Received July 2, 2020, accepted July 13, 2020, date of publication July 16, 2020, date of current version July 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009917

Input-Dependent Error Sketching Model Enabled Information Theoretical Secure Sketch

YEN-LUNG LAI AND ZHE JIN, (Member, IEEE)
Monash University Malaysia, Subang Jaya 47500, Malaysia

Corresponding author: Zhe Jin (jin.zhe@monash.edu)

This research was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (FRGS/1/2018/ICT02/MUSM/03/3).

ABSTRACT Secure sketch conceals any random string w by generating a helper string ss (known as a sketch). It allows the exact recovery of w from ss given another value w' that is close to w. A secure sketch can be utilized to protect any error-prone secret, e.g., biometrics, stored in secret storage to promote secure authentication. When error tolerance is demanded, a secure sketch can be used as an error correction code to tolerate the noise over an unreliable, noisy communication channel. However, when both security and error tolerance are of interest, the error tolerance property of a secure sketch imposes entropy loss. It leads to a weak security guaranty on a low entropy input string. Recent work by Fuller *et al.* (2016) has exploited the structure of the input string. They showed that having precise knowledge over the input strings' distribution is essential to construct a secure sketch for an input string of low entropy. We formalized a new model for secure sketch construction to realize precise knowledge of the input distribution setting. With the formalized new model, we devised an explicit secure sketch construction to a large family of noisy sources. The devised secure sketch can tolerate an error rate close to 1/2 in *polynomial time*, i.e., $O(n^4)$, and meets the best possible secure sketch's security bound with optimal entropy loss.

INDEX TERMS Coding theory, fuzzy extractor, information theory, secure sketch.

I. INTRODUCTION

Traditional cryptography systems rely on uniformly distributed and recoverable random strings for secret. For example, random passwords, tokens, and keys, all are secrets that must be presented precisely on every query for a user to be authenticated and get accessed into the system. Besides, it must also consist of high enough entropy, making it very long and complicated, further resulting in the difficulty in memorizing it. On the other hand, there existed plentiful non-uniform strings to be utilized for secrets in practice. For instance, biometrics (e.g., human iris, fingerprint) can be used for human recognition/identification purposes. Similarly, long passphrase [1], answering several questions for secure access [2] or personal entropy system [3], and list of favorite movies [4], all are non-uniformly distributed random strings that can be utilized for secrets.

The availability of non-uniform information prompted the generation of uniform random string from non-uniform materials. Bennett *et al.*, [5], identified two major steps in deriving a uniform string from noisy non-uniform sources. The first step is *information-reconciliation*, by tolerating the errors in the sources without leaking any information. The second step refers to the *privacy amplification*, which converts high entropy input into a uniformly random input. The information-reconciliation process can be classified into interactive (includes multi messages) and non-interactive (only includes single message) versions. For the non-interactive line of work, it has been firstly defined by Dodis *et al.*, [6] called the fuzzy extractor.

A standard fuzzy extractor works as follow. Given an input string $w \in \{0,1\}^{k^*}$. A randomness extractor algorithm is first used to generate a nearly uniform random string $r \in \{0,1\}^{\ell}$ from w. Then, a secure sketch that consists of a sketching (SS) and a recover (Rec) algorithms is adopted. The sketching algorithm SS accepts input w to generate a helper string, namely sketch, $ss \in \{0,1\}^n$ ($k^* < n$). The sketch ss is necessary to store side information (i.e., add redundancy) for future recovery of w using another random string $w' \in \{0,1\}^{k^*}$. If w' and w are similar, the differences (errors) between them can be tolerated by running the recover algorithm Rec with inputs ss and w' to recover w exactly. Once w is recovered, the same randomness extractor can be used to regenerate r from w.

Secure sketch is defined as information theoretical object. Under the information-theoretical setting, the information

The associate editor coordinating the review of this manuscript and approving it for publication was Junaid Arshad.

stored in ss imposes entropy loss, leads to the extracted random string r may be too short for secret (or key). A goal of fuzzy extractor constructions is to reduce the entropy loss for longer uniform random string r extraction. In view of this, good design of secure sketch construction is therefore needed.

Recently, Fuller *et al.*, [7] incorporated the knowledge over the input structure to reduce the entropy loss in a secure sketch. They defined a new entropy, namely *fuzzy min-entropy*, to measure the security of a secure sketch under distribution-sensitive setting, i.e., the input distribution is precisely known. Specifically, fuzzy min-entropy is defined as the min-entropy of a distribution W with maximized chances to look for a sample $w \in W$ within distance t of other random sample w' follows.

$$\mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) \stackrel{\mathrm{def}}{=} -\log\left(\max_{w'} \Pr\left[W \in B_t(w')\right]\right),$$

where $B_t(w')$ denotes a hamming ball of radius t around w'.

Super-logarithm fuzzy min-entropy, i.e., $\mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) = \omega(\log(n))$ is necessary and sufficient for security measurement of a secure sketch given the input distribution W is precisely known and only computational security is of interested (e.g., holds for computationally bounded adversaries). Most importantly, Fuller *et al.* also showed that even for a secure sketch whose defined information theoretically with security bounded in term of average min-entropy (a.k.a. conditioned min-entropy $\tilde{\mathrm{H}}_\infty(W \mid ss)$ given a sketch ss), the best one can hope for recovering w with error at most $\gamma > 0$ is

$$\tilde{\mathrm{H}}_\infty(W \mid ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1-\gamma). \tag{1}$$

They gave an explicit construction called layered hashing [7] that achieved close to the optimal result with average min-entropy

$$\tilde{\mathrm{H}}_\infty(W \mid ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1/\gamma) - \log(H_0(W)) - 1,$$

where $H_0(W)$ denotes the Hartley entropy, which is the logarithm of the supports of W, i.e., $\log(|\mathsf{supp}|(W))$. In short, layered hashing incorporates universal hashing to disambiguate the points over a known distribution, under the same layer, to achieve error tolerance. Every point under the same layer possess the same probability to be utilized for sketching.

Woodage *et al.* [8] proposed an improvement over layered hashing construction by introducing a brute force checker to enumerate the probable points and check their correctness during recovery. In particular, they proposed to hide the layer information with appropriate number of random bits padding and yielded a saving of $\log(H_0(W))$ bits entropy. This improved the security bound of a secure sketch to

$$\tilde{\mathrm{H}}_\infty(W \mid ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1/\gamma) - 1. \tag{2}$$

Open question: an important open question retained is whether one can build a secure sketch that replaces $\log(1/\gamma)$ with the optimal $\log(1-\gamma)$ depicted in Eq. 1.

A. MOTIVATIONS AND CONTRIBUTIONS

The main motivation of this paper is to tackle the open question in constructing a secure sketch that achieves optimal entropy loss follows Eq. 1. It is a non-trivial task to build a secure sketch with optimal entropy loss. Fuller *et al.*, [9] identified that there is a family of noisy sources described using a family of distributions. Despite individual members of distributions in this family consists of super-logarithm $\omega(\log(n))$ fuzzy min-entropy, no fuzzy extractor can provide meaningful security to *all* of the members in this family. This result implies that the security property of a secure sketch must hold only under the case when the fuzzy min-entropy is equal to the min-entropy of the sources. In light of this, one must be allowed to design the secure sketch as a function of the input distribution where the precise knowledge over the input distribution is believed to be necessary. Motivated by the above reasons, we present IDES-model, that encapsulates the precise knowledge over the input distribution setting for secure sketch construction. Loosely speaking, the IDES-model allows the realization of precise knowledge on arbitrary random distributions. The precise knowledge on the distribution setting can be realized by using a recovery algorithm (our devised IDES-recovery algorithm) that acts as a distribution distinguisher that can reveal all the possible readings $w \in W$ lied under an unknown distribution W. Detail formalization of IDES-model is given in Section V.

1) OUR CONTRIBUTIONS

The open question in constructing a secure sketch outlined above primarily motivates us to construct a secure sketch for achieving optimal entropy loss follows Eq. 1. We summarized the main two contributions of this work as follows.

New Efficient Model for Secure Sketch: We formalized a new model, namely input-dependent error sketching (IDES) model, for secure sketch. With the formalized IDES-model, we devised a pair of efficient sketching and recovery algorithms, operating in $O(n^4)$, for large family of enumerable distributions $\mathcal{W}$. The devised IDES-secure sketch can achieve arbitrary (on average-cases) minimum entropy loss at least $k-n^*$, which are the parameters to be set during the sketching phase. Most importantly, we showed that the proposed secure sketch enjoys the optimal upper bound of entropy loss follows Eq. 1 (see Proposition 2) that responds to our main motivation early.

Efficient Error Correction Code With Optimal Parameter: Benefited from the correctness property of the devised IDES-secure sketch, it offers efficient error correction to any random string w' that is close to the input w in $O(n^4)$ operations. We showed that with IDES-secure sketch, we could efficiently derive arbitrary $[n^*, k-n^*, d]_2$ binary error correction code $\mathcal{C}$ that works as asymptotic good error correction code. Most importantly, we showed that $\mathcal{C}$ is capable of attaining the optimal trade-off in between the code rate $R = 1-(k-n^*)n^*$ and the relative distance $(d-1)/n^*$ on the average selection of the parameter $k-n^* \geq 1$. The details discussion can be found in Section VII.

2) PAPER ORGANIZATION

The rest of the paper is organized as follow. We review the existing secure sketch constructions in Section II. The preliminaries and background of secure sketch is given in Section III. Section IV introduces the resilient vector used in our construction. Section V details our formalization of new IDES-model. Then, we propose a pair of IDES-sketching and IDES-recovery algorithm based on IDES-model in Section VI. A brief discussion on the application of IDES-secure sketch for error correction is given in Section VII. In Section VIII, we benchmark our proposed method with the existing works in the security bound, which serves the comparison on the security parameter between our method and the existing constructions. Lastly, a conclusive remark of this work in given in Section IX

II. RELATED WORKS

In this section, we review the existing secure sketch constructions. Various secure sketch constructions can be found in the literature. Some notable works involved the code-offset construction proposed by Juels and Wattenberg [10] that operates perfectly over hamming matric space. Besides, Juels and Sudan [4] proposed another construction for set metric space called the fuzzy vault. Then, Dodis *et al*., [6], offered an improved version of the fuzzy vault (with strengthened parameters). At the same time, they also introduced the Pin-sketch construction for non-fixed length input over a universe $\mathcal{U}$.

There are a number of implementations of secure sketch for various noisy sources. For instance, secure sketch have been implemented to protect human fingerprint and dynamic signature [11], [12]–[15], for user privacy preserving in body sensor network and tele-health system [16], [17]. Secure sketch also protect human-gait feature [18] and promote secure authentication between multi parties [19], [20]. Most of these implementations used fuzzy vault due to its unordered property in accepting any unordered sets as input, which is common for biometric templates. To ensure correctness, tolerating large number of errors in the input feature imposes high entropy loss, further leads to low attack complexity [21]–[23], and no meaningful security can be claimed.

The tension between security and error tolerance capability in a secure sketch is very strong. Given some non-uniform sources with low entropy, especially, when the sources consist of *more error than entropy* itself, deducting the entropy loss from the sources' entropy always outputs a negative value; hence, show no security. Thus, tolerating a large number of errors would need to assume sufficient high entropy to the input sources. It is reported that the min-entropy of the sources must at least half of the input length itself [24].

For clearer illustration on the more error than entropy sources, we consider an example of noisy source, i.e, the human iris. Human iris is believed to offer ≈ 249 bits of entropy [25]. However, most of the iriscode generated from human iris requires about 10% - 30% error tolerance of the human iriscode [26], [27], which is approximately 400 bits of errors. We can observe that the errors (400 bits) in an iriscode are typically more than the entropy (249 bits). Hence, tolerating all the error in an iriscode gives us no security [28]. In the light of this, the traditional way of security analysis based on merely the min-entropy of the sources and the number of the error to be tolerated is insufficient for when the sources have more error than entropy. The fuzzy min-entropy measurement, which takes consideration over the input sources' structure, is viewed as a necessity.

III. PRELIMINARIES

There are some preliminaries to introduce the background of a secure sketch, entropy, and error correction code.

For two random variables X and Y over some metric space (i.e., $\mathcal{M} = \{0, 1\}$). The min-entropy of X is $\mathrm{H}_\infty(X) = -\log(\max_x \Pr[X = x])$. The average min-entropy of X given Y is $\tilde{\mathrm{H}}_\infty(X \mid Y) = -\log(\mathbb{E}_{y\in Y} \max_x \Pr[X = x \mid Y = y])$.

