimally secure PRFs using block ciphers. In: ASIACRYPT 2020, Part I. LNCS, vol. 12491, pp. 754–784
13. Cogliati, B., Patarin, J.: Mirror theory: A simple proof of the pi+pj theorem with xi_max=2. IACR Cryptol. ePrint Arch. p. 734
14. additional public comments, N.: NIST SP 800-38A Additional Public Comments, https://csrc.nist.gov/csrc/media/Projects/crypto-publication-review-project/documents/decision-proposal-comments/sp800-38a-decision-proposal-comments-2022.pdf
15. initial public comments, N.: NIST SP 800-38A Initial Public Comments, https://csrc.nist.gov/CSRC/media/Projects/crypto-publication-review-project/documents/initial-comments/sp800-38a-initial-public-comments-2021.pdf
16. public comments, N.: NIST SP 800-38A Review, https://csrc.nist.gov/News/2023/decision-to-revise-nist-sp-800-38a
17. Damgård, I., Nielsen, J.B.: Expanding pseudorandom functions; or: From known-plaintext security to chosen-plaintext security. In: Advances in Cryptology - CRYPTO 2002, Proceedings. pp. 449–464
18. Dutta, A., Nandi, M., Saha, A.: Proof of mirror theory for $\xi_{\max}=2$. IACR Cryptol. ePrint Arch. p. 669
19. Dutta, A., Nandi, M., Talnikar, S.: Beyond birthday bound secure MAC in faulty nonce model. In: EUROCRYPT 2019, Part I. LNCS, vol. 11476, pp. 437–466
20. Gueron, S., Lindell, Y.: GCM-SIV: Full nonce misuse-resistant authenticated encryption at under one cycle per byte. In: ACM CCS 2015. pp. 109–119
21. Iwata, T.: New blockcipher modes of operation with beyond the birthday bound security. In: FSE 2006. LNCS, vol. 4047, pp. 310–327
22. Iwata, T., Minematsu, K.: Stronger security variants of GCM-SIV. IACR Trans. Symm. Cryptol. **2016**(1), 134–157, https://tosc.iacr.org/index.php/ToSC/article/view/539
23. Iwata, T., Minematsu, K., Peyrin, T., Seurin, Y.: ZMAC: A fast tweakable block cipher mode for highly secure message authentication. In: CRYPTO 2017, Part III. LNCS, vol. 10403, pp. 34–65
24. Iwata, T., Seurin, Y.: Reconsidering the security bound of AES-GCM-SIV. IACR Trans. Symm. Cryptol. **2017**(4), 240–267
25. Jha, A., Nandi, M.: A survey on applications of h-technique: Revisiting security analysis of PRP and PRF. Entropy **24**(4), 462
26. Krovetz, T., Rogaway, P.: The software performance of authenticated-encryption modes. In: FSE 2011. LNCS, vol. 6733, pp. 306–327
27. Leech, D.P., Ferris, S., Scott, J.T.: The Economic Impacts of the Advanced Encryption Standard, 1996-2017. In: (National Institute of Standards and Technology, Gaithersburg, MD), NIST Grant/Contract Reports (GCR) 18-017. https://doi.org/10.6028/NIST.GCR.18-017
28. List, E., Nandi, M.: Revisiting full-PRF-secure PMAC and using it for beyond-birthday authenticated encryption. In: CT-RSA 2017. LNCS, vol. 10159, pp. 258–274

29. McGrew, D.A., Viega, J.: The security and performance of the Galois/counter mode (GCM) of operation. In: INDOCRYPT 2004. LNCS, vol. 3348, pp. 343–355
30. Minematsu, K., Tsunoo, Y.: Expanding weak PRF with small key size. In: Information Security and Cryptology - ICISC 2005, Revised Selected Papers. pp. 284–298
31. Naito, Y.: The exact security of PMAC with two powering-up masks. IACR Trans. Symm. Cryptol. **2019**(2), 125–145
32. Naito, Y., Sasaki, Y., Sugawara, T.: Lightweight authenticated encryption mode suitable for threshold implementation. In: EUROCRYPT 2020, Part II. LNCS, vol. 12106, pp. 705–735
33. Naito, Y., Sasaki, Y., Sugawara, T.: LM-DAE: Low-memory deterministic authenticated encryption for 128-bit security. IACR Trans. Symm. Cryptol. **2020**(4), 1–38
34. Naito, Y., Sasaki, Y., Sugawara, T.: Secret can be public: Low-memory AEAD mode for high-order masking. In: CRYPTO 2022, Part III. LNCS, vol. 13509, pp. 315–345
35. Naito, Y., Sugawara, T.: Lightweight authenticated encryption mode of operation for tweakable block ciphers. IACR TCHES **2020**(1), 66–94, https://tches.iacr.org/index.php/TCHES/article/view/8393
36. Namprempre, C., Rogaway, P., Shrimpton, T.: Reconsidering generic composition. In: EUROCRYPT 2014. LNCS, vol. 8441, pp. 257–274
37. Patarin, J.: The "coefficients H" technique. In: Selected Areas in Cryptography - SAC '08. Revised Selected Papers. pp. 328–345
38. Patarin, J.: A proof of security in o(2n) for the xor of two random permutations. In: Information Theoretic Security, Third International Conference, ICITS 2008. Proceedings. pp. 232–248
39. Patarin, J.: Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. IACR Cryptology ePrint Archive **2010**, 287
40. Patarin, J.: Mirror theory and cryptography. Appl. Algebra Eng. Commun. Comput. **28**(4), 321–338
41. Patarin, J., Montreuil, A.: Benes and butterfly schemes revisited. In: Information Security and Cryptology - ICISC '05. Revised Selected Papers. pp. 92–116
42. Peyrin, T., Seurin, Y.: Counter-in-tweak: Authenticated encryption modes for tweakable block ciphers. In: CRYPTO 2016, Part I. LNCS, vol. 9814, pp. 33–63
43. Rogaway, P., Shrimpton, T.: A provable-security treatment of the key-wrap problem. In: EUROCRYPT 2006. LNCS, vol. 4004, pp. 373–390
44. Wegman, M.N., Carter, L.: New classes and applications of hash functions. In: 20th Annual Symposium on Foundations of Computer Science, San Juan, Puerto Rico, 29-31 October 1979. pp. 175–182
45. Wegman, M.N., Carter, L.: New hash functions and their use in authentication and set equality. J. Comput. Syst. Sci. **22**(3), 265–279
46. agenda for the workshop, N.: Third NIST Workshop on Block Cipher Modes of Operation 2023, https://csrc.nist.gov/csrc/media/Events/2023/third-workshop-on-block-cipher-modes-of-operation/documents/agenda-bcm-workshop-2023.pdf
47. Yasuda, K.: A new variant of PMAC: Beyond the birthday bound. In: CRYPTO 2011. LNCS, vol. 6841, pp. 596–609

