情報理論的に安全な検索可能暗号

吉澤 貴博 † 渡邉 洋平 † 四方 順司 †‡

† 横浜国立大学 大学院環境情報学府/研究院
‡ 横浜国立大学 先端科学高等研究院
240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-7
yoshizawa-takahiro-my@ynu.jp, watanabe-yohei-xs@ynu.jp, shikata@ynu.ac.jp

あらまし 検索可能暗号は，復号することなく暗号化された文書の検索を行うことができる方式であり，計算量的安全性の枠組みで盛んに研究が進められてきた．近年，様々な個人情報の電子化が進んでおり，中には長期的安全性を必要とするものも増えてきているが，計算量的安全性では長期的安全性を保証することは難しい．本稿では，情報理論的安全性に基づく共通鍵検索可能暗号方式を提案する．具体的には，モデル，安全性の定式化を行い，鍵長の下界を導出し，その下界の等号を満たす最適な構成法を提案する．本方式により，長期的安全性を保証しつつデータの情報を漏らさずに検索が可能となる．

Unconditionally Secure Searchable Encryption

Takahiro Yoshizawa† Yohei Watanabe† Junji Shikata†‡

†Graduate School of Environment and Information Sciences, Yokohama National University
‡Institute of Advanced Sciences, Yokohama National University
79-7 Tokiwadai, Hodogaya-ku, Yokohama 240-8501, Japan
yoshizawa-takahiro-my@ynu.jp, watanabe-yohei-xs@ynu.jp, shikata@ynu.ac.jp

Abstract Recently, various information including highly senstive one is computerized and used in a wide range of areas such as cloud computing. Searchable encryption, which is expected to use in cloud computing, has been well studied in terms of computational security. Although some sensitive information such as genome data requires a long-term security (e.g. at least 50 years), such computationally secure schemes can guarantee the confidentiality of data for only decades at most. Therefore, in this paper we propose searchable encryption with unconditional security. Specifically, we first define a model and security of unconditionally secure searchable enctyption. Next, we derive a tight lower bound on the secret-key size, and propose an efficient construction in terms of its secret-key size.

1 はじめに

検索可能暗号はドキュメントを暗号化したまま検索可能な方式であり，2000 年代から研究が進められ [8, 11]，クラウドコンピューティング等での利用が期待されている．検索可能暗号には大きく分けて公開鍵型（Public key Encryption with Keyword Search: PEKS）[1, 3] と共通鍵型（Symmetric Searchable Encryption: SSE）[8, 11] の 2 種類が存在し，本稿では後者を取り扱う．SSE は，ドキュメントの暗号化と検索用のタグ生成に共通の鍵が使われる方式であり，

ユーザはキーワードに対応するタグを鍵を用いて生成し，サーバ側はそのタグを用いて対応したドキュメントを検索可能である．一方で，サーバに対して，検索されたキーワードの情報，また保存してある（暗号化された）ドキュメントの情報を漏らさないという安全性を持つ．

公開鍵暗号等の計算量的安全性を有する暗号技術は利便性が高い一方で，長期的（数十年以上）な安全性を保証することはできない．実際，NIST の発表 [4] によると，2,048 ビットの RSA 暗号であっても，2030 年までしか安全性を保証できないことがわかっている．

近年，クラウドコンピューティング環境で利用される，または利用が期待されている情報の中には長期的な安全性が必要な情報が存在する．例えば，ゲノム解析等の技術もクラウドコンピューティング環境での利用が期待されており [2, 12]，ゲノムデータは長期的な安全性が必要な情報の代表例である．更には，我が国ではマイナンバー制の導入が 2015 年 10 月より導入され，マイナンバー漏洩時には再発行の処置が取られるものの，実際に漏洩にいち早く気付き迅速に対応できる保証はない．もちろん公開鍵暗号等で暗号化しておき，その安全性が危殆化する前に漏洩に気付き再発行すれば対処可能ではあるが，最初から長期的に安全性を保証できる暗号化方式で暗号化しておけば，漏洩時のリスクを更に軽減することができる．