Error Correction Code [29]: Let $q \geq 2$ be an integer, let $[q] = \{1, \ldots, q\}$, we called an $[n, k, d]_q$-ary code $\mathcal{C}$ consists of following properties:

- $\mathcal{C}$ is a subset of $[q]^n$, where n is an integer referring to the *block length* of $\mathcal{C}$.
- The *dimension* of code $\mathcal{C}$ can be represented as $|\mathcal{C}| = [q]^k = V$
- The *rate* of code $\mathcal{C}$ is the normalized quantity $\frac{k}{n}$
- The *min-distance* between different codewords defined as $d = \min_{c,c^*\in\mathcal{C}} \mathsf{dis}(c, c^*)$

It is convenient to view code $\mathcal{C}$ as a function $\mathcal{C} : [q]^k \to [q]^n$. An element of V can be considered as a message $v \in V$ and the process to generate its associated codeword $\mathcal{C}(v) = c$ is called *encoding*. Encoding a message v of length k, always adds redundancy to produce codeword $c \in [q]^n$ of longer length n. Nevertheless, for any codeword c with at most $t = \lfloor \frac{d-1}{2} \rfloor$ symbols is modified to form c', it is possible to uniquely recover c from c' by using certain function f s.t. $\mathsf{f}(c') = c$. The procedure to find the unique $c \in \mathcal{C}$ that satisfied $\mathsf{dis}(c, c') \leq t$ by using f is called as *decoding*. A code $\mathcal{C}$ is efficient if there exists a polynomial time algorithm for encoding and decoding.

Linear Error Correction Code [29]: Linear error correction code is a linear subspace of a field $\mathbb{F}_q^n$. A q-ary linear code of block length n, dimension k and minimum distance d is represented as $[n, k, d]_q$ code $\mathcal{C}$. For a linear code, a string with all zeros 0^n is always a codeword. It can be specified into one of two equivalent ways with a generator matrix $G \in \mathbb{F}_q^{n\times k}$ or parity check matrix $H \in \mathbb{F}_q^{(n-k)\times n}$:

- a $[n, k, d]_q$ linear code $\mathcal{C}$ can be specified as the set $\{Gv : v \in \mathbb{F}_q^k\}$ for an $n \times k$ metric which is known as the *generator matrix* of $\mathcal{C}$.
- a $[n, k, d]_q$ linear code $\mathcal{C}$ can also be specified as the subspace $\{x : x \in \mathbb{F}_q^n$ and $Hx = 0^{n-k}\}$ for an $(n-k) \times n$ metric which known as the *parity check matrix* of $\mathcal{C}$.

The conversion from G to H and its reverse can be completed with $O(n^3)$ operations. For any linear code, the linear combination of any codewords is also considered as a codeword

over $\mathbb{F}_q^n$. Often, the encoding of any message $v \in \mathbb{F}_q^k$ can be completed with $O(n^2)$ operations (by multiplying it with the generator matrix, i.e., Gv). The distance between two linear codewords refers to the number of disagreed elements between them, also known as the *hamming distance*. Sometime, we refer $[n, k, d]$ code $\mathcal{C}$ as $[n, k, t]$ code $\mathcal{C}$ if the error tolerance distance t is of interested rather than its minimum distance d.

Random Linear Code (Binary) [29]: Denote $h_2(x) = x\log_2(x) - (1-x)\log_2(1-x)$. The following theorem is a standard in coding theory for the existence of random linear code

Theorem 1: For $\varepsilon \in (0, 1/2)$, i.e., $0 < \varepsilon < 1 - h_2(\varepsilon)$ and large enough n, the following holds for $k = \lceil(1 - h_2(\varepsilon) - \varepsilon)n\rceil$. If $G \in \mathbb{F}_2^{n\times k}$ is drawn uniformly at random, then the linear code with G as a generator matrix has rate at least $(1 - h_2(\varepsilon) - \varepsilon)n$ and relative distance at least ε with probability at least $1 - \exp(-\Omega(n))$

Secure Sketch: [6] An $(\mathcal{M}, m, \tilde{m}, t)$-secure sketch is a pair of randomized procedures "sketch" (SS) and "Recover" (Rec), with the following properties:

- SS: takes input $w \in \mathcal{M}$ returns a secure sketch (e.g., helper string) $ss \in \{0, 1\}^*$.
- Rec: takes an element $w' \in \mathcal{M}$ and ss. If $\mathsf{dis}(w, w') \le t$ for some tolerance threshold t, then $\mathsf{Rec}(w', ss) = w$ with probability $1 - \gamma$, where γ is a negligible quantity. If $\mathsf{dis}(w, w') > t$, then no guarantee is provided about the output of Rec.

The ***correctness*** property of secure sketch ensures the successful recovery of w given ss and w' with probability $1 - \gamma$ for all $w \in \mathcal{M}$, $\mathsf{dis}(w, w') \le t$. The ***security*** property of secure sketch guarantees that for any distribution W over $\mathcal{M}$ with min-entropy m, the values of W can be recovered by the adversary who observes ss with probability no greater than $2^{-\tilde{m}}$. That is the average min-entropy (or conditioned min-entropy) $\mathrm{H}_\infty(W \mid ss) \ge \tilde{m}$.

IV. RESILIENT VECTOR: PROPERTIES AND GENERATION

The usa of the resilient vector (RV) into cryptography is first introduced by Rivest [30] in 2016. Its main concept is derived from Locality Sensitive Hashing (LSH) defined below.

Locality Sensitive Hashing [31] Given that $P_2 > P_1$, while $w, w' \in \mathcal{M}$, and $\mathcal{H} = h_i : \mathcal{M} \to U$, where U refers to the output metric space (after hashing), which comes along with a similarity function S, where i is the number of hash functions h_i. A locality sensitive hashing can be viewed as a probability distribution over a family $\mathcal{H}$ of hash functions follows $P_{h\in\mathcal{H}}[h(w) = h(w')] = S(w, w')$. In particular, the similarity function S described the hashed collision probability in between w and w'.

$$P_{h\in\mathcal{H}}(h_i(w) = h_i(w')) \le P_1, \quad \text{if } S(w, w') < R_1$$
$$P_{h\in\mathcal{H}}(h_i(w) = h_i(w')) \ge P_2, \quad \text{if } S(w, w') > R_2$$

LSH transforms input w and w' to its output metric space U with property that ensuring similar inputs render higher probability of collision over U, and vice versa.

For RV generation, we only focus on a particular LSH family called hamming-hash [32]. The hamming hash is considered one of the easiest ways to construct an LSH family by bit sampling technique.

We denote the hamming-hash algorithm as $\Omega : \{0, 1\}^{k^*} \times [k^*]^n \to \{0, 1\}^n$, which serves to sample the input binary string of length k^* into a longer binary string a.k.a resilient vector of length $n > k^*$.

Given input $w \in \{0, 1\}^{k^*}$, and $N \in [k^*]^n$, algorithm Ω can be described as follow (where || denotes concatenation):

$\Omega(w, N)$

1: $\phi \leftarrow \emptyset$
2: **for** $i = 1,\ldots,n$ **do**
3: **parse** $x = w(N(i))/x$ is the $N(i)$-th bits of w
4: $\phi = \phi \| x$
5: **endfor**
6: **return** ϕ

Theorem 2: Denote $\xi = \|w \oplus w'\| (k^*)^{-1}$. Given w, w' and the RV generated follow $\phi \leftarrow \Omega(w, N)$, and $\phi' \leftarrow \Omega(w', N)$ (with same N). Then $\mathbb{E}\big[\|\phi \oplus \phi'\|\big] = n\xi$.

Proof: By LSH definition (for $j \in \{1, \ldots, n\}$):

$$\Pr\big[\phi(j) \ne \phi'(j)\big] = \|w \oplus w'\| (k^*)^{-1} = \xi.$$

Hence, the RV pairs (ϕ, ϕ') follow i.i.d. (binomial distribution) with expected number of different position (hamming distance) equal to $\mathbb{E}\big[\|\phi \oplus \phi'\|\big] = n\xi$. □

V. IDES-MODEL

The preliminaries and notations used in IDES-model formalization are given as follows. Our focus is on the binary matrix space.

Preliminaries and Notations: Let $\mathcal{M}_1 = \{0, 1\}^{k^*}$, and $\mathcal{M}_2 = \{0, 1\}^n$ denote two different sizes of matrix spaces where $n > k^*$. Any two readings $w \in W$ and $w' \in W'$ are under some random distributions W and W' respectively over $\mathcal{M}_1$. The distance between different binary strings w and w' denoted as $\|w \oplus w'\|$ is the binary hamming distance (e.g., the number of disagreed elements) where $\|.\|$ is the hamming weight operation that counts the number of non-zero elements, and $\oplus$ is the addition modulo two operation (XOR).

In IDES-model, we consider the family of distributions $\mathcal{W} \in \mathcal{M}_1$ and $\Psi \in \mathcal{M}_2$ of same size $|\Psi| = |\mathcal{W}| = |\mathsf{supp}(\mathcal{E}_{ss})|$ that depend upon the error distribution $\mathcal{E}_{ss} \in \mathcal{M}_1$ introduced in the sketching phase. In the recovery phase, we consider another two family of distributions $\mathcal{W}' \in \mathcal{M}_1$ and $\Psi' \in \mathcal{M}_2$ of same size $|\Psi'| = |\mathcal{W}'| = |\mathsf{supp}(\mathcal{E}_{rec})|$ depend upon another error distribution $\mathcal{E}_{rec} \in \mathcal{M}_1$. We denote the members of the families as $W_e \in \mathcal{W}$ and $W'_{e'} \in \mathcal{W}'$,

which are essentially two random distributions of some noisy strings $w_e \in W_e$ and $w'_{e'} \in W'_{e'}$ respectively. w_e and $w'_{e'}$ are generated follow $w_e = w \oplus e$ and $w'_{e'} = w' \oplus e'$, using different error readings $e \in \mathcal{E}_{ss}$ and $e' \in \mathcal{E}_{rec}$ sampled from the error distributions $\mathcal{E}_{ss}$ and $\mathcal{E}_{rec}$ respectively. Let $\Phi_e \in \Psi$ and $\Phi'_{e'} \in \Psi'$ be two random RV distributions, viewed as the members of Ψ and Ψ' respectively over $\mathcal{M}_2$. The RV generations used inputs w_e and $w'_{e'}$ are denoted as $\phi \leftarrow \Omega(w, N)$ and $\phi'_{e'} \leftarrow \Omega(w'_{e'}, N)$ (with same random string N) where $\phi_e \in \Phi_e$ and $\phi'_{e'} \in \Phi'_{e'}$.

Denote $B_{t^*}(w'_{e'})$ as hamming ball of radius t^* centred on $w'_{e'}$. We generally express $W_e \in B_{t^*}(w'_{e'})$, which means $\forall w_e \in W_e, \left\|w_e \oplus w'_{e'}\right\| \leq t^*$. Same notation also used for the RV distribution s.t. $\Phi_e \in B_t(\phi'_{e'})$, which means $\forall \phi_e \in \Phi_e, \left\|\phi_e \oplus \phi'_{e'}\right\| \leq t$.

To achieve error tolerance, we used two random linear codes, denoted as $[n^*, k^*, t^*]_2$- $\mathcal{C}_{in}$ and $[n, k, t]_2$- $\mathcal{C}_{out}$. More precisely, we called $\mathcal{C}_{in}$ the 'inner code' that accepts any input string w of length k^* to output a codeword c^* of length n^*, i.e., $\mathcal{C}_{in}w = c^*$. On the other hand, we called $\mathcal{C}_{out}$ the 'outer code' that accepts any input string y of length k to output a codeword c of length n, i.e., $\mathcal{C}_{out}y = c$. $\mathcal{C}_{in}$ and $\mathcal{C}_{out}$ consist of error tolerance distance $t^* \geq 0$ and $t \geq 0$ respectively. The encoding processes is done by multiplying the generator matrix of the given code with the input string. The decoding is simply a Gaussian elimination process. All logarithm without a base are considered base 2, that is, $\log(x) = \log_2(x)$.

A. WHY IDES-MODEL: A DISTRIBUTION DUSTINGUISHER IS NECESSARY

Intuitively, the IDES-model models the sketching and recovery process by using a random error adding into the input strings $w \in W$ and $w' \in W'$. All the claims for IDES-model can be found in the Appendix Section. We start by giving our inspiration for proposing the IDES-model described as follows.

A secure sketch should work for any random inputs $w \in \mathcal{M}_1$ under any random distribution W (over $\mathcal{M}_1$). Given any random string $w' \in W'$, the correctness property of a secure sketch (see Section III) must hold for all $w \in W$ follows $\left\|w \oplus w'\right\| \leq t^*$. This can easily achieved if all $w \in W$ that are close to w' within distance t^* have stored in prior using a sketching algorithm. In such a case, precise knowledge on the distribution W s.t. $\forall w \in W, \left\|w \oplus w'\right\| \leq t^*$ is necessary. Since $W \in B_{t^*}(w')$, gaining precise knowledge on $B_{t^*}(w')$ straight away implies a precise knowledge on W.