Appendix

A Proof of F^* Security Theorem

First, we recall some relevant notions of universality. It would be convenient to fix an arbitrary sequence of distinct inputs $X_1, \ldots, X_q \in \mathcal{X}$ for any $q \geq 2$ such that $\ell = \max_i |X_i|$ and $\sum_{i \in [q]} |X_i| = \sigma$.

Almost Universal Hash Function: A $(\mathcal{K}, \mathcal{X}, \mathcal{B})$-keyed function H is said to be a $(q, \ell, \sigma, \epsilon)$-*almost universal (AU) hash function* if:

$$\Pr\left(\exists i \neq j \;:\; \mathsf{H}_K(X_i) = H_K(X_j)\right) \leq \epsilon. \tag{23}$$

holds for some $\epsilon \in [0, \infty)$ and $K \leftarrow\!\!\$\, \mathcal{K}$.

Almost Collision-free Universal Hash Function [12]: A $(\mathcal{K}, \mathcal{X}, \mathcal{B})$-keyed function H is said to be $(q, \ell, \sigma, \epsilon)$-*Almost θ-Collision-free Universal* (or ACU$_\theta$) if one has $\Pr\left(C \geq \theta\right) \leq \epsilon$ for some $\epsilon \in [0, \infty)$, where

$$C \coloneqq |\{(i,j) \,:\, i < j \in [q], \mathsf{H}_K(X_i) = \mathsf{H}_K(X_j)\}|.$$

Diblock ACU$_q$ Hash Function [12]: A $(\mathcal{K}, \mathcal{X}, \mathcal{T})$-keyed function H is said to be a $(q, \ell, \sigma, \epsilon_1, \epsilon_2)$-*Diblock ACU$_q$* (or DbACU$_q$) if H is $(q, \ell, \sigma, \epsilon_1)$-AU and H_1 and H_2 are $(q, \ell, \sigma, \epsilon_2)$-ACU$_q$, where H_1 and H_2 denote the leftmost and rightmost n bits of H, respectively.

The Proof. The following result is a minor variation of [12, Theorem 7.3], and a proof follows much in the same manner as the proof of [12, Theorem 7.3 and Theorem 4.4] and [11, Theorem 3].

Proposition A.1 (Extension of Theorem 7.3 from [12]). *Suppose* PHASH^* *is a $(q, \ell, \sigma, \epsilon_1, \epsilon_2)$-DbACU$_q$ hash function, $n^3 \leq 2^{\frac{n}{2}-1}$ and $n^2 q \leq 2^{n-4}$. Then, for any (q, ℓ, σ)-PRF distinguisher $\mathcal{A}$*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{F}^*}(\mathcal{A}) \leq \frac{128q^2}{2^{3n}} + \frac{136q^2}{2^{2n}} + \frac{8q}{2^n} + \epsilon_1 + 2\epsilon_2.$$

Thus it is sufficient to show that $\mathsf{PHASH}^*_{\boldsymbol{\pi}_1}$ is a DbACU$_q$ hash function for $\boldsymbol{\pi}_1 \leftarrow\!\!\$\, \mathcal{P}(\mathcal{B})$.

Fix any sequence of distinct input $(A_1, C_1), \ldots, (A_q, C_q) \in \mathcal{H} \times \mathcal{M}$. Since the mapping $(A, C) \mapsto \mathtt{ozs}_n(A) \parallel \mathtt{ozs}_n(C)$ is injective, it would be convenient to simply consider the combined input, $D_i \coloneqq \mathtt{ozs}_n(A_i) \parallel \mathtt{ozs}_n(C_i)$ for all $i \in [q]$. From [47, Lemma 2 and 3], for distinct $D \neq D'$ such that $|D|, |D'| \leq n2^{n-2}$, we have

$$\Pr\left(\mathsf{PHASH1}^*_{\boldsymbol{\pi}_1}(D) = \mathsf{PHASH1}^*_{\boldsymbol{\pi}_1}(D')\right) \leq \frac{16\max\{\|D\|, \|D'\|\}}{2^n}$$

$$\Pr\left(\mathsf{PHASH2}^*_{\boldsymbol{\pi}_1}(D) = \mathsf{PHASH2}^*_{\boldsymbol{\pi}_1}(D')\right) \leq \frac{16\max\{\|D\|, \|D'\|\}}{2^n}$$

where $\mathsf{PHASH1}^*$ and $\mathsf{PHASH2}^*$ are the leftmost and rightmost n bits of PHASH^*. Then, a simple application of Markov's inequality gives