PEKS，SSE 両方を含む既存の検索可能暗号は全て計算量的安全性の枠組みで研究されており，長期的な安全性を保証することは難しい．そこで本稿では，情報理論的安全性を満たす検索可能暗号を提案する．具体的には，情報理論的に安全な SSE のモデル，安全性の定式化を行い，鍵長の下界を導出する．また，その鍵長の下界を満たす最適な構成法も提案する．

関連研究． SSE は，[8, 11] 等の初期研究段階を経て，[9] によって現在の SSE の標準的なモデル，安全性の定式化がなされた．[10] では，[9] に比べ検索に必要な計算量を改善した方式を提案している．また，[6] では検索に必要な計算量と暗号化ドキュメントのデータサイズの間にはトレードオフが存在することを示している．

その他，付加的な機能をもつ SSE もいくつか提案されている．[7] では，サーバが正しい検索結果を返しているか検証可能な方式を提案している．[5] では，あるキーワードだけでなく，それと関連するキーワードも検索可能な方式を提案している．これらの付加機能については本稿では議論しない．

2 モデル

本節では情報理論的に安全な検索可能暗号のモデルについて述べる．まず簡単に本方式の流れを説明する．ドキュメントを保存するユーザは保存したいドキュメントを暗号化し，それらはサーバに送られ，保存される．各ドキュメントにはキーワードが 1 つ以上紐付いているものとする．また，検索を行うユーザは検索したいキーワードのタグを生成，サーバに安全な通信路を通じて送る．サーバはそのタグを用いて検索を行い，対応するキーワードと紐付いたドキュメントの暗号文（厳密には，そのインデックス）を出力する．ドキュメントを保存するユーザと検索するユーザは同一でも異なっていても良く，異なる場合は秘密鍵をあらかじめ（ドキュメントが暗号化される前に）共有しているものとする．この時，サーバは必ず正しくアルゴリズムを実行し，またサーバ内に保存されている暗号化されたドキュメントを改ざんすることは無いことを仮定する．ただし，暗号化されているドキュメントや検索に用いられたタグから元のドキュメントやキーワードが何であるかを推測しようとするものとする．このようなモデルはしばしば honest-but-curious モデルと呼ばれる[1]．以下，サーバに保存されるドキュメントの数を n，プロトコル内で検索される（異なる）キーワードの数を高々 τ とする．

まず，本稿で用いる記法を定義する．

記法． $\mathcal{M}$ をドキュメントの集合とし，$\mathcal{W} = \{w_1, \ldots, w_\tau\}$（すなわち $|\mathcal{W}| = \tau$）をキーワードの

[1] サーバが，サーバ内に保存してあるドキュメントを改ざんすることは無いが，検索結果を改ざんする攻撃，すなわち検索されたキーワードに紐付いていないドキュメントを検索結果に含めたりする攻撃を仮定するモデルを，semi-honest-but-curious モデル [7] という．

集合とし，各 w_i $(1 \leq i \leq \tau)$ は辞書式順序に従うものとする．$\mathcal{K}$ を秘密鍵の集合，$\mathcal{C}$ を暗号文の集合，$\mathcal{T}$ をタグの集合とする．任意の要素数 n の部分集合 $D = \{m_1, \ldots, m_n\} \subseteq \mathcal{M}$ に対して，$\mathcal{I} = \{id(m_1), \ldots, id(m_n)\}$ をインデックスの集合とし，各 $id(m_i)$ $(1 \leq i \leq n)$ は m_i のインデックスを表し，ドキュメントとは独立に選ばれる（ドキュメントの情報を一切含まない）．また，写像 $f : \mathcal{M} \to 2^{\mathcal{W}}$ を考え，$\mathcal{W}(m) = f(m)$ とする．すなわち $\mathcal{W}(m)$ は $m \in \mathcal{M}$ にひもづくキーワードの集合を表しており，またここで $2^{\mathcal{W}}$ は $\mathcal{W}$ のベキ集合を表す．任意の $D = \{m_1, \ldots, m_n\} \subseteq \mathcal{M}$，任意の $w \in \mathcal{W}$ に対して，$D(w) = \{id(m) \in \mathcal{I} \mid w \in \mathcal{W}(m), m \in \mathcal{M}\}$ とする．すなわち，D の中でキーワード w と紐付いているドキュメントの（インデックスの）集合を表している．