To obtain precise knowledge on $B_{t^*}(w')$, we need an algorithm that able to reveal all the points in $B_{t^*}(w')$. We formally call such algorithm as a *distribution distinguisher* which is necessary to distinguish $B_{t^*}(w')$ given a random string $w' \in W'$. It is convenient to describe any $w \in W \in B_{t^*}(w')$ as a point. The process of revealing all the points in $B_{t^*}(w')$ given w' can be formally described as *distinguish* $B_{t^*}(w')$. The following claim is given to conclude that a distribution distinguisher is necessary to achieve precise knowledge on any random input distribution W and show correctness of a secure sketch.

Claim 1: Given a random string $w' \in W'$. Given a hamming ball $B_{t^*}(w')$ consists of 2^m points. The distinguishability of $B_{t^*}(w')$, i.e., $B_{t^*}(w')$ can be distinguished, implies the correctness of a secure sketch, holds for any random $W \in B_{t^*}(w')$ and all $w \in W$.

B. MODEL FORMALISATION

In this subsection, we give the details on the way we formalize IDES-model. It can be categorized into three major parts, which are the

1) Realization of enumerable input-dependent family of distributions
2) Distingusher for large hamming ball
3) Definition and measurement of the distinguishability, conditioned on correctness, with bounded errors.

1) REALIZATION OF ENUMERABLE INPUT-DEPENDENT FAMILY OF DISTRIBUTIONS

We want our model to accept any $w \in W$ in arbitrary random distribution $W \in \mathcal{M}_1$.

However, it is impossible to precisely model and distinguishes every $B_{t^*}(w')$ in $\mathcal{M}_1$ with an arbitrary number of points. Especially for $B_{t^*}(w')$ that consists of an exponentially large number of points described as 2^{k^*}, a large value of input length k^* would lead to a lifetime of distinguishing process. Therefore, to ensure the meaningful distinguishing result, we only interested in the efficiently computable distributions.

In light of this, we put our focus over an *input-dependent* family of distributions $\mathcal{W}$ which is enumerable given some random error parameter $\varepsilon_{ss} > 0$. In particular, we define such enumerable family of distributions consists of sizes that is super-polynomial with referred to some polynomial function $p(n)$ in the input sketch size n, i.e., $|\mathcal{W}| = \omega(p(n))$.

The input dependent family distributions $\mathcal{W}$ can be realized by using any random string $w \in W \in B_{t^*}(w')$ with an error vector $e \in \mathcal{E}_{ss}$ sampled uniformly at random from a list $\mathcal{E}_{ss} = \{e_{(1)}, \ldots, e_{(|\mathsf{supp}(\mathcal{E}_{ss})|)}\}$. Such a list, without loss of generality, can described in lexicographical order where $\forall e \in \mathcal{E}_{ss}, \|e\| = \lceil k^* \varepsilon_{ss} \rceil$. More precisely, for any w, the error vector e is added with w, forming a noisy string $w_e = w \oplus e$. There are exactly $|\mathsf{supp}(\mathcal{E}_{ss})|$ possible ways to form w_e where $w_e \in W_e$ is distribute in some random distribution W_e depending on w and e. Formally, $w_e \in \mathcal{W} = \{W_{e(1)}, \ldots, W_{e(|\mathsf{supp}(\mathcal{E}_{ss})|)}\}$. The way of forming w_e can be interpreted as a random sampling process to sample a random distribution W_e uniformly at random from $\mathcal{W}$.

The following claim is given to conclude that distinguish $B_{t^*}(w')$ is sufficient to achieve precise knowledge on any random distribution $W_e \in B_{t^*}(w')$ and show correctness of a secure sketch to all $w \in W \in B_{t^*}(e \oplus w')$.

Claim 2: Given a random string $w' \in W'$. If $B_{t^*}(w')$ can be distinguished, then $B_{t^*}(e \oplus w')$ can be distinguished and show correctness of a secure sketch to all $w \in W \in B_{t^*}(e \oplus w')$.

2) DISTINGUISHER FOR LARGE HAMMING BALL IS NECESSARY AND SUFFICIENT

Recall our Claim 2, distinguish $B_{t^*}(w')$ is sufficient to show correctness of a secure sketch. Nonetheless, above result only shows correctness to arbitrary random distribution $W \in B_{t^*}(e \oplus w')$ of size (number of points) that is bounded by the volume of the hamming ball $B_{t^*}(.)$, depending on the radius $t^* \leq k^*$.

Additionally, the error vector e can be conveniently viewed as any random error naturally existing in an input source, which is *unknown* in prior. For large e, the distance between the points (w_e) in any random distribution W_e will be large, leads to many points in W_e cannot be found within $B_{t^*}(w')$. Hence, the correctness of a secure sketch cannot show to all these missing points over $B_{t^*}(w')$. Therefore, it is desirable to consider arbitrary random value for e, especially when e is large, to construct a distinguisher to distinguish $B_{t^*}(w')$ of larger radius. Indeed, we want to distinguish a larger hamming ball $B_t(.)$ of radius $t > t^*$ that is not limited by the input length k^* to show correctness over a larger family of random distribution. For this reason, we adopted RV usage to realize a distinguisher for $B_t(.)$ over larger metric space $\mathcal{M}_2$, which is not bounded by the input length k^*.

Since there are exactly $|\mathsf{supp}(\mathcal{E}_{ss})|$ possible ways to form w_e, it means there are exactly $|\mathsf{supp}(\mathcal{E}_{ss})|$ possible RV can be generated from w_e given N. Any RV generated follows $\phi_e \leftarrow (w_e, N)$ is distribute in a random distribution Φ_e over a family of RV distributions Ψ. Formally, $\phi_e \in \Psi = \{\Phi_{e(1)}, \ldots, \Phi_{e(|\mathsf{supp}(\mathcal{E}_{ss})|)}\}$. Noting that $|\Psi| = |\mathcal{W}| = |\mathsf{supp}(\mathcal{E}_{ss})| = \omega(p(n))$ and any possible RV in distribution Φ_e shall correlated to W_e by the same error vector $e \in \mathcal{E}_{ss}$.

Given another random string $w' \in W'$ (over $\mathcal{M}_1$) and the same N, its corresponding RV can be generated follows $\phi' \leftarrow \Omega(w', N)$. Distinguish $B_t(\phi')$ is necessary to reveal any random RV distribution $\Phi_e \in B_t(\phi')$. The following Claim 3 is given to show that distinguish $B_t(\phi')$ is necessary and sufficient to distinguish $B_{t^*}(e \oplus w')$ for arbitrary large value of e, and show correctness of a secure sketch to all $w \in W \in B_{t^*}(e \oplus w')$ over $\mathcal{M}_1$ and all $\phi_e \in \Phi_e \in B_t(\phi')$ over $\mathcal{M}_2$.

Claim 3: For any $t \geq t^*$. Given a random RV $\phi' \in \Phi'$ and N. If $B_t(\phi')$ can be distinguished, then $B_{t^*}(e \oplus w')$ can be distinguished and show correctness of a secure sketch to all $w \in W \in B_{t^*}(e \oplus w')$ over $\mathcal{M}_1$ and all $\phi_e \in \Phi_e \in B_t(\phi')$ over $\mathcal{M}_2$.

Noting that any random RV distribution Φ_e corresponding to the error vector e. Without the knowledge on e, no guaranty that $B_t(\phi')$ can be distinguished. Hence, it is appropriate to sample another enumerable input-dependent family of distribution, say $\mathcal{W}'$, which can be used to realize another family of RV distribution Ψ' and show correlation in between the random RV distributions $\Phi_e \in \Psi$ and $\Phi'_{e'} \in \Psi'$. Obviously, Φ_e and $\Phi'_{e'}$ are correlated if their corresponding error vector are the same, i.e., $e = e'$. More precisely, given $w' \in W'$, we define another error parameter $\varepsilon_{rec} > 0$ which is necessary for the formation of another noisy string $w'_{e'} = w' \oplus e'$. The error vector $e' \in \mathcal{E}'$ is chosen from the list $\mathcal{E}'_{rec} = \{e'_{(1)}, \ldots, e'_{(|\mathsf{supp}(\mathcal{E}_{rec})|)}\}$ where $\forall e' \in \mathcal{E}_{rec}$, $\|e'\| = \lceil k^* \varepsilon_{rec} \rceil$. Then, any RV generated follows $\phi'_{e'} \leftarrow (w'_{e'}, N)$ shall distribute randomly in $\Phi'_{e'} \in \Psi'$ and $|\mathcal{W}'| = |\Psi'| = |\mathsf{supp}(\mathcal{E}_{rec})|$. Based on the above setting, we give the following claim to conclude that large enough value of ε_{rec} allows $B_t(\phi'_{e'})$ to be distinguished in $O(p(n))$ steps and show correctness of a secure sketch to all $w_e = w \in W \in B_{t^*}(w')$ over $\mathcal{M}_1$ and all $\phi_e \in \Phi_e \in B_t(\phi'_{e'})$ over $\mathcal{M}_2$.

Claim 4: Let $p(n) = 2^{\lceil k^* h_2(\varepsilon_{rec}) \rceil} - 1$ with any error parameter $\varepsilon_{ss} \in (0, 1/2]$, $\varepsilon_{rec} \in [1/k^*, 1/2]$. For $B_{t^*}(e \oplus w'_{e'})$ consists of $p(n)$ number of points. With large enough n s.t. $\varepsilon_{rec} \geq \varepsilon_{ss} > 0$, $B_t(\phi'_{e'})$ can be distinguished in $O(p(n))$ steps if and only if $B_{t^*}(e \oplus w'_{e'})$ can be distinguished in $O(p(n))$ steps. In particular, with large enough n, distinguish $B_{t^*}(e \oplus w'_{e'})$ is equivalent to distinguish $B_{t^*}(w')$ to show correctness of a secure sketch to all $w_e = w \in W \in B_{t^*}(e \oplus w'_{e'})$ over $\mathcal{M}_1$ and all $\phi_e \in \Phi_e \in B_t(\phi'_{e'})$ over $\mathcal{M}_2$.

3) DISTINGUISHABILITY WITH BOUNDED ERRORS CONDITIONED ON CORRECTNESS

Follow our result on Claim 4, we measure the error of distinguish $B_t(\phi'_{e'})$ conditioned on the case when the correctness holds for any random distribution W follows $W \in B_{t^*}(e \oplus w'_{e'})$ (over $\mathcal{M}_1$). The following definition formally described the procedure in distinguish $B_t(\phi'_{e'})$.

Definition 3: Given any random distributions W, $W' \in \mathcal{M}_1$. For all $w \in W$ and $w' \in W'$, consider the following procedure with $\varepsilon_{ss} \in (0, 1/2]$, $\varepsilon_{rec} \in [1/k^*, 1/2]$,

1) *Sample uniformly at random an error vector e where $\|e\| = \lceil k^* \varepsilon_{ss} \rceil$*
2) *Sample an error vector e' where $\|e'\| = \lceil k^* \varepsilon_{rec} \rceil$*
3) *compute $w_e = w \oplus e$*
4) *compute $w'_{e'} = w' \oplus e'$*
5) *Generate the RVs pair $(\phi_e, \phi'_{e'})$ follow $\phi_e \leftarrow \Omega(w_e, N)$ and $\phi'_{e'} \leftarrow \Omega(w'_{e'}, N)$. Then, compute their distance $\|\phi_e \oplus \phi'_{e'}\|$*

Given two tolerance distances $t^* \geq 0$, and $t_{\min} > 0$. For all $\phi_e \in \Phi_e$ and $\phi'_{e'} \in \Phi'_{e'}$ in some random RV distributions Φ_e and $\Phi'_{e'}$ (over $\mathcal{M}_2$), we say $B_t(\phi'_{e'})$ is $(t_{\min}, t^*, \beta)$-distinguishable with error at most $\beta > 0$ if for all $t \geq t_{\min}$:

$$\mathbb{E}_{w' \leftarrow W'}\left[\Pr\left[\Phi_e \notin B_t(\phi'_{e'}) \;\middle|\; W \in B_{t^*}(e \oplus w'_{e'})\right]\right] \leq \beta.$$

The following lemma is given to show that the distinguishability error is at most $\beta = \exp(-2n\varepsilon_d^2)$ with any parameter $t^* \geq 0$ and any $t \geq t_{\min}$. Its proof is given in the Appendix Section.