Proposition A.2. *For $\|\ell\| \leq 2^{n-2}$, we have*

- $\mathsf{PHASH1}^*_{\boldsymbol{\pi}_1}$ *is* $(q, \ell, \sigma, \epsilon_2)$*-*$ACU_q$*,*
- $\mathsf{PHASH2}^*_{\boldsymbol{\pi}_1}$ *is* $(q, \ell, \sigma, \epsilon_2)$*-*$ACU_q$*,*

where $\epsilon_2 \coloneqq (\|\sigma\| + q)/2^{n-4}$*.*

All that remains is to show that $\mathsf{PHASH}^*_{\boldsymbol{\pi}_1}$ is a $(q, \ell, \sigma, \epsilon_1)$-AU for an appropriate ϵ_1, where the sequence of messages is denoted $(D_1, \ldots, D_q)$. We extend and reuse notations from Algorithm 1 whenever necessary. Let $E_i[j] = \Delta_j \oplus D_i[j]$. Let

- $\texttt{ZRO}$: $\boldsymbol{\pi}_1(10^{n-1}) = 0^n$,
- $\texttt{AUX}_1$: $\exists\, (i, j) \in [q] \times [\|\ell_i\|]$ such that $E_i[j] = 0^n$,
- $\texttt{AUX}_2$: $\exists\, (i, j) \in [q] \times [\|\ell_i\|]$ such that $E_i[j] = 10^{n-1}$,
- $\texttt{AUX}_3$: $\exists\, i \in [q]$ and pairwise distinct $j_1, j_2, j_3 \in [\|\ell_i\|]$ such that

$$E_i[j_1] = E_i[j_2] = E_i[j_3],$$

and $\texttt{AUX} \coloneqq \texttt{ZRO} \cup \texttt{AUX}_1 \cup \texttt{AUX}_2 \cup \texttt{AUX}_3$. For all $i \neq j \in [q]$ and $b_0, b_1 \in \{0, 1\}$, let $\texttt{COL}_{i,j,b_0,b_1}$ be the event

$$\mathsf{PHASH}^*_{\boldsymbol{\pi}_1}(D_i) \oplus \mathsf{PHASH}^*_{\boldsymbol{\pi}_1}(D_j) = b_0 \parallel 0^{n-1} \parallel b_1 \parallel 0^{n-1},$$

and $\texttt{COL} \coloneqq \bigcup_{\substack{i \neq j \\ b_0, b_1}} \texttt{COL}_{i,j,b_0,b_1}$. Then, we have

$$\begin{aligned}
\Pr(\texttt{COL}) &\leq \Pr(\texttt{AUX}) + \Pr(\texttt{COL} \mid \neg\texttt{AUX}) \\
&\leq \Pr(\texttt{ZRO}) + \Pr(\texttt{AUX}_1) + \Pr(\texttt{AUX}_2) + \Pr(\texttt{AUX}_3) + \Pr(\texttt{COL} \mid \neg\texttt{AUX}) \\
&\leq \frac{1}{2^n} + \frac{2(\|\sigma\| + q)}{2^n - 1} + \Pr(\texttt{AUX}_3) + \Pr(\texttt{COL} \mid \neg\texttt{AUX}) \qquad (24)
\end{aligned}$$

where the first two terms correspond to $\Pr(\texttt{ZRO})$ and $\Pr(\texttt{AUX}_1) + \Pr(\texttt{AUX}_2)$, respectively.

Now, consider the event $\texttt{AUX}_3$. Fix any $i \in [q]$ and any pairwise distinct $j_1, j_2, j_3 \in [\|\ell_i\|]$. We can rewrite $E_i[j_1] = E_i[j_2] = E_i[j_3]$ as

$$\begin{aligned}
(2^{j_1} \oplus 2^{j_2})\boldsymbol{\pi}_1(0^n) \oplus (2^{2j_1} \oplus 2^{2j_2})\boldsymbol{\pi}_1(10^{n-1}) &= D_i[j_1] \oplus D_i[j_2] \\
(2^{j_1} \oplus 2^{j_3})\boldsymbol{\pi}_1(0^n) \oplus (2^{2j_1} \oplus 2^{2j_3})\boldsymbol{\pi}_1(10^{n-1}) &= D_i[j_1] \oplus D_i[j_3].
\end{aligned}$$

Given that $j_1, j_2, j_3 < 2^{n-1}$, the values $2^{j_1}, 2^{j_2}, 2^{j_3}$ are pairwise distinct. Thus, the aforementioned system has a unique solution that holds with probability at most $1/2^n(2^n - 1)$. Summing over all possible choices yield

$$\Pr(\texttt{AUX}_3) \leq \sum_{i=1}^{q} \frac{\|\ell_i\|^3}{2^n(2^n - 1)} \leq \sum_{i=1}^{q} \frac{\|\ell_i\|}{2^n - 1} \leq \frac{\|\sigma\| + q}{2^n - 1}, \qquad (25)$$

where the second inequality follows from $\ell \leq n2^{\frac{n}{2}-1}$. For $\mathtt{COL} \mid \neg\mathtt{AUX}$, we claim

$$\Pr\left(\mathtt{COL} \mid \neg\mathtt{AUX}\right) \leq \frac{8(\|\sigma\| + q)(\|\sigma\| + 15q)}{2^{2n}} + \frac{12\|\sigma\| + 28q}{2^n - 2}. \tag{26}$$