(τ, n)-SSE Π を以下のように定義する．

定義 1 ((τ, n)-SSE)**.** (τ, n)-*SSE* Π は以下の 5 つのアルゴリズム *(Gen, Enc, Tag, Search, Dec)* と 5 つの有限集合 $\mathcal{M}, \mathcal{K}, \mathcal{W}, \mathcal{C}, \mathcal{T}, \mathcal{I}$ からなる．*Gen* は確率的アルゴリズム，*Enc, Tag, Search, Dec* は確定的アルゴリズムである．

1. $k \leftarrow Gen(1^\lambda, \tau, n)$*:* セキュリティパラメータ λ とキーワードの数 τ，ドキュメントの数 n を入力として秘密鍵 $k \in \mathcal{K}$ を出力する．
2. $(\mathcal{I}, ED) \leftarrow Enc(k, D)$*:* 秘密鍵 k とドキュメント $D = \{m_1, \ldots, m_n\} \subseteq \mathcal{M}$ を入力としてインデックス $\mathcal{I} = \{id(m_1), \ldots, id(m_n)\}$ と暗号文 $ED = \{c_1, \ldots, c_n\} \subseteq \mathcal{C}$ を出力する．ただし，c_i を m_i の暗号文とし，鍵 k のもとで平文 m_i から c_i が生成される．
3. $t \leftarrow Tag(k, w)$*:* 秘密鍵 k とキーワード $w \in \mathcal{W}$ を入力としてタグ $t \in \mathcal{T}$ を出力する．
4. $\mathcal{X} \leftarrow Search(\mathcal{I}, t)$*:* インデックス $\mathcal{I}$ とタグ t を入力として，t に対応するキーワードと紐付いたドキュメントのインデックスの集合 $\mathcal{X} \subseteq \mathcal{I}$ を出力する．
5. $m_i \leftarrow Dec(k, c_i)$*:* 秘密鍵 k と暗号文 c_i を入力としてドキュメント m_i を出力する．

(τ, n)-SSE Π は以下の search correctness と decryption correctness を必要とする．

- Search correctness: 全ての $k \leftarrow Gen(1^\lambda, \tau, n)$，全ての $D \subseteq \mathcal{M}$ と $\mathcal{I}$，全ての $w \in \mathcal{W}$ に対して，

$$\Pr\left[D(w) = \mathcal{X} : \mathcal{X} \leftarrow Search(\mathcal{I}, Tag(k, w))\right] = 1.$$

  これはすなわち，任意のキーワードに対して必ず正しい検索結果が出力されることを意味する．

- Decryption correctness: 全ての $k \leftarrow Gen(1^\lambda, \tau, n)$ と全ての $D = \{m_1, \ldots, m_n\} \subseteq \mathcal{M}$，全ての $ED = \{c_1, \ldots, c_n\} \leftarrow Enc(k, D)$，全ての $i \in \{1, \ldots, n\}$ に対して，

$$\Pr\left[m_i \leftarrow Dec(k, c_i)\right] = 1.$$

  これはすなわち，正しく暗号化された任意の暗号文に対して，必ず正しいドキュメントを復号出来ることを意味する．

安全性として，サーバに対する完全秘匿性を考える．M_i $(1 \leq i \leq n)$ を $\mathcal{M}$ から独立に同一の確率分布に従う確率変数 (i.i.d.) とし，C_i は M_i の暗号文の確率変数であり，$\mathcal{C}$ に値をとる．同様に，W_i $(1 \leq i \leq \tau)$ を $\mathcal{W}$ から独立に同一の確率分布に従う確率変数 (i.i.d.) とし，T_i は W_i のタグの確率変数であり，$\mathcal{T}$ に値をとる．I は $\mathcal{I}$ に値をとる確率変数である．以下，$Z := (I, T_1, \ldots, T_\tau, C_1, \ldots, C_n)$ とする．