Lemma 1: Denote $\varepsilon_d = \|e \oplus e'\|(k^*)^{-1}$, $\xi' = \|w_e \oplus w'_{e'}\|(k^*)^{-1}$, and $\xi = \|w \oplus w'\|(k^*)^{-1}$. For $t_{\min} = n(\xi' + \varepsilon_d)$ and $t^* = (\xi - \varepsilon_d)k^*$. It follows that $B_t(\phi'_{e'})$ is $(t_{\min}, t^*, \beta)$-distinguishable with error at most $\beta = \exp(-2n\varepsilon_d^2)$.

VI. IDES-SECURE SKETCH

With the formalized IDES-model, we proposed an explicit construction of secure sketch, namely IDES-secure sketch, for any inputs strings $w \in \mathcal{M}_1$ and $w' \in \mathcal{M}_1$. A graphical

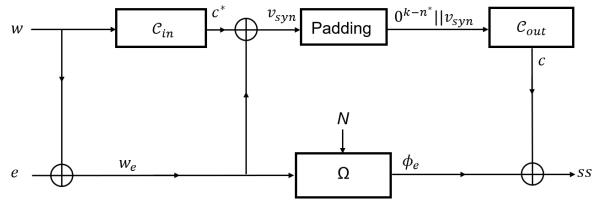

FIGURE 1. IDES-sketching.

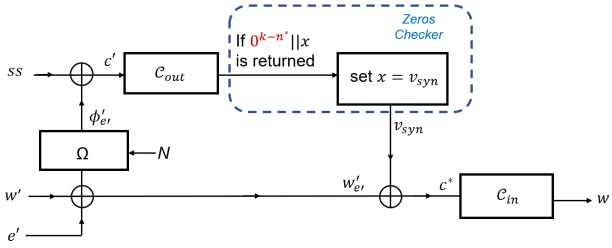

FIGURE 2. IDES-recovery.

illustration of IDES-sketching and IDES-recovery algorithms are presented as Figure 1 and Figure 2.

Given two random generator matrices $G_{in} \in \mathbb{F}_2^{n^* \times k^*}$ and $G_{out} \in \mathbb{F}_2^{n \times k}$ for $\mathcal{C}_{in}$ and $\mathcal{C}_{out}$ respectively. The IDES-sketching algorithm ($\mathsf{SS}_{\mathsf{IDES}}$) with random inputs w, N, and ε_{ss} is described as follow:

IDES-Sketching Algorithm: $\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss})$

Step 1: Sample e uniformly at random, where $\|e\| = \lceil k^* \varepsilon_{ss} \rceil$
Step 2: Compute $w_e = w \oplus e$
Step 3: Compute a codeword $c^* = G_{in} w$
Step 4: Generate $v_{syn} = c^* \oplus w_e$.
Step 5: Pad $k - n^*$ zeros in front v_{syn}, i.e., $0^{k-n^*}||v_{syn}$
Step 6: Compute a codeword $c = G_{out}(0^{k-n^*}||v_{syn})$
Step 7: Generate the RV $\phi_e \leftarrow \Omega(w_e, N)$
Step 8: Generate $ss = c \oplus \phi_e$.
Step 9: Output the public data as (ss, N)

Referring to the same generator matrices $G_{in} \in \mathbb{F}_2^{n^* \times k^*}$ and $G_{out} \in \mathbb{F}_2^{n \times k}$ for $\mathcal{C}_{in}$ and $\mathcal{C}_{out}$ respectively. Referring to the same random string N, the IDES-recovery algorithm ($\mathsf{Rec}_{\mathsf{IDES}}$) with inputs $ss, w', N, \varepsilon_{rec}$ to recover w is described as follow:

IDES-Recovery Algorithm: $\mathsf{Rec}_{\mathsf{IDES}}(ss, w', N, \varepsilon_{rec})$

Step 1: Sample e' where $\|e'\| = \lceil k^* \varepsilon_{rec} \rceil$
Step 2: Compute $w'_{e'} = w' \oplus e'$
Step 3: Generate the RV $\phi'_{e'} \leftarrow \Omega(w'_{e'}, N)$
Step 4: Compute $c' = ss \oplus \phi'_{e'}$
Step 5: Decode c' with G_{out} (via Gaussian elimination). If the output is $0^{k-n^*}||x$ with a random string $x \in \{0,1\}^{n^*}$, then set $x = v_{syn}$, else, return to Step 1
Step 6: Compute $c^* = w'_{e'} \oplus v_{syn}$
Step 7: Decode c^* with G_{in} (via Gaussian elimination).
Step 8: Outputs the decoded result as w

A. CORRECTNESS OF IDES-SECURE SKETCH

Follow the definition of a secure sketch (see Section III). The correctness of IDES-secure sketch characterizes the *success rate* in using $\mathsf{Rec}_{\mathsf{IDES}}$ to recover w from a sketch ss generated by $\mathsf{SS}_{\mathsf{IDES}}$. It is desirable to show that the probability of success to recover w using algorithm pair $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}} \rangle$ is $1 - \gamma$ with some probability $\gamma > 0$ described as follows

$$\Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) = w\right] = 1 - \gamma$$

The theorem below characterizes the correctness of IDES-secure sketch. Its proof is given in the Appendix Section.

Theorem 4: For $p(n) = 2^{\lceil k^* h_2(\varepsilon_{rec}) \rceil} - 1$, $\varepsilon_{ss} \in (0, 1/2)$, $\varepsilon_{rec} \in [1/k^*, 1/2]$, $k^* \leq n^* < k \leq n$, *where* $k^* \geq \log(p(n) + 1) \geq 1$, $n \geq 2$. *If* $\|w \oplus w'\| \leq k^* \varepsilon_{ss} \leq k^* \varepsilon_{rec}$ *(can be achieved if large enough n), and when* $\lfloor k^* h_2(\varepsilon_{ss}) \rfloor = k - n^*$, *then*

$$\Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) = w\right] = 1 - 2^{-(k-n^*)}. \quad (3)$$

Both sketching and recovery operate in $O(p(n)(n^3))$ *running time. Setting* $p(n) = n$, *then both sketching and recovery operate in* $O(n^4)$ *running time (for worst-case).*

B. SECURITY OF IDES-SECURE SKETCH

Follow the definition of a secure sketch (see Section III), we want to show the security of IDES-secure sketch to any adversary who observed ss and N, then w can be recovered with probability no greater than $2^{-\tilde{m}} = 2^{-(k-n^*)}$, depends upon the number of zero padding.

Viewing w' and w as a point in two random distributions W' and W respectively (over $\mathcal{M}_1$). During sketching with IDES-sketching algorithm, a total number of 2^{k-n^*} points in an unknown distribution U (over $\{0,1\}^{k-n^*}$) can be stored via zeros padding. Given a point in W', the correctness of a secure sketch can be showed to all $w_e = w \in W \in B_{t^*}(e \oplus w'_{e'})$ in IDES-model (see Claim 4). It means some of the points in U can be revealed as $w_e = w \in U$. Clearly, given U is precisely known, it means all the points w in distribution U are revealed, hence, recover any $w \in U$ is trivial. Since U can only be known precisely via distinguish $B_{t^*}(e \oplus w')$. Therefore, we analyze the security of IDES-secure sketch in term of the distinguishability of $B_{t^*}(e \oplus w'_{e'})$ conditioned on the case when the correctness of a secure sketch holds for all $\phi_e \in \Phi_e \in B_t(\phi'_{e'}) \in U$. Such setting can be achieved easily given an adversary have *unlimited* computation resources, and doing so is necessary to *complete* our security argument to reveal U precisely with any point under a random distribution $W \notin B_{t^*}(e \oplus w'_{e'}) \in U$ that cannot be found within the ball $B_{t^*}(e \oplus w'_{e'})$.

We formally defined the error in distinguish $B_{t^*}(e \oplus w'_{e'})$ conditioned on the case when the correctness of a secure sketch holds for any random distribution $\Phi_e \in B_t(\phi'_{e'}) \in U$ follows the definition given below.

Definition 5: Given any random distributions $W, W' \in \mathcal{M}_1$ *and two tolerance distances* $t^*_{\min} > 0$, *and* $t \geq 0$. *Given any random strings* $w \in W$, $w' \in W'$ *and error parameters* $\varepsilon_{ss} \in (0, 1/2]$, $\varepsilon_{rec} \in [1/k^*, 1/2]$. *Follow the procedure described in Definition 3 (Step 1 to 5). For all* $\phi_e \in \Phi_e$ *and* $\phi'_{e'} \in \Phi'_{e'}$ *over some random distribution* $\Phi_e \in \Psi$ *and*

$\Phi'_{e'} \in \Psi'$, *we say* $B_{t^*}(e \oplus w'_{e'})$ *is* $(t, t^*_{\min}, \beta')$*-distinguishable with error at most* $\beta' > 0$ *if for all* $t^* \geq t^*_{\min}$*:*

$$\mathbb{E}_{w' \leftarrow W'}\left[\Pr\left[W \notin B_{t^*}(e \oplus w'_{e'}) \mid \Phi_e \in B_t(\phi'_{e'})\right]\right] \leq \beta'$$

The following Theorem shows that given unlimited computation resources, the distinguishability error follows Definition 5 is *no greater than* $2^{-(k-n^*)}$ with any parameter $t^* \geq t^*_{\min}$ and any $t \geq 0$. The proof is given in Appendix Section.

Theorem 6: Denote $\varepsilon'_d = \left\|e \oplus e'\right\|(n)^{-1}$, $\xi' = \left\|\phi_e \oplus \phi'_{e'}\right\|(n)^{-1}$, *and* $\xi = \|\phi \oplus \phi'\|(n)^{-1}$. *Let* $t^*_{\min} = (\xi' + \varepsilon'_d)k^*$ *and* $t = (\xi - \varepsilon'_d)n$. *For* $p(n) = 2^{\lceil k^* h_2(\varepsilon_{rec})\rceil} - 1$, *with large enough* k^* *and* n, *s.t.* $k^*/n \approx 1$ *(where* $k^* \leq n^* < k \leq n$*), when* $\lfloor k^* h_2(\varepsilon_{ss})\rfloor > k - n^*$, *then* $B_{t^*}(e \oplus w'_{e'})$ *is* $(t, t^*_{\min}, \beta')$*-distinguishable with error* $\beta' < 2^{-(k-n^*)}$ *follows (holds for all* $t^* \geq t^*_{\min}$*)*

$$\mathbb{E}_{w' \leftarrow W'}\left[\Pr\left[W \notin B_{t^*}(e \oplus w'_{e'}) \mid \Phi_e \in B_t(\phi'_{e'})\right]\right] < 2^{-(k-n^*)}. \tag{4}$$

In particular, choosing $t^*_{\min} = 1$, $t = 0$ *are suffice.*

Conceivably, the security property of IDES-secure sketch contains a stronger notion in the sense that *both* k^* and n must be large enough compared to the correctness property which only require n to be large enough. Then, we give a proposition to conclude our construction is an efficient secure sketch. Its proof is trivial with Theorem 4 (for correctness and efficiency claims) and Theorem 6 (for security claim)

Proposition 1: For $p(n) = 2^{\lceil k^* h_2(\varepsilon_{rec})\rceil} - 1$, $\varepsilon_{ss} \in (0, 1/2)$, $\varepsilon_{rec} \in [1/k^*, 1/2]$, $k^* \leq n^* < k \leq n$, *where* $k^* \geq \log(p(n) + 1) \geq 1$. *With large enough* k^* *and* n *where* $k^*/n \approx 1$, *the algorithm pair* $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}}\rangle$ *is an efficient* $(\mathcal{M}_1, m, \tilde{m}, k^*\varepsilon_{rec})$ *secure sketch with* $\tilde{m} = k - n^*$ *(no. zeros padding). In particular, by setting* $k^* = n^*$ *and* $k = n$, *the achievable rate* $k^*/n = 1-(n-k^*)/n$ *is maximum.*

1) ENTROPY LOSS IN IDES-SECURE SKETCH

Based on the derived security bound in Eq. 4, it is convenient to express the minimum entropy loss in term of average min-entropy follows $\tilde{\mathrm{H}}_\infty(W \mid ss) = -\log(2^{-(k-n^*)})$. Doing so means the min-entropy of the sources must be equal to $\mathrm{H}_\infty(W) = k - n^*$ to ensure positive entropy is remained in the sources for meaningful security claim, i.e., $\mathrm{H}_\infty(W) - \tilde{\mathrm{H}}_\infty(W \mid ss) \geq 0$.