A proof of this claim follows from the proof of [12, Claim 6.2]. Combining Eq. (24)-(26) gives

Proposition A.3. *For $q, \ell, \sigma \in \mathbb{N}^+$ and $\|\ell\| \leq 2^{\frac{n}{2}-1}$, $\mathsf{PHASH}^*_{\boldsymbol{\pi}_1}$ is $(q, \ell, \sigma, \epsilon_1, \epsilon_2)$-DbACU$_q$ hash function, where*

$$\epsilon_1 \leq \frac{8(\|\sigma\| + q)(\|\sigma\| + 15q)}{2^{2n}} + \frac{15\|\sigma\| + 32q}{2^n - 2},$$
$$\epsilon_2 \leq \frac{16(\|\sigma\| + q)}{2^n}.$$

Theorem 3.1 follows from Proposition A.1 and A.3.

B Proof of IV2 Security Theorem

First, triangle inequality gives

$$\begin{aligned}\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV2}}(\mathcal{A}) &= \mathbf{CD}(\mathsf{IV2}[\mathsf{G}].\mathsf{Enc}_{K,\boldsymbol{\pi}} - \$_{\mathsf{e}} \mid \mathcal{A}) \\ &\leq \mathbf{CD}(\mathsf{IV2}[\mathsf{G}].\mathsf{Enc}_{K,\boldsymbol{\pi}} - \mathsf{IV2}[\$_{\mathsf{g}}].\mathsf{Enc}_{\boldsymbol{\pi}} \mid \mathcal{A}) \\ &\quad + \mathbf{CD}(\mathsf{IV2}[\$_{\mathsf{g}}].\mathsf{Enc}_{\boldsymbol{\pi}} - \$_{\mathsf{e}} \mid \mathcal{A})\end{aligned} \tag{27}$$

For the first term on the r.h.s. of Eq. (27), we construct a $(q, 2\ell + 2n, 2\sigma + 6nq)$-PRBG distinguisher $\mathcal{B}$ as follows: it runs $\mathcal{A}$ in a black box manner, answering its i-th query by first generating $W_1 \parallel \ldots \parallel W_{\|\ell_i\|_r}$ (see step 5 of Algorithm 4) using its own oracle, and then following Algorithm 4 from step 6 onwards, where Star is instantiated with $\boldsymbol{\pi}$. Then clearly, $\mathcal{B}$ correctly simulates $\mathsf{IV2}[\mathsf{G}].\mathsf{Enc}_{K,\boldsymbol{\pi}}$ when its oracle is G_K, and it correctly simulates $\mathsf{IV2}[\$_{\mathsf{g}}].\mathsf{Enc}_{\boldsymbol{\pi}}$ when its oracle is $\$_{\mathsf{g}}$, as long as there is no collision in the T values. Moreover, $\mathcal{B}$ simply relays the output of $\mathcal{A}$ at the end of the query-response phase. Thus we have

$$\begin{aligned}\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{G}}(\mathcal{B}) &= \mathbf{CD}(\mathsf{G}_K - \$_{\mathsf{g}} \mid \mathcal{A}) \\ &\geq \mathbf{CD}(\mathsf{IV2}[\mathsf{G}].\mathsf{Enc}_{K,\boldsymbol{\pi}} - \mathsf{IV2}[\$_{\mathsf{g}}].\mathsf{Enc}_{\boldsymbol{\pi}} \mid \mathcal{A}) - \frac{q^2}{2^{2n}},\end{aligned} \tag{28}$$

where the second term on the r.h.s. is due to the collision probability of T values.

Next for the second term on the r.h.s. of Eq. (27), we construct a $(\|\sigma\|_r + 2q, rn, \sigma + 4rnq)$-Priv\$ distinguisher $\mathcal{C}$ as follows: it runs $\mathcal{A}$ in a black box manner, answering its i-th query using the following steps:

1. Set $W_1 \parallel \ldots \parallel W_{\|\ell_i\|_r}$ as $\$_{\mathsf{g}}(T^i, 2\|\ell_i\|_r)$.
2. For $j \in [\|\ell_i\|_r - 1]$ do:
   (a) Query (W_j, M^i_j) and suppose the response is C^i_j, where $M^i_j = (M^i_j[1], \ldots, M^i_j[r])$ denotes j-th chunk of the input message and $r_{i,j} = r$.

3. Query $(W_j, M^i_{\|\ell_i\|_r})$ and suppose the response is $C^i_{\|\ell_i\|_r}$, where $M^i_j = (M^i_j[1], \ldots, M^i_j[r_{i,\|\ell_i\|_r}])$ denotes the last chunk of the input message and $r_{i,\|\ell_i\|_r} \in [r]$.

Once again, $\mathcal{C}$ correctly simulates $\mathsf{IV2}[\$_{\mathsf{g}}].\mathsf{Enc}_{\boldsymbol{\pi}}$ when its oracle is $\mathsf{IV0}.\mathsf{Enc}_{\boldsymbol{\pi}}$, and it correctly simulates $\$_{\mathsf{e}}$ when its oracle is $\$_{\mathsf{e}}$, as long as there is no collision in the W values. It relays the output of $\mathcal{A}$ as it is at the end of the query-response phase. Thus we have

$$\begin{aligned}\mathbf{Adv}^{\mathsf{priv\$}}_{\mathsf{IV0}}(\mathcal{C}) &= \mathbf{CD}(\mathsf{IV0}.\mathsf{Enc}_{\boldsymbol{\pi}} - \$_{\mathsf{e}} \mid \mathcal{A}) \\ &\geq \mathbf{CD}(\mathsf{IV2}[\$_{\mathsf{g}}].\mathsf{Enc}_{\boldsymbol{\pi}} - \$_{\mathsf{e}} \mid \mathcal{A}) - \frac{4(\|\sigma\|_r + 2q)^2}{2^{2n}}, \qquad (29)\end{aligned}$$

where the second term on the r.h.s. is due to the collision probability of W values. The result now follows by combining Eq. (27)-(29) followed by some simplification.