定義 2 (安全性)**.** 以下の条件を満たすとき，(τ, n)-*SSE* Π は (τ, n)-*secure* であるという．

$$H(W_1, \ldots, W_\tau, M_1, \ldots, M_n) = H(W_1, \ldots, W_\tau, M_1, \ldots, M_n | Z).$$

上記の定義は，インデックス，タグ，暗号文からキーワードとドキュメントの情報が何も漏れないことを示している．

3 鍵長の下界

(n, τ)-secure SSE Π の鍵長の下界を導出する．ドキュメントとキーワードの確率変数は独立

に同一の確率分布に従うので，以下では簡単に $H(M) = H(M_1) = H(M_2) = \cdots = H(M_n)$, $H(W) = H(W_1) = H(W_2) = \cdots = H(W_\tau)$ と書く．また，K は $\mathcal{K}$ に値をとる確率変数とする．

定理 1. 任意の (n,τ)-*secure SSE* Π に対して，以下が成り立つ．

$$H(K) \geq nH(M) + \tau H(W).$$

証明. まず，以下の補題を示す．

補題 1. $Tag : \mathcal{K} \times \mathcal{W} \to \mathcal{T}$ を $k \in K$ をインデックスとした $\{Tag_k\}_{k\in K}$ と表す．この時，$\forall k \in K$ に対して，$Tag_k : \mathcal{W} \to \mathcal{T}$ は単射である．

証明. $w \neq w'$ であり，$t \leftarrow Tag_k(w) \in \mathcal{T}$，$t' \leftarrow Tag_k(w') \in \mathcal{T}$，$t = t'$ を満たす $w, w' \in \mathcal{W}$ の存在を仮定する．このとき，明らかに search correctness に矛盾する．従って，任意の $w, w' \in \mathcal{W}$ $(w \neq w')$ に対して $Tag_k(w) \neq Tag_k(w')$ であり，Tag_k は単射である． □

補題 1 より，$Tag_k : \mathcal{W} \to Tag_k(\mathcal{T})$ は全単射であり，すなわち逆写像が計算可能なため，次の補題を得る．

補題 2. (n,τ)-*secure SSE* Π において，$H(W|K,T) = 0$.

従って，

$$\begin{aligned}
&H(K)\\
&= H(K) + H(M_1,\ldots,M_n \mid Z,K)\\
&\quad + H(W_1,\ldots,W_\tau \mid Z,K,M_1,\ldots,M_n) \quad (1)\\
&\geq H(K \mid Z) + H(M_1,\ldots,M_n \mid Z,K)\\
&\quad + H(W_1,\ldots,W_\tau \mid Z,K,M_1,\ldots,M_n)\\
&= H(W_1,\ldots,W_\tau,M_1,\ldots,M_n,K \mid Z)\\
&\geq H(W_1,\ldots,W_\tau,M_1,\ldots,M_n \mid Z)\\
&= H(W_1,\ldots,W_\tau,M_1,\ldots,M_n) \quad (2)\\
&= \tau H(W) + nH(M)
\end{aligned}$$

(1) は補題 2 と decryption correctness から従う．
(2) は安全性（定義 2）から従う． □

4 節で述べる構成法は定理 1 の不等式における等号を満たす．すなわち，上記の下界はタイトである．従って，次のような (τ,n)-SSE の構成法の最適性を定義する．

定義 3. (τ,n)-*secure SSE* Π の構成法が定理 *1* の等号を満たす時，その構成法は最適であるという．

4 最適構成法

以下では，$\mathbb{F}_{2^\lambda}$ を要素数 $2^\lambda (> n)$ 個の有限体とし，また τ を素数とし，$\mathbb{F}_\tau$ を要素数 τ 個の有限体とする．$\mathcal{M} = \mathcal{C} = \mathbb{F}_{2^\lambda}$，$\mathcal{W} = \mathcal{