Besides, the security of the IDES-secure sketch can also be relaxed in the sense that it only shows security to computationally bounded attackers, i.e., polynomial time-bounded. More precisely, Theorem 4 characterizes the correctness of IDES-recovery to recover w in polynomial running time with certain parameterization (i.e., large enough n and when $\lfloor k^* h_2(\varepsilon_{ss})\rfloor = k - n^*$). Otherwise, there is no correctness guarantee for the IDES-recovery algorithm operating in polynomial time unless the problem of decoding a random linear code can be solved in polynomial time, which is reported to be an NP-hard problem [33]. In particular, for super-logarithm fuzzy min-entropy, setting $\mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) = k - n^* = \omega(\log(n))$ is suffice. However, such relaxation gives *no advantages* since doing so is sufficient to yield a secure sketch, defined information-theoretically, for any sources with min-entropy equal to the fuzzy min-entropy, i.e., $\mathrm{H}_\infty(W \mid ss) = \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) = k - n^*$.

VII. IDES-SECURE SKETCH FOR ERROR CORRECTION

This section gives the discussion on the usage of IDES-secure sketch as an error correction code.

Noted that the correctness property of the algorithm pair $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}}\rangle$ allows the recovery of w using w' when $\left\|w \oplus w'\right\| \leq k^*\varepsilon_{rec}$. Indeed, one of the necessary condition to support the correctness property of the algorithm pair $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}}\rangle$ is the value of n that must be large enough. Such necessity suggesting the algorithm pair $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}}\rangle$ (IDES-secure sketch) renders large classes of asymptotic good error correction codes to tolerate the errors in between w and w' given $\left\|w \oplus w'\right\| \leq k^*\varepsilon_{rec}$. In particular, it offers error correction for arbitrary error rate bounded by $p(n)$ if $\varepsilon_{ss} \leq \varepsilon_{rec} \leq h_2^{-1}(\log(p(n)+1)/k^*) \leq 1/2$.

To illustrate how IDES-secure sketch renders large classes of asymptotic good error correction code, we only focus on the setting when $k^* = n^*$ and $k = n$. Let define the rate R of arbitrary (non-trivial) random $[n^*, k-n^*, d]_2$ code $\mathcal{C}$ follows:

$$R = 1 - (k - n^*)/n^*.$$

Here, R measures the information symbol per code symbol. Large R means less redundancy due to less zeros padding. Note that $n^* = 2(k-n^*)$. Without loss of generality, since the inner code's generator matrix G_{in} is random, we can define the generator matrix of the random code $\mathcal{C}$ as $G \in \mathbb{F}_2^{n^* \times (k-n^*)}$, which is a sub-matrix formed by taking only the first $k - n^*$ columns (or the first half) of G_{in}. Given large enough k^* and n where $k^*/n \approx 1$, the inequality $\lfloor k^* h_2(\varepsilon_{ss})\rfloor > k - n^*$ is *necessary* to show security of the IDES-secure sketch (see Theorem 6) that can be generally described using $k^* h_2(\varepsilon_{ss}) \geq \lfloor k^* h_2(\varepsilon_{ss})\rfloor > k - n^*$. Therefore we yielded the inequality described below as one of the necessary condition to show security of the IDES-secure sketch with code $\mathcal{C}$ of rate

$$R > 1 - h_2(\varepsilon_{ss}) \ (security).$$

On the other hand, given large enough n, the equality $\lfloor k^* h_2(\varepsilon_{ss})\rfloor = k - n^*$ is *necessary* to show correctness of IDES-secure sketch (see Theorem 4) that must follow only if $k^* h_2(\varepsilon_{ss}) = \lfloor k^* h_2(\varepsilon_{ss})\rfloor = k - n^*$, yielding one of the necessary condition to show correctness of IDES-secure sketch with code $\mathcal{C}$ of rate R described as:

$$R = 1 - h_2(\varepsilon_{ss}) \ (correctness).$$

Recall that the input error parameter ε_{ss} is viewed as the "natural" random noise incurred in the noisy source, which is unknown in prior. Hence, for arbitrary distance d between the codewords $c \in \mathcal{C}$, we can conveniently express $(d-1)/n^* = 2\varepsilon_{ss}$. It should be obvious that by the proof of Theorem 4 (subsequent paragraph after Eq. 5), $(d-1)/n^*$ equal to the

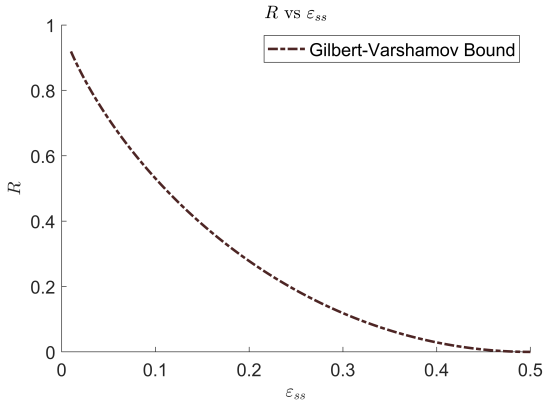

FIGURE 3. Graph of R vs ε_{ss}.

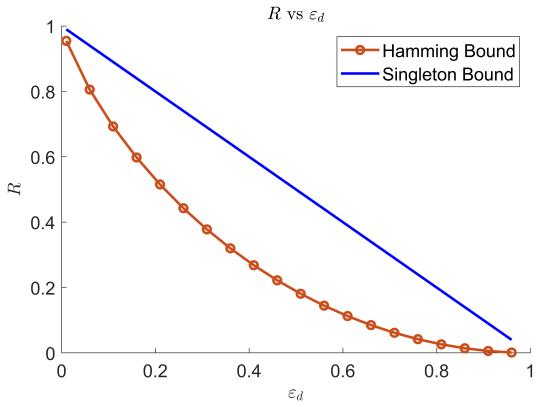

FIGURE 4. Graph of R vs ε_d.

non-trivial minimum rate $(d-1)/n^* = \varepsilon_d = 2\varepsilon_{ss}$ that gives us the minimum solution of $\varepsilon_{rec} = \varepsilon_{ss}$. Thus, $d = 2n^*\varepsilon_{ss} + 1$ is the *minimum* distance between the codewords of code $\mathcal{C}$. The IDES-secure sketch is able to tolerate any error rate of $\varepsilon_{ss} \in (0, 1/2)$ that would mean $\varepsilon_d \in (0, 1)$ can be determined. The graph of R vs ε_{ss} is plotted in Figure 1. It showed that IDES-secure sketch offers a code $\mathcal{C}$ whose attains the trade-off curve in between R and ε_{ss} follows Figure 3 which is also known as *Gilbert-Varshamov* bound [34]. On the other hand, the curve of R vs ε_d where

$$R = 1 - h_2(\varepsilon_d/2) = 1 - h_2(\varepsilon_{ss})$$

is plotted in Figure 4, indicating the *Hamming bound* ([35], p. 83). Noting that since the non trivial minimum rate $\varepsilon_d = 2\varepsilon_{ss}$ can be described using ε_{ss}, and sufficiently large n will reduce the error in recovering w from $\exp(-2n\varepsilon_d^2)$ to $2^{-(k-n^*)}$ (see proof of Theorem 4, Eq. 5). Therefore, it is safe to say that any code $\mathcal{C}$ offered by IDES-secure sketch with arbitrary parameter $[n^*, k-n^*, d]_2$ lies on the curve of Gilbert-Varshamov bound also lies on the curve of Hamming bound. Such result resolved the long-standing open conjecture saying that the Gilbert-Varshamov bound is the best trade-off curve in between R and $(d-1)/n^*$ (for binary codes) regardless the computation resources in constructing the code itself (see [36], Chap. 17, Theorem 30).

The theorem below formalized (IDES-secure sketch) the algorithm pair $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}} \rangle$ as an optimal linear code that meets the singleton bound (see Figure 4, blue solid line) with minimum distance $d \geq n - k^* + 1$. Its proof is given in the Appendix Section.

Theorem 7: For $n = k$, $n^* = k^*$, and $k - n^* \geq 2$, with large enough n, the algorithm pair $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}} \rangle$ is a $[n, k^*, d]$-linear code with minimum distance

$$d \geq n - k^* + 1 \ (\textit{singeton bound}).$$

In particular, it can tolerate arbitrary number of error at least $(d-1)/2 \geq \lfloor (d-1)/2 \rfloor = t = 1$ with probability of success at least 0.75.

In fact, it is not difficult to verify that under the same setting, $k^* = n^*$, $k = n$, and large enough n, the code $\mathcal{C}$ also attains the singleton bound with minimum distance $d = 2n^*\varepsilon_{ss} + 1$. Nevertheless, we now show that the above derived coding bounds are subjected to a stronger bound, known as the *Shannon bound* [37]. Specifically, by Theorem 4, a successful decoding result in running the IDES-recovery algorithm would mean the correctness holds with $k^* h_2(\varepsilon_{ss}) = k - n^*$. Therefore, by setting large enough k^* and n with arbitrary value of $k - n^* \geq 1$ s.t. $k^*/n \approx 1$, it follows $k^* h_2(\varepsilon_{ss})/n \approx h_2(\varepsilon_{ss}) \approx k - n^*/n$. More precisely, due to the fact that $k^* < n$, it means $h_2(\varepsilon_{ss}) < (k - n^*)/n$ must follow. Therefore, let denote $C = 1 - h_2(\varepsilon_{ss})$ be the channel capacity and $R = 1 - (n - k^*)/n$, the correctness property of IDES-secure sketch holds iff $h_2(\varepsilon_{ss}) < (n - k^*)/n$, which leads us to the inequality below

$$R < C.$$

In contrary, no correctness guaranty given $k^* h_2(\varepsilon_{ss}) > k - n^*$ under the same setting (large k^* and n, where $k^*/n \approx 1$), which refers to the case when

$$R > C.$$

Noting that the latter inequality exactly described the security requirement of IDES-secure sketch follows Theorem 6.

Last but not least, IDES-model models the input noise in worst-case. The input error rate ε_{ss}, corresponding to the maximum number of bits changed caused by the error e, is maximum. Therefore, the code derived using the IDES-secure sketch does not restrict being used for particular communication channels, such as the discrete memoryless channel and the additive white Gaussian noise (AWGN) channel.

Remark: When $\varepsilon_{ss} = \varepsilon_{rec} = 1/2$, the recovery algorithm gives no advantages compared to brute-forcing all 2^{k^*} possible results for w. Therefore, in Theorem 4, we are only interested in $\varepsilon_{ss} \in (0, 1/2)$ to show meaningful correctness without brute forcing.

Example: We here give an example of stimulation on IDES-secure sketch in verifying the existence of the code $\mathcal{C}$ with minimum distance $d = 2n^*\varepsilon_{ss} + 1$. To do so, one shall choose an arbitrary number of zeros padding $k - n^* \geq 1$ with fixed $k^* = n^*$, $k = n$ and a large $p(n) = 2^{\lceil k^* h_2(\varepsilon_{rec}) \rceil} - 1$.

TABLE 1. Summary of security bound of existing secure sketch in terms of average min-entropy (with error $\gamma > 0$). Only the secure sketch over the first three rows can show meaningful security to more error than entropy sources based on fuzzy min-entropy notion. Clearly, the proposed approach is able to attain the best possible security with parameters $k - n^* = 1$.

Security Bound of Existing Secure Sketch Construction	
Best possible security [9]	$\tilde{\mathrm{H}}_\infty(W\|ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1-\gamma)$
FRS sketch (universal hash functions) [9]	$\tilde{\mathrm{H}}_\infty(W\|ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1/\gamma) - \log\log(\mathsf{supp}(W)) - 1$
Layer hiding hash (strong universal hash function) [8]	$\tilde{\mathrm{H}}_\infty(W\|ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1/\gamma) - 1$
Fuzzy commitment [10]	$\tilde{\mathrm{H}}_\infty(W\|ss) \geq t\log(n)$ (when $t \ll n$)
Fuzzy vault [4]	$\tilde{\mathrm{H}}_\infty(W\|ss) \geq t\log(n)$
Improved Fuzzy vault [6]	$\tilde{\mathrm{H}}_\infty(W\|ss) > t\log(n) + 2$
Pinsketch [6]	$\tilde{\mathrm{H}}_\infty(W\|ss) \geq t\log(n+1)$
Proposed	$\tilde{\mathrm{H}}_\infty(W\|ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1-\gamma) = 0$ (achievable with $k - n^* = 1$)

Since the recovery algorithm must halt in $O(p(n)n^3)$, if the first decoding (Step 5 of IDES-recovery) is success, it means $0^{k-n^*}||v_{syn}$ is returned. Then c^* can be computed using v_{syn} and $w'_{e'}$ (Step 6 of IDES-recovery).