C Proofs for the Dependency Graph Results

C.1 Proof of Proposition 8.2

Since $\mathcal{T}$ is a k-snowflake, we must have k red edges in $\mathcal{T}$. Suppose there exists a star, say $\mathcal{S}^i_j$, such that $0 < \mathcal{V}(\mathcal{T}) \cap \mathcal{V}(\mathcal{S}^i_j) < r_{i,j}+1$. Let $\mathcal{V}' = \mathcal{V}(\mathcal{T}) \cap \mathcal{V}(\mathcal{S}^i_j)$. Since $\mathcal{S}^i_j$ is connected, any $u \in \mathcal{V}'$ and $v \in \mathcal{V}(\mathcal{S}^i_j)$ are connected. More importantly, $\{u, v\}$ is a blue edge, whence $v \in \mathcal{V}(\mathcal{T})$ for all $v \in \mathcal{V}(\mathcal{S}^i_j)$, otherwise it contradicts the maximally blue property of $\mathcal{T}$. Therefore, we have $\mathcal{V}(\mathcal{S}^i_j) \subseteq \mathcal{V}(\mathcal{T})$, leading to a contradiction. Thus, we must have $\mathcal{V}(\mathcal{T}) = \bigsqcup_{(i,j)\in\mathcal{I}} \mathcal{V}(\mathcal{S}^i_j)$, for some $\mathcal{I} \subseteq \{(i,j) \in [q] \times [c_i]\}$. Let $|\mathcal{I}| = k'$. Then, $|\mathcal{V}(\mathcal{T})| = \sum_{(i,j)\in\mathcal{I}}(r_{i,j}+1) = k' + \sum_{(i,j)\in\mathcal{I}} r_{i,j}$, and $\mathcal{E}(\mathcal{T}) = k' - 1 + \sum_{(i,j)\in\mathcal{I}} r_{i,j}$. Now, exactly $\sum_{(i,j)\in\mathcal{I}} r_{i,j}$ edges are blue, accounting for the $r_{i,j}$ edges in each (i,j)-th star in $\mathcal{T}$. Rest of the $(k'-1)$ edges must all be red as edges between the stars are always red, whence $k' = k+1$. □

C.2 Proof of Proposition 8.3

First consider the number of snowflakes. From Proposition 8.2 we know that there are exactly $k+1$ stars and k red edges in any k-snowflake. There are at most μ^{k+1} ways to choose these stars. Further, there are at most $(r+1)^{2k}$ ways to choose the vertices that incident on the red edges. Once these indices are chosen the k-snowflake is fixed. Thus, the number of k-snowflakes is at most $(4r^2\mu)^{k+1}$.

Next, we show the bound for k-line. The bound for $(k+1)$-circle can be derived similarly. Partition the k-line p into a sequence of sub-paths $(p_1, \ldots, p_{2k+1})$ such that p_i is a red edge (res. a blue path) for all even (res. odd) $i \in [2k+1]$. Since p is a path, each p_{2i+1} belongs to a distinct star, for all $i \in (k]$, leading to $k+1$ distinct stars. Further, from Proposition 8.1, we have that each p_{2i+1} can either be of length 1 or 2. Suppose, exactly s blue sub-paths are of length 2, and the rest are all of length 1. Each blue edge in any (i,j)-star incidents on $(i,j,0)$. Therefore, for any sub-paths of length 2, we have to fix just two vertices, and for sub-paths of length 1, we have to fix just one vertex in the star. And, once we have fixed these vertices, the k-line is fixed. Thus, the number of such k-lines is at most $\mu^{k+1}r^{2s}r^{k+1-s}$. Summing over all s, we have

$$\mu^{k+1}\sum_{s=0}^{k+1}\binom{k+1}{s}r^{2s}r^{k+1-s} \leq (\mu r(r+1))^{k+1} \leq (2\mu r^2)^{k+1}. \qquad \square$$

C.3 Proof of Proposition 8.4

If there is no valid labeling then the statement is vacuous. So, suppose there is at least one valid labeling Y. First, suppose $Y^i_j[k] = Y^{i'}_{j'}[k']$ for some $(i,j,k) \neq (i'j',k')$. Then, we must have an edge $e = \{(i,j,k),(i',j',k')\}$, otherwise $Y^i_j[k] \neq Y^{i'}_{j'}[k']$ (by condition 1). Furthermore, suppose $\lambda(e)$ is non-zero. Then, using condition 2, $Y^i_j[k] \oplus Y^{i'}_{j'}[k'] = \lambda(e) \neq 0$. This is impossible by hypothesis. Thus, $\lambda(e) = 0$, which means $X^i_j[k] = X^{i'}_{j'}[k']$. Now, suppose $X^i_j[k] = X^{i'}_{j'}[k']$ for some $(i,j,k) \neq (i'j',k')$. Then, there is an edge $e = \{(i,j,k),(i',j',k')\}$ with label $\lambda(e) = 0$. Using condition 2, $Y^i_j[k] = \lambda(e) \oplus Y^{i'}_{j'}[k'] = Y^{i'}_{j'}[k']$. □