The validity of the codeword c, s.t. $c \in \mathcal{C}$ is the codeword of code $\mathcal{C}$, can be verified using G with any non-zero random vector $x \in \{0,1\}^{k-n^*}$ follows $Gx = c$. More precisely, it can be described as

$$Gx = c^* \oplus (w_e \oplus w'_{e'}) = c^* \oplus (w \oplus w') \oplus (e \oplus e') = c.$$

The code $\mathcal{C}$ is a good error correction code iff the correctness of IDES-secure sketch (Theorem 4) holds. Specifically, the prove of Theorem 4 emphasised on the case when $w_e = w'_{e'}$, i.e., $\|w_e \oplus w'_{e'}\|(n^*)^{-1} = \|(w \oplus w') \oplus (e \oplus e')\|(n^*)^{-1} = \xi - \varepsilon_d = 0$, which implies $Gx = c = c^*$. By Eq. 5, with minimum $\varepsilon_{rec} = \varepsilon_{ss}$, we have $\lfloor k^* h_2(\varepsilon_{rec}) \rfloor = \lfloor k^* h_2(\varepsilon_{ss}) \rfloor = k - n^*$, therefore, the validity of $Gx = c = c^*$ can be verified using G, c^*, and a non-zero random vector $x \in \{0,1\}^{k-n^*}$ in at most $2^{k-n^*} - 1 \leq 2^{k^* h_2(\varepsilon_{rec})} - 1 \leq p(n)$ operations, i.e., polynomial time.

To argue the minimum distance of code $\mathcal{C}$. Since $\xi - \varepsilon_d = 0$, it is succinct to express $(d-1)/n^* = \xi = \varepsilon_d = 2\varepsilon_{ss}$, where $d = 2n^*\varepsilon_{ss} + 1$ is the minimum distance between the codewords of code $\mathcal{C}$, corresponding to the minimum distance $\|w \oplus w'\| = \|e \oplus e'\|$.

Summing up, the existence of code $\mathcal{C}$ can be verified in $O(p(n))$ operations follows $Gx = c^*$ iff the correctness of IDES-secure sketch holds (means c^* can be computed in $O(p(n)n^3)$). The generator matrix G of code $\mathcal{C}$ can be computed in polynomial time using G_{in} (recall that G is the first $k - n^*$ columns of G_{in}). Otherwise, there is no such G can be computed from G_{in} in polynomial time.

VIII. COMPARISON: SECURITY BOUND OF SECURE SKETCH

Table 1 depicted the security bound of existing secure sketch construction. Compared to the existing secure sketch constructions, our construction is capable of claiming security to all noisy sources with min-entropy (equivalent to the fuzzy min-entropy) equal to $k - n^* \geq 1$.

At the end, we give a proposition below to show that the best possible secure sketch with *optimal* entropy loss implies IDES-secure sketch with $k - n^* = 1$. The proof is given in the Appendix Section.

Proposition 2: For any random distribution $W \in \mathcal{M}_1$ with average min-entropy $\mathrm{H}_\infty(W \mid ss) \geq \tilde{m} \geq 0$, an IDES-secure sketch with $k - n^* = 1$ is a secure sketch that meets the best possible security bound with single bit of fuzzy min-entropy follows (i.e., $\gamma = 1/2$)

$$\tilde{\mathrm{H}}_\infty(W \mid ss) \geq \mathrm{H}^{\mathrm{fuzz}}_{t,\infty}(W) - \log(1-\gamma) = 0.$$

IX. CONCLUSION

In this paper, we studied the limitation of existing secure sketch. Traditional way of analyzing the security of secure sketch, solely relying on min-entropy of a source's distribution W and the error tolerance parameter t, has not considered the input structure of the distribution, i.e., distance between the points in W. This way of analysis shows no security guaranty over input sources with large number of errors, especially when there has more errors than the entropy of the sources. Fuller *et al.*, [9] have showed impossibility in constructing a secure sketch under distribution uncertainty setting. Therefore, motivated by the precise knowledge on the input distribution is believed to be necessary for secure sketch construction. We formalized a new sketching model, namely IDES-model, to model the input distribution with a distinguisher. We claimed (with proofs) that IDES-model is necessary to realize precise knowledge on the input distribution setting. Besides, we devised an explicit secure sketch construction, called IDES-secure sketch, that meets the best possible secure sketch's security bound with optimal entropy loss. In particular IDES-secure sketch shows security to any random source's distribution W with min-entropy at least a single bit.

APPENDIX A
PROOFS OF THE CLAIMS OVER IDES-MODEL FORMALIZATION

Proof of Claim 1:

Proof: The prove uses a sketching algorithm that stores all 2^m neighbour points within distance t^* for a given $w' \in W'$, which already implies the need of a distinguisher to distinguish $B_{t^*}(w')$ that consists of 2^m points. It follows that any points $w \in W \in B_{t^*}(w')$ must have stored by the sketching

algorithm, means precise knowledge on $W \in B_{t^*}(w')$ and correctness of a secure sketch follows. □

Proof of Claim 2:

Proof: Our prove uses a sketching algorithm that stores all $w_e \in W_e \in B_{t^*}(w')$ that is close to w' within distance t^*, means $B_{t^*}(w')$ can be distinguished and implies any random distribution $W_e \in B_{t^*}(w')$ is precisely known. Because $\|w_e \oplus w'\| = \|w \oplus (e \oplus w')\|$, therefore, all $w \in W \in B_{t^*}(e \oplus w')$ that is close to $(e \oplus w')$ within distance t^* should have stored as well, which means any random distribution $W \in B_{t^*}(e \oplus w')$ is precisely known and implies $B_{t^*}(e \oplus w')$ can be distinguished and shows correctness of a secure sketch to all $w \in W \in B_{t^*}(e \oplus w')$. □

Proof of Claim 3: *Proof:* Our prove uses a sketching algorithm that stores all $\phi_e \in \Phi_e \in B_t(\phi')$ that is close to ϕ' within distance $t \geq t^*$, means $B_t(\phi')$ can be distinguished and implies any random RV distribution $\Phi_e \in B_t(\phi')$ is precisely known. This shows correctness of a secure sketch to all $\phi_e \in \Phi_e \in B_t(\phi')$ over $\mathcal{M}_2$. On the other hand, given any RV $\phi_e \in \Phi_e$ and N, it is not difficult to verify that w_e can be recovered by "reverse permutation" using N (long enough N is suffice). Hence, precise knowledge of $\Phi_e \in B_t(\phi')$ implies precise knowledge on $W_e \in B_{t^*}(w')$, then Claim 2 follows for the correctness claims of a secure sketch for all $w \in W \in B_{t^*}(e \oplus w')$. □

Proof of Claim 4:

Proof: We want to show that $B_t(\phi'_{e'})$ can be distinguished in $O(p(n))$ steps if n is large enough. To do so, we first have to show correlation in between Φ_e and $\Phi'_{e'}$. Obviously, the only way to show correlation in between these two distributions is to show their error vectors are the same, i.e., $e = e'$. Recall the error vector $e \in \mathcal{E}_{ss} = \{e_{(1)}, \ldots, e_{(|\mathsf{supp}(\mathcal{E}_{ss})|)}\}$ and $e' \in \mathcal{E}_{rec} = \{e'_{(1)}, \ldots, e'_{(|\mathsf{supp}(\mathcal{E}_{rec})|)}\}$. We can explicitly describe the error vector $e = e_{(i)}$ is with an index value $i \in \{1, \ldots, |\mathsf{supp}(\mathcal{E}_{ss})|\}$. Same goes to the error vector $e' = e'_{(j)}$ is with another index value $j \in \{1, \ldots, |\mathsf{supp}(\mathcal{E}_{rec})|\}$. Clearly, $e_{(i)} = e'_{(j)}$ iff $i = j$, and $j \geq i = |\mathsf{supp}(\mathcal{E}_{ss})|$ is necessary to reveal any possible $i \leq |\mathsf{supp}(\mathcal{E}_{ss})|$. Hence, choosing large enough n for large $p(n)$ is suffice. More precisely, for $\varepsilon_{rec} \in [1/k^*, 1/2]$, we use *Stirling's approximation* and obtain (for any $i \in \{1, \ldots, |\mathsf{supp}(\mathcal{E}_{ss})|\}$):

$$\begin{aligned} i \leq j \leq |\mathsf{supp}(\mathcal{E}_{rec})| &= \binom{k^*}{\lceil k^* \varepsilon_{rec} \rceil} \\ &\leq 2^{\lceil k^* h_2(\varepsilon_{rec}) \rceil} = p(n) + 1. \end{aligned}$$

Given $i = j$ (means $e = e'$), it follows that distinguish $B_{t^*}(e \oplus w'_{e'}) = B_{t^*}((e \oplus e') \oplus w')$ is equal to distinguish $B_{t^*}(w')$ (since $e = e'$ cancel off). Hence, distinguish $B_{t^*}(e \oplus w'_{e'})$ in $O(p(n))$ steps is suffice to reveal any random distribution $W_e = W \in B_{t^*}(e \oplus w'_{e'})$ of at most $p(n)$ points precisely, and show correctness of a secure sketch to all $w_e = w \in W \in B_{t^*}(e \oplus w'_{e'})$ over $\mathcal{M}_1$. Since W is precisely known, its corresponding RV distribution Φ is also known precisely given N. Then, for any random distribution $\Phi_e \in B_t(\phi'_{e'})$, Φ_e can be revealed precisely iff $\Phi_e = \Phi_{e'} = \Phi$. This implies $B_t(\phi'_{e'})$ can be distinguished iff $B_{t^*}(e \oplus w'_{e'})$ can be distinguished. Because of both $B_{t^*}(e \oplus w'_{e'})$ and $B_t(\phi'_{e'})$ consist of not more than $p(n)$ points. Therefore, precise knowledge on Φ with $p(n)$ points means $B_t(\phi'_{e'})$ can be distinguished, and this show correctness of a secure sketch to all $\phi_e \in \Phi_e \in B_t(\phi'_{e'})$ over $\mathcal{M}_2$. □

APPENDIX B
PROOFS OF THE LEMMA FOR IDES-SECURE SKETCH

Proof of Lemma 1:

Proof: Given a random distribution $\Phi_e \notin B_t(\phi'_{e'})$, it means for all $\phi_e \in \Phi_e$, $\|\phi_e \oplus \phi'_{e'}\| > t$. For any random $W \in B_{t^*}(e \oplus w'_{e'})$, it means $W \in B_{t^*}(e \oplus w' \oplus e')$. For all $w \in W$, it follows

$$\|w_e \oplus w'_{e'}\| = \|(w \oplus w') \oplus (e \oplus e')\| \leq t^*.$$

With $\varepsilon_d = \|e \oplus e'\| (k^*)^{-1} > 0$ and $t^* = (\xi - \varepsilon_d)k^*$, $\|w_e \oplus w'_{e'}\|(k^*)^{-1} \leq \xi - \varepsilon_d$ follows. Thus, $n\xi' \leq n\xi - n\varepsilon_d$ (by multiplying both sides of the inequality with n). With $t_{\min} = n(\xi' + \varepsilon_d)$, one yields $n\xi \geq t_{\min}$. Then, let $t = n\xi$, the probability for $\|\phi_e \oplus \phi'_{e'}\| > t \geq t_{\min}$ given $\|w_e \oplus w'_{e'}\| \leq t^*$ can be computed as

$$\begin{aligned} &\mathbb{E}_{w' \leftarrow W'}\left[\Pr\left[\Phi_e \notin B_t(\phi'_{e'}) \mid W \in B_{t^*}(e \oplus w'_{e'})\right]\right] \\ &= \Pr\left[\|\phi_e \oplus \phi'_{e'}\| > t \mid \|w_e \oplus w'_{e'}\| \leq t^*\right] \\ &= \Pr\left[\|\phi_e \oplus \phi'_{e'}\| \geq t_{\min} \mid \|w_e \oplus w'_{e'}\| \leq t^*\right] \\ &= \Pr\left[\|\phi_e \oplus \phi'_{e'}\| \geq n(\xi' + \varepsilon_d) \mid \|w_e \oplus w'_{e'}\| \leq t^*\right] \\ &\leq \exp\left(-2n\varepsilon_d^2\right) \end{aligned}$$

The last two lines of the above equations follow *Hoeffding's inequality* and Theorem 2, stating that $\mathbb{E}\left[\|\phi_e \oplus \phi'_{e'}\|\right] = n\xi'$ (given $\|w_e \oplus w'_{e'}\| = k^*\xi'$), holds for all $w \in W$, $t^* \geq 0$, and $t \geq t_{\min}$. □

APPENDIX C
PROOFS OF THE THEOREMS FOR IDES-SECURE SKETCH

Proof of Theorem 4:

Proof: We first measure the probability of failure to recover w given $w_e = w'_{e'}$, means $t^* = 0$. Obviously, the recovery algorithm will fails in this case if $\phi_e \neq \phi'_{e'}$. Such error can be described as

$$\begin{aligned} &\Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) \neq w\right] \\ &\qquad = \Pr\left[\phi_e \neq \phi'_{e'} \mid w_e = w'_{e'}\right] \end{aligned}$$

We then show a reduction for the distinguishability error follows Lemma 1 to the probability described at above. In particular, choosing $t_{\min} = 1$ is suffice to describe (holds for all $t \geq t_{\min}$):

$$\begin{aligned} &\Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) \neq w\right] \\ &= \Pr\left[\phi_e \neq \phi'_{e'} \mid w_e = w'_{e'}\right] \\ &= \Pr\left[\|\phi_e \oplus \phi'_{e'}\| > 0 \mid \|w_e \oplus w'_{e'}\| = 0\right] \\ &= \Pr\left[\|\phi_e \oplus \phi'_{e'}\| \geq 1 \mid \|w_e \oplus w'_{e'}\| \leq 0\right] \\ &= \Pr\left[\|\phi_e \oplus \phi'_{e'}\| \geq t_{\min} \mid \|w_e \oplus w'_{e'}\| \leq t^*\right] \\ &\leq \exp(-2n\varepsilon_d^2) = \exp(-8n\varepsilon_{ss}^2) \leq 1/(p(n)+1) \\ &= 2^{-\lceil k^* h_2(\varepsilon_{rec}) \rceil} \leq 2^{-k^* h_2(\varepsilon_{rec})} = 2^{-k^* h_2(\varepsilon_{ss})} \\ &\leq 2^{-\lfloor k^* h_2(\varepsilon_{ss}) \rfloor} = 2^{-(k-n^*)}. \end{aligned} \tag{5}$$

The last line of Eq. 5 follows the equivalent of $\lfloor k^* h_2(\varepsilon_{ss})\rfloor = k - n^*$. The last second line of Eq. 5 follows if n is large enough, implies large enough $\varepsilon_{rec} \geq \varepsilon_{ss}$. Particularly, we refer to minimum $\varepsilon_{rec} = \varepsilon_{ss}$ for our prove. Then, the last third line of Eq. 5 follows with large n, the error term $\exp(-2n\varepsilon_d^2)$ decreases (very fast) exponentially but the error term $p(n)$ decreases (slowly) polynomially depends on $p(n)$. More precisely, since $\varepsilon_{rec} \geq \varepsilon_{ss}$ is necessary to reveal any e with e' (follows claim 4). In light of this, for minimum $\varepsilon_{rec} = \varepsilon_{ss}$, it follows $\varepsilon_d = 2\varepsilon_{ss}$ is minimum with the expression of $\varepsilon_d = \varepsilon_{ss} \pm \varepsilon_{rec}$ (recall $\varepsilon_d = \|e \oplus e'\| (k^*)^{-1}$). Otherwise, $\varepsilon_d = 0$ and we obtain an error $\exp(-2n\varepsilon_d^2) = 1$ which lead to the whole prove obsoleted. Argued in this way, given large enough n and $\lfloor k^* h_2(\varepsilon_{ss})\rfloor = k - n^*$, the error $\exp(-2n\varepsilon_d^2)$ can be described as $\exp(-8n\varepsilon_{ss}^2)$ and reduced to $2^{-(k-n^*)}$ with minimum $\varepsilon_{rec} = \varepsilon_{ss}$.

Reasoned as above, for $k - n^* \geq 1$, the recovery algorithm can recover w successfully with probability

$$\begin{aligned}&\Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) = w\right]\\ &= 1 - \Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) \neq w\right]\\ &= 1 - 2^{-(k-n^*)} \geq 1/2. \end{aligned} \tag{6}$$

We then continue to prove that the sketching and recovery can always be done in polynomial time. Recall in IDES-model, we are only interested in $\mathcal{W}$ of size $|\mathcal{W}| = |\mathsf{supp}(\mathcal{E}_{ss})| = \omega(p(n))$. For a given error parameter ε_{rec}, there are exactly $|\mathsf{supp}(\mathcal{E}_{rec})|$ number of possible error vector $e' \in \mathcal{E}_{rec}$. More precisely, for $\varepsilon_{rec} \in [1/k^*, 1/2]$, one can use *Stirling's approximation* to obtain

$$|\mathsf{supp}(\mathcal{E}_{rec})| = \binom{k^*}{\lceil k^* \varepsilon_{rec}\rceil} \leq 2^{\lceil k^* h_2(\varepsilon_{rec})\rceil} = p(n) + 1.$$

Follows above equation, the IDES-recovery algorithm could repeat Step 1 to Step 5 at most $p(n)$ number of times where the remaining steps can be done in the order of $O(n^3)$. The overall complexity of the recovery algorithm is asymptotically bounded in $O(p(n)(n^3))$. It follows the running time of the IDES-sketching algorithm is bounded in $O(n^3)$. Therefore, the IDES-secure sketch works in polynomial time running time with worst-case running time $O(p(n)(n^3))$ and best-case running time $\Omega(n^3)$. It should be noted that one can also set $p(n) = n$ to yield a polynomial worst-case running time for the IDES-secure sketch described as $O(n^4)$.

The remaining is to prove $\|w \oplus w'\| \leq k^* \varepsilon_{ss} \leq k^* \varepsilon_{rec}$ and $k^* \geq \log(p(n) + 1) \geq 1$. Since $t^* = (\xi - \varepsilon_{ss})k^* = 0$, if w can be recovered (correctness holds) in $O(p(n)(n^3))$ steps, it means $\varepsilon_{ss} = \xi$. Follow the previous argument for last third line of Eq. 5, with sufficient large n, we get minimum $\varepsilon_{rec} = \varepsilon_{ss}$. Hence, the minimum error rate $\xi = \varepsilon_{ss} = \varepsilon_{rec}$ can be bounded follows

$$0 < \xi = \varepsilon_{ss} = \varepsilon_{rec} \leq h_2^{-1}(\log(p(n) + 1)/k^*) \leq 1/2.$$

Therefore, $\|w \oplus w'\| = \lfloor k^* \xi \rfloor \leq k^* \xi = k^* \varepsilon_{ss} = k^* \varepsilon_{rec}$. To prove that above statement holds for any $k^* \geq 1$ and $n \geq 2$: Noted that $k^* \leq n^* < k \leq n$, and $h_2^{-1}(\log(p(n) + 1)/k^*) \geq \varepsilon_{rec}$, where $\log(p(n)+1)/k^* \leq 1$. Therefore, for any minimum polynomial $p(n) = n > k^* \geq k - n^*$, the minimum $k^* \geq 1$ and $n \geq 2$ follows. □

Proof of Theorem 6:

Proof: We first show that the distinguishability error is at most $\exp(-2n(\varepsilon'_d)^2)$ follows

$$\mathbb{E}_{w' \leftarrow W'}\left[\Pr\left[W \notin B_{t^*}(e \oplus w'_{e'}) \mid \Phi_e \in B_t(\phi'_{e'})\right]\right] \leq \exp(-2n(\varepsilon'_d)^2).$$

For a random distribution $W \notin B_{t^*}(e \oplus w'_{e'})$, it means for any $w \in W$, $\|w \oplus (e \oplus w_{e'})\| = \|w_e \oplus w'_{e'}\| > t^*$.

For any two random RVs ϕ and ϕ' with $\xi = \|\phi \oplus \phi'\| (n)^{-1}$. Given $\Phi_e \in B_t(\phi'_{e'})$, It means

$$\|\phi_e \oplus \phi'_{e'}\| = \|(\phi \oplus \phi') \oplus (e \oplus e')\| \leq t.$$

With $\varepsilon'_d = \|e \oplus e'\| (n)^{-1}$ and $t = (\xi - \varepsilon_d)n$, $\|\phi_e \oplus \phi'_{e'}\|(n)^{-1} \leq \xi - \varepsilon_d$ follows. Thus, $k^*\xi' \leq k^*\xi - k^*\varepsilon'_d$ (by multiplying both sides of the inequality with k^*). With $t^*_{\min} = (\xi' + \varepsilon'_d)k^*$, one yields $t^*_{\min} \leq k^*\xi$. Then, let $t^* = k^*\xi$, the probability for $\|w_e \oplus w'_{e'}\| > t^* \geq t^*_{\min}$ given $\|\phi_e \oplus \phi'_{e'}\| \leq t$ can be computed as

$$\begin{aligned}&\mathbb{E}_{w' \leftarrow W'}\left[\Pr\left[W \notin B_{t^*}(e \oplus w'_{e'}) \mid \Phi_e \in B_t(\phi'_{e'})\right]\right]\\ &= \Pr\left[\|w_e \oplus w'_{e'}\| > t^* \mid \|\phi_e \oplus \phi'_{e'}\| \leq t\right]\\ &= \Pr\left[\|w_e \oplus w'_{e'}\| \geq t^*_{\min} \mid \|\phi_e \oplus \phi'_{e'}\| \leq t\right]\\ &= \Pr\left[\|w_e \oplus w'_{e'}\| \geq (\xi' + \varepsilon_d)k^* \mid \|\phi_e \oplus \phi'_{e'}\| \leq t\right]\\ &\leq \exp(-2n(\varepsilon'_d)^2)\end{aligned}$$

The last two lines of above the equations follow using *Hoeffding's inequality* and Theorem 2, stating that $\mathbb{E}\left[\|w_e \oplus w'_{e'}\|\right] = k^*\xi'$ (given $\|\phi_e \oplus \phi'_{e'}\| = n\xi'$), holds for all $w \in W$, $t \geq 0$ and $t^* \geq t^*_{\min}$.

Then, we show a reduction for the above distinguishability error to the probability described as

$$\Pr\left[w_e \neq w'_{e'} \mid \phi_e = \phi'_{e'}\right] < 2^{-(k-n^*)}.$$

Using $t^*_{\min} = 1$ and $t = 0$ is suffice. The reduction follows if large enough k^* and n:

$$\begin{aligned}&= \Pr\left[w_e \neq w'_{e'} \mid \phi_e = \phi'_{e'}\right]\\ &= \Pr\left[\|w_e \oplus w'_{e'}\| > 0 \mid \|\phi_e \oplus \phi'_{e'}\| = 0\right]\\ &= \Pr\left[\|w_e \oplus w'_{e'}\| \geq 1 \mid \|\phi_e \oplus \phi'_{e'}\| \leq 0\right]\\ &= \Pr\left[\|w_e \oplus w'_{e'}\| \geq t^*_{\min} \mid \|\phi_e \oplus \phi'_{e'}\| \leq 0\right]\\ &= \Pr\left[\|w_e \oplus w'_{e'}\| \geq (\xi' + \varepsilon_d)k^* \mid \|\phi_e \oplus \phi'_{e'}\| \leq t\right]\\ &\leq \exp(-2n(\varepsilon'_d)^2) = \exp(-8(k^*)^2\varepsilon_{ss}^2/n) \leq 1/(p(n) + 1)\\ &= 2^{-\lceil k^* h_2(\varepsilon_{rec})\rceil} \leq 2^{-k^* h_2(\varepsilon_{rec})} = 2^{-k^* h_2(\varepsilon_{ss})}\\ &\leq 2^{-\lfloor k^* h_2(\varepsilon_{ss})\rfloor} < 2^{-(k-n^*)}.\end{aligned} \tag{7}$$

The last line of Eq. 7 follows if the inequality $\lfloor k^* h_2(\varepsilon_{ss})\rfloor > k - n^*$ holds. The last second line of Eq. 7 follows by large enough n, implies large enough $\varepsilon_{rec} \geq \varepsilon_{ss}$. In particular, we refer to the minimum $\varepsilon_{rec} = \varepsilon_{ss}$ for our prove.