C.4 Proof of Proposition 8.5

The largest component in $\mathcal{G}$ is of course maximally blue, otherwise one can just add the missing blue edges without disconnecting it. Thus, it is sufficient to show the result for an arbitrary maximally blue component. First, using a similar line of arguments as used in the proof of Proposition 8.2, we can establish that $\mathcal{V}(\mathcal{C}) = \bigsqcup_{(i,j)\in\mathcal{I}} \mathcal{V}(\mathcal{S}^i_j)$, where $\mathcal{I} \subseteq \{(i,j) \in [q] \times [c_i]\}$ such that $|\mathcal{I}| = k$ for some $k \geq 1$. Clearly, $k \geq \lceil \xi/(r+1) \rceil$. Now, we use induction to show the existence of a k'-snowflake that spans exactly $k'+1$ stars in $\mathcal{C}$ for all $k' \in (k-1]$. This, in combination of the fact that $k \geq \lceil \xi/(r+1) \rceil$ proves the lemma. For $k' = 0$, any (i,j)-star in $\mathcal{C}$ is a 0-snowflake, and $\lceil \xi/(r+1) \rceil = \lceil r_{i,j} + 1/r + 1 \rceil = 1$. Now, suppose we have a k'-snowflake $\mathcal{S}$ for some $k' < k-1$. We construct a $(k'+1)$-snowflake by connecting $\mathcal{S}$ with a (i',j')-star not in $\mathcal{S}$, but which shares a red edge with $\mathcal{S}$. Note that, at least one such star exists, since $k' < k-1$ and $\mathcal{C}$ is connected. The result follows. □

D Proof of Claim 9.1

We prove the three bounds one by one as follows:

- $\underline{\Pr(\neg\texttt{GSF}_\xi \mid \neg\texttt{ZRO})}$: Proposition 8.5 establishes that, without loss of generality, it is sufficient to bound the probability of existence of an n-snowflake $\mathcal{T}$

in $\mathcal{G}$. Such a $\mathcal{T}$ contains exactly n red edges connecting exactly $(n+1)$ stars, which satisfy the following system of equations:

$$\begin{gathered} X_1^{i_{11}}[k_{11}] = X_1^{i_{12}}[k_{12}] \\ \vdots \\ X_1^{i_{n1}}[k_{n1}] = X_1^{i_{n2}}[k_{n2}], \end{gathered}$$

Since $\mathcal{T}$ is acyclic, the set of indices $i_{j_1 1}, i_{j_1 2}, \ldots, i_{j_l 1}, i_{j_l 2}$ must contain at least $l+1$ distinct elements for any $l \in [n]$. Let $\mathcal{I} = i_{11}, i_{12}, \ldots, i_{n1}, i_{n2}$. We know that each T^j is uniformly distributed and independent of $T^{j'}$ for all $j \neq j' \in \mathcal{I}$. Combining this with the 2-wise independence of H, the probability that the system of equations holds is at most 2^{-n^2}. Using union bound and Proposition 8.3, we get

$$\Pr\left(\neg\mathtt{GSF}_\xi \mid \neg\mathtt{ZRO}\right) \leq \frac{\left(4r^2 \left\lceil \frac{\|\sigma\|+q}{r} \right\rceil\right)^{n+1}}{2^{n^2}} \leq \frac{4r(\|\sigma\|+q+r)}{2^n}, \tag{30}$$

where the last inequality follows from $r(\|\sigma\|+q+r) \leq 2^{n-3}$.

- $\underline{\Pr\left(\neg\mathtt{CF} \mid \neg\mathtt{ZRO} \wedge \mathtt{GSF}_\xi\right)}$: The graph does not contain any component of size greater than $(n+1)(r+1)$. Thus, it is sufficient to bound the probability that $\mathcal{G}$ contains an l-circle for some $l \leq n$. Any such l-circle must satisfy the following system of equations corresponding to the l red edges in the circle:

$$\begin{gathered} X_1^{i_1}[k_{11}] = X_1^{i_2}[k_{22}] \\ \vdots \\ X_1^{i_l}[k_{l1}] = X_1^{i_1}[k_{12}] \end{gathered}$$

When $l = 1$, there must be a red edge between two vertices within a star. Using the 2-wise independence property of H, the probability of this happening is at most 2^{-n}. For $l \geq 2$, we can apply a similar argument as in the previous case, concluding that the first $l-1$ equations hold with probability at most $2^{(1-l)n}$. Conditioned on this event, the final equation holds with probability at most 2^{-n} due to the 2-wise independence of H. Applying the union bound and using Proposition 8.3 over all $l \in [n]$, we get

$$\begin{aligned} \Pr\left(\neg\mathtt{CF} \mid \neg\mathtt{ZRO} \wedge \mathtt{GSF}_\xi\right) &\leq \sum_{l=1}^{n} \left(\frac{2r^2 \left\lceil \frac{\|\sigma\|+q}{r} \right\rceil}{2^n}\right)^l \\ &\leq \sum_{l=1}^{\infty} \left(\frac{2r(\|\sigma\|+q+r)}{2^n}\right)^l \\ &\leq \frac{4r(\|\sigma\|+q+r)}{2^n}, \end{aligned} \tag{31}$$

where we used the fact that $r(\|\sigma\|+q+r) \leq 2^{n-2}$.