For the last third line of Eq. 7, it follows with large $k^* \leq n^* < k \leq n$, the error term $\exp(-2n(\varepsilon'_d)^2)$, which

can be described as $\exp(-8(k^*)^2\varepsilon_{ss}^2/n)$, decreases (very fast) exponentially with increment of k^* and the expression of $\varepsilon_d' = (k^*/n)(\varepsilon_{ss} \pm \varepsilon_{rec})$. More precisely, since $\varepsilon_{rec} \geq \varepsilon_{ss}$ is necessary to reveal any e with e' (follows claim 4). For minimum $\varepsilon_{rec} = \varepsilon_{ss}$, we can recall the previously derived minimum $\varepsilon_d = 2\varepsilon_{ss}$ in the proof of Theorem 4, and yield $\varepsilon_d' = 2k^*\varepsilon_{ss}/n = k^*\varepsilon_d/n$. Obviously, in such a case, $\varepsilon_d' < \varepsilon_d$, hence it is more appropriate to refer ε_d' as the minimum rate. Noting that $\varepsilon_d' \to \varepsilon_d$ given $k^* \to n$. Therefore, for arbitrary large enough value of n, given k^* is also large enough, close to n, i.e., $k^*/n \approx 1$, the security property of IDES-secure sketch is obsoleted by its correctness property. In other words, for large n and k^* where $k^*/n \approx 1$, the security property of IDES-secure sketch holds iff $\lfloor k^* h_2(\varepsilon_{ss})\rfloor > k - n^*$. □

Proof of Theorem 7:

Proof: With $n = k$ and $n^* = k^*$, it follows $n - k^* = k - n^*$, thus we yield the code rate of algorithm pair $\langle \mathsf{SS}_{\mathsf{IDES}}, \mathsf{Rec}_{\mathsf{IDES}}\rangle$ (works as an $[n, k^*, d]$ linear code) $R = k^*/n = 1 - (k - n^*)/n$. Define $(d-1)/n = \varepsilon_d' = 2k^*\varepsilon_{ss}/n$, meaning that $d = 2n^*\varepsilon_{ss}+1$. It should be noted that $R = 1 - (d-1)/n$ by letting $d = k - n^* + 1$. Then, we can reduce the value of d to any $t \geq t_{\min}$ for arbitrary value of $k - n^* = n - k^* \geq 2$. It follows that $d \geq n - k^* + 1 \geq 3$, which means it can tolerate arbitrary number of error equal to (for $t_{\min} = 1$)

$$n^*\varepsilon_{ss} = \frac{d-1}{2} \geq \left\lfloor \frac{d-1}{2} \right\rfloor = t = 1 = t_{\min}$$

with the probability of success in decoding (recover the input w) is at least $1 - 2^{-(d-1)} = 0.75$ if n is large enough and $\lfloor k^* h_2(\varepsilon_{ss})\rfloor = k - n^*$ by Theorem 4. Clearly the value of $t = t_{\min} = 1$ is minimum, therefore d must be minimum as well. □

APPENDIX D
PROOFS OF THE PROPOSITION FOR IDES-SECURE SKETCH

Proof of Proposition 2:

Proof: We first argue on the **correctness**. Follow the correctness result obtained in Theorem 4 (Eq. 3), for $k - n^* \geq 1$, we have

$$\begin{aligned}&\Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) = w\right]\\ &= 1 - \Pr\left[\mathsf{Rec}_{\mathsf{IDES}}(\mathsf{SS}_{\mathsf{IDES}}(w, N, \varepsilon_{ss}), w', N, \varepsilon_{rec}) \neq w\right]\\ &= 1 - 2^{-(k-n^*)} \geq 1/2.\end{aligned}$$

We then argue on the **security**. Follow Theorem 6, for large enough k^* and n s.t. $k^*/n \approx 1$ and $\lfloor k^* h_2(\varepsilon_{ss})\rfloor > k - n^*$, it is suffice to use $k - n^* = 1$ to show the worst-case security with maximum distinguishibility error $\beta' = \exp(-2n(\varepsilon_d')^2) < 2^{-(k-n^*)} = 1 - 2^{-(k-n^*)} = 1/2$. Letting $\mathrm{H}_{t,\infty}^{\mathrm{fuzz}}(W) = -\log(2^{-(k-n^*)})$, it follows

$$\mathrm{H}_{t,\infty}^{\mathrm{fuzz}}(W) = -\log(2^{-(k-n^*)}) = -\log(1 - 2^{-(k-n^*)}) = 1.$$

Then, let $\gamma = 2^{-(k-n^*)}$, one can use the notion of average min-entropy, i.e., $\tilde{\mathrm{H}}_\infty(W \mid ss) = \tilde{m}$ and yield

$$\tilde{\mathrm{H}}_\infty(W \mid ss) = \tilde{m} \geq -\log(1-\gamma) = \mathrm{H}_{t,\infty}^{\mathrm{fuzz}}(W) = 1.$$

For $\tilde{\mathrm{H}}_\infty(W \mid ss) = \tilde{m} \geq 0$, above equation can be simplified as $\tilde{\mathrm{H}}_\infty(W \mid ss) = \tilde{m} + \log(1-\gamma) \geq \mathrm{H}_{t,\infty}^{\mathrm{fuzz}}(W) = 0$, thus it follows that with $\mathrm{H}_{t,\infty}^{\mathrm{fuzz}}(W) = 1$ and $\gamma = 1/2$:

$$\tilde{m} = \tilde{\mathrm{H}}_\infty(W \mid ss) \geq \mathrm{H}_{t,\infty}^{\mathrm{fuzz}}(W) - \log(1-\gamma) = 0,$$

meets the best possible security bound (see Table 1 or Eq. 1) and complete the prove. □

REFERENCES

[1] S. N. Porter, "A password extension for improved human factors," *Comput. Secur.*, vol. 1, no. 1, pp. 54–56, Jan. 1982.
[2] N. Frykholm and A. Juels, "Error-tolerant password recovery," in *Proc. 8th ACM Conf. Comput. Commun. Secur. (CCS)*, 2001, pp. 1–9.
[3] C. Ellison, C. Hall, R. Milbert, and B. Schneier, "Protecting secret keys with personal entropy," *Future Gener. Comput. Syst.*, vol. 16, no. 4, pp. 311–318, Feb. 2000.
[4] A. Juels and M. Sudan, "A fuzzy vault scheme," *Des., Codes Cryptogr.*, vol. 38, no. 2, pp. 237–257, Feb. 2006.
[5] C. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM J. Comput.*, vol. 17, no. 2, pp. 210–229, Apr. 1988.
[6] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.* Berlin, Germany: Springer, 2004, pp. 523–540.
[7] B. Fuller, X. Meng, and L. Reyzin, "Computational fuzzy extractors," in *Proc. Int. Conf. Theory Appl. Cryptol. Inf. Secur.* Berlin, Germany: Springer, 2013, pp. 174–193.
[8] J. Woodage, R. Chatterjee, Y. Dodis, A. Juels, and T. Ristenpart, "A new distribution-sensitive secure sketch and popularity-proportional hashing," in *Proc. Annu. Int. Cryptol. Conf.* Cham, Switzerland: Springer, 2017, pp. 682–710.
[9] B. Fuller, L. Reyzin, and A. Smith, "When are fuzzy extractors possible?" in *Proc. 22nd Int. Conf. Theory Appl. Cryptol. Inf. Secur.* Hanoi, Vietnam: Springer, Dec. 2016, pp. 277–306.
[10] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. 6th ACM Conf. Comput. Commun. Secur. (CCS)*, 1999, pp. 28–36.
[11] W. Yang, J. Hu, and S. Wang, "A delaunay triangle group based fuzzy vault with cancellability," in *Proc. 6th Int. Congr. Image Signal Process. (CISP)*, Dec. 2013, pp. 1676–1681.
[12] W. Yang, J. Hu, and S. Wang, "A delaunay quadrangle-based fingerprint authentication system with template protection using topology code for local registration and security enhancement," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1179–1192, Jul. 2014.
[13] W. Yang, J. Hu, S. Wang, and M. Stojmenovic, "An alignment-free fingerprint bio-cryptosystem based on modified Voronoi neighbor structures," *Pattern Recognit.*, vol. 47, no. 3, pp. 1309–1320, Mar. 2014.
[14] P. Li, X. Yang, K. Cao, X. Tao, R. Wang, and J. Tian, "An alignment-free fingerprint cryptosystem based on fuzzy vault scheme," *J. Netw. Comput. Appl.*, vol. 33, no. 3, pp. 207–220, May 2010.
[15] W. Ponce-Hernandez, R. Blanco-Gonzalo, J. Liu-Jimenez, and R. Sanchez-Reillo, "Fuzzy vault scheme based on fixed-length templates applied to dynamic signature verification," *IEEE Access*, vol. 8, pp. 11152–11164, 2020.
[16] R. K. Mahendran and P. Velusamy, "A secure fuzzy extractor based biometric key authentication scheme for body sensor network in Internet of medical things," *Comput. Commun.*, vol. 153, pp. 545–552, Mar. 2020.
[17] S. Pirbhulal, P. Shang, W. Wu, A. K. Sangaiah, O. W. Samuel, and G. Li, "Fuzzy vault-based biometric security method for tele-health monitoring systems," *Comput. Electr. Eng.*, vol. 71, pp. 546–557, Oct. 2018.
[18] L. A. Elrefaei and A. M. Al-Mohammadi, "Machine vision gait-based biometric cryptosystem using a fuzzy commitment scheme," *J. King Saud Univ.-Comput. Inf. Sci.*, early access, Nov. 2, 2019, doi: 10.1016/j.jksuci.2019.10.011.

[19] S. Barman, A. K. Das, D. Samanta, S. Chattopadhyay, J. J. P. C. Rodrigues, and Y. Park, "Provably secure multi-server authentication protocol using fuzzy commitment," *IEEE Access*, vol. 6, pp. 38578–38594, 2018.
[20] S. Barman, H. P. H. Shum, S. Chattopadhyay, and D. Samanta, "A secure authentication protocol for multi-server-based E-healthcare using a fuzzy commitment scheme," *IEEE Access*, vol. 7, pp. 12557–12574, 2019.
[21] P. Mihăilescu, A. Munk, and B. Tams, "The fuzzy vault for fingerprints is vulnerable to brute force attack," in *Proc. Biometrics Electron. Signatures (BIOSIG)*, A. Brömme, C. Busch, and D. Hühnlein, Eds. Bonn, Germany: Gesellschaft für Informatik e.V., 2009, pp. 43–54.
[22] J. Merkle, M. Niesing, M. Schwaiger, H. Ihmor, and U. Korte, "Security capacity of the fuzzy fingerprint vault," *Int. J. Adv. Secur.*, vol. 3, no. 3, pp. 146–168, 2010.
[23] B. Tams, P. Mihăilescu, and A. Munk, "Security considerations in minutiae-based fuzzy vaults," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 985–998, May 2015.
[24] Y. Dodis and D. Wichs, "Non-malleable extractors and symmetric key cryptography from weak secrets," in *Proc. 41st Annu. ACM Symp. Theory Comput.*, 2009, pp. 601–610.
[25] J. Daugman, "Probing the uniqueness and randomness of IrisCodes: Results from 200 billion iris pair comparisons," *Proc. IEEE*, vol. 94, no. 11, pp. 1927–1935, Nov. 2006.
[26] B. Fuller, S. Simhadri, and J. Steel, "Reusable authentication from the iris," Cryptol. ePrint Arch., Tech. Rep. 2017/1177, 2017. [Online]. Available: https://eprint.iacr.org/2017/1177
[27] S. Simhadri, J. Steel, and B. Fuller, "Cryptographic authentication from the iris," in *Proc. Int. Conf. Inf. Secur.* Cham, Switzerland: Springer, 2019, pp. 465–485.
[28] M. Blanton and W. M. Hudelson, "Biometric-based non-transferable anonymous credentials," in *Proc. Int. Conf. Inf. Commun. Secur.* Berlin, Germany: Springer, 2009, pp. 165–180.
[29] V. Guruswami, *List Decoding of Error-Correcting Codes: Winning Thesis of the 2002 ACM Doctoral Dissertation Competition*, vol. 3282. Berlin, Germany: Springer, 2004.
[30] R. L. Rivest, "Symmetric encryption via keyrings and ECC," Northernmost Crypto Workshop, Longyearbyen, Norway, Midnight Lect., 2016.
[31] M. S. Charikar, "Similarity estimation techniques from rounding algorithms," in *Proc. 34th Annu. ACM Symp. Theory Comput.*, 2002, pp. 380–388.
[32] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *Proc. Vldb*, 1999, vol. 99, no. 6, pp. 518–529.
[33] E. Berlekamp, R. McEliece, and H. van Tilborg, "On the inherent intractability of certain coding problems (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 384–386, May 1978.
[34] E. N. Gilbert, "A comparison of signalling alphabets," *Bell Syst. Tech. J.*, vol. 31, no. 3, pp. 504–522, May 1952.
[35] S. Ling and C. Xing, *Coding Theory: A First Course*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[36] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*. Amsterdam, The Netherlands: Elsevier, 1977.
[37] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.

YEN-LUNG LAI received the B.Sc. degree in physics from University Tunku Abdul Rahman (UTAR), Malaysia, in 2015. He is currently pursuing the Ph.D. degree with Monash University Malaysia. His research interests include information security and biometrics.

ZHE JIN (Member, IEEE) received the B.I.T. degree (Hons.) in software engineering and the M.Sc. (I.T.) degree from Multimedia University, Malaysia, in 2007 and 2011, respectively, and the Ph.D. degree in engineering from University Tunku Abdul Rahman Malaysia, in 2016. He is currently a Senior Lecturer with the School of Information Technology, MONASH University, Malaysia Campus. He has published more than 40 refereed journals and conference articles, including the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, *DSC*, and *Pattern Recognition*. His research interests include biometric security, computer vision, and machine learning. He was awarded Marie Sk?odowska-Curie Research Exchange Fellowship and visited the University of Salzburg, Austria, and the University of Sassari, Italy, respectively, as a Visiting Scholar under the EU Project IDENTITY 690907.