- $\underline{\Pr\left(\neg\mathtt{LF} \mid \neg\mathtt{ZRO} \wedge \mathtt{GSF}_\xi\right)}$: This case can be handled in a manner similar to the previous two, noting that X and Z are statistically independent, and $Z^i_j[k]$ is uniformly distributed on $\mathcal{B} \setminus \{0^n\}$. We then obtain:

$$\Pr\left(\neg\mathtt{LF} \mid \neg\mathtt{ZRO} \wedge \mathtt{GSF}_\xi\right) \leq \frac{8r(\|\sigma\| + q + r)}{2^n}. \tag{32}$$

The result then follows from Eq. (30)-(32). □

E Proof of Claim 10.1

Suppose the graph has the said d-interesting path for some fixed $d \in [n]$. Any such d-interesting path must satisfy the following system of equations corresponding to the d red edges in the interesting path:

$$\begin{gathered} X^{i_1}_{j_{11}}[k_{11}] = X^{i_2}_{j_{21}}[k_{22}] \\ \vdots \\ X^{i_d}_{j_{d1}}[k_{d1}] = X^{i_1}_{j_{12}}[k_{12}] \end{gathered}$$

Now, we can have two cases:

- $\mathsf{Case\ A}$: $d = 1$. there must be a red edge between two query-related vertices. Due to the specific algebraic structure of gwc hash, this is only possible if $T^{i_1}[2] = 0^n$. The probability of this happening is at most 2^{-n}, which gives

$$\Pr\left(\neg\mathtt{AUX} \wedge d = 1 \mid \neg\mathtt{ZRO}\right) \leq \frac{q}{2^n}. \tag{33}$$

- $\mathsf{Case\ B}$: $d \geq 2$. For this case we can apply a similar argument as in the proof of 9.1, concluding that the first $d-1$ equations hold with probability at most $2^{(1-d)n}$. Conditioned on this event, the final equation holds with probability at most 2^{-n} due to the 2-wise independence of H. Now, for any such d-interesting path, we have at most $(2r(\|\sigma\|+q+r))^{d-1}$ choices for the intermediate vertices, and for any fixed choice of the intermediate vertices the two query-related endpoints can be chosen in at most $2\|\ell\|(\|\sigma\|+q)$ ways. Applying the union bound over all $d \in [n]$, we get

$$\begin{aligned} \Pr\left(\neg\mathtt{AUX} \wedge d \geq 2 \mid \neg\mathtt{ZRO}\right) &\leq 2\|\ell\| \sum_{d=2}^{n} \left(\frac{2r(\|\sigma\| + q + r)}{2^n}\right)^d \\ &\leq \frac{4r\|\ell\|(\|\sigma\| + q + r)}{2^n} \sum_{d=1}^{\infty} \left(\frac{2r(\|\sigma\| + q + r)}{2^n}\right)^d \\ &\leq \frac{16r^2\|\ell\|(\|\sigma\| + q + r)^2}{2^{2n}}, \end{aligned} \tag{34}$$

where we used the fact that $r(\|\sigma\| + q + r) \leq 2^{n-2}$.

F Proof of $\mathbf{G}^*$ Security Lemma

We extend the notations from Algorithm 5 steps 11–14 to multiple queries. Specifically, we write $X_j^i[k]$ for $(i,j,k) \in \mathcal{V}$, where $\mathcal{V} := [q] \times [\|\ell_i\|] \times (1]$.

Considering the proofs of Theorems 4.1 and 5.1, and by applying the same strategy, it is clear that it suffices to prove that $\mathtt{MTC}_\xi[X,W]$ holds with high probability for $\xi = 2(n+1)$. In particular, we have

$$\begin{aligned}\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{G}^*}(\mathcal{A}) &\leq \Pr\left(\mathtt{ZRO} \vee \neg\mathtt{MTC}_\xi[X,W]\right) \\ &\leq \Pr\left(\mathtt{ZRO}\right) + \Pr\left(\neg\mathtt{MTC}_\xi[X,W] \mid \neg\mathtt{ZRO}\right),\end{aligned}$$

where $\mathtt{ZRO}:\ \exists\,(i,j) \in [q] \times [\|\ell_i\|]$, such that $W_j^i[1] = 0^n$, and the probabilities are computed in the ideal world. Then, using uniformity of $W_j^i[1]$, we have

$$\mathbf{Adv}^{\mathsf{prbg}}_{\mathsf{G}^*}(\mathcal{A}) \leq \frac{\|\sigma\| + q}{2^n} + \Pr\left(\neg\mathtt{MTC}_\xi[X,W] \mid \neg\mathtt{ZRO}\right), \tag{35}$$

and furthermore

$$\begin{aligned}\Pr\left(\neg\mathtt{MTC}_\xi \mid \neg\mathtt{ZRO}\right) \leq \Pr\left(\neg\mathtt{GSF}_\xi \mid \neg\mathtt{ZRO}\right) &+ \Pr\left(\neg\mathtt{CF} \mid \neg\mathtt{ZRO} \wedge \mathtt{GSF}_\xi\right) \\ &+ \Pr\left(\neg\mathtt{LF} \mid \neg\mathtt{ZRO} \wedge \mathtt{GSF}_\xi\right).\end{aligned} \tag{36}$$

The three terms on the r.h.s. can be upper bounded by using a similar strategy as used in [12, Lemma 4.2, 4.3 and 4.4]. In particular, by reusing the notations from [12], we define

- $\mathtt{Fresh}:\ \forall\, i,j \in [q],\ \left(\widehat{T}^i[1], \widehat{T}^i[2]\right) \neq \left(\widehat{T}^{i'}[1], \widehat{T}^{i'}[2]\right)$.
- $\mathtt{Lpairs}:\ \left|\{(i,i')\ :\ i < i' \in [q],\, \widehat{T}^i[1] = \widehat{T}^{i'}[1]\}\right| < \|\sigma\|$.
- $\mathtt{Rpairs}:\ \left|\{(i,i')\ :\ i < i' \in [q],\, \widehat{T}^i[2] = \widehat{T}^{i'}[2]\}\right| < \|\sigma\|$.

where $\widehat{x} := \lfloor x \rfloor_{n-1}$ for any $x \in \{0,1\}^n$. Let $\mathtt{Triv} = \neg(\mathtt{Fresh} \cap \mathtt{Lpairs} \cap \mathtt{Rpairs})$. Then, using the uniformity of T^i and independence of T^i and T^j, we have

$$\begin{aligned}\Pr\left(\mathtt{Triv}\right) &= \Pr\left(\neg(\mathtt{Fresh} \cap \mathtt{Lpairs} \cap \mathtt{Rpairs})\right) \\ &\leq \Pr\left(\neg\mathtt{Fresh}\right) + \Pr\left(\neg\mathtt{Lpairs}\right) + \Pr\left(\neg\mathtt{Rpairs}\right) \\ &\leq \frac{q^2}{2^{2n}} + \frac{2(\|\sigma\| + q)}{2^n}.\end{aligned} \tag{37}$$

Now, we bound the probabilities of the three remaining terms on the r.h.s. of Eq. (37) conditioned on $\neg\mathtt{Triv}$, i.e.

$$\begin{aligned}\Pr\left(\neg\mathtt{MTC}_\xi \mid \neg\mathtt{ZRO}\right) \leq \Pr\left(\mathtt{Triv}\right) &+ \Pr\left(\neg\mathtt{GSF}_\xi \mid \neg\mathtt{ZRO} \wedge \neg\mathtt{Triv}\right) \\ &+ \Pr\left(\neg\mathtt{CF} \mid \neg\mathtt{ZRO} \wedge \neg\mathtt{Triv} \wedge \mathtt{GSF}_\xi\right)\end{aligned}$$

$$+\Pr\left(\neg\texttt{LF}\,|\,\neg\texttt{ZRO}\wedge\neg\texttt{Triv}\wedge\texttt{GSF}_\xi\right). \tag{38}$$

First consider $\Pr\left(\neg\texttt{GSF}_\xi\,|\,\neg\texttt{ZRO}\wedge\neg\texttt{Triv}\right)$. By Proposition 8.5 it is sufficient to bound the probability of existence of an n-snowflake $\mathcal{T}$. Such a snowflake must contain exactly n red edges, which in turn must satisfy the following system of equations:

$$\begin{gathered}X^{i_{11}}_{j_{11}}[k_{11}] = X^{i_{12}}_{j_{12}}[k_{12}]\\ \vdots\\ X^{i_{n1}}_{j_{n1}}[k_{n1}] = X^{i_{n2}}_{j_{n2}}[k_{n2}].\end{gathered}$$

Now, we must have one of the following two cases:

1. *The system has full rank:* Using a similar argumentation as used in the proof of Claim 9.1, we bound the probability to $(8\|\sigma\|+q+2)/2^n$.
2. *The system does not have full rank:* There exists a subsystem of l equations for some $l<n$ such that there exists an arrangement of the equations where the first $l-1$ equations are independent, and the final equation is a consequence of the previous equations. Without loss of generality, we assume that the first l equations satisfy this property. Furthermore, in the last equation, we can also assume $k_{l1}=k_{l2}=1$. Since this last equation is dependent on the previous equations, we must have distinct (i_{a1},j_{a1}) and (i_{b2},j_{b2}) such that

$$\begin{aligned}\widehat{T}^{i_{a1}}[1]\oplus\widehat{T}^{i_{l1}}[1] &= \langle j_{a1}\rangle_{n-1}\oplus\langle j_{l1}\rangle_{n-1},\\ \widehat{T}^{i_{b2}}[1]\oplus\widehat{T}^{i_{l2}}[1] &= \langle j_{b2}\rangle_{n-1}\oplus\langle j_{l2}\rangle_{n-1}.\end{aligned}$$

   Since $\texttt{Lpairs}$ holds we must have at most $(\|\sigma\|+q)^2$ choices for $(i_{a1},j_{a1}),(i_{b2},j_{b2}),(i_{l1},j_{l1}),(i_{l2},j_{l2})$. Then, using the independence of the first $l-1$ equations, the probability in this case is bounded as follows:

$$\sum_{l'=4}^{\infty}\frac{4^{l'-2}(\|\sigma\|+q)^{l'-2}}{2^{(l'-2)n}}\leq\frac{16(\|\sigma\|+q)^2}{2^{2n}},$$

   where the last equation follows from the fact $(\|\sigma\|+q)\leq 2^{n-2}$.

Combining the two cases yields:

$$\Pr\left(\neg\texttt{GSF}_\xi\,|\,\neg\texttt{ZRO}\wedge\neg\texttt{Triv}\right)\leq\frac{8\|\sigma\|+q+2}{2^n}+\frac{16(\|\sigma\|+q)^2}{2^{2n}} \tag{39}$$

Using similar argumentation, we have

$$\Pr\left(\neg\texttt{CF}\,|\,\neg\texttt{ZRO}\wedge\neg\texttt{Triv}\wedge\texttt{GSF}_\xi\right)\leq\frac{16(\|\sigma\|+q)^2}{2^{2n}} \tag{40}$$

$$\Pr\left(\neg\texttt{LF}\,|\,\neg\texttt{ZRO}\wedge\neg\texttt{Triv}\wedge\texttt{GSF}_\xi\right)\leq\frac{16(\|\sigma\|+q)^2}{2^{2n}} \tag{41}$$

Finally, the result follows from Eq. (35)-(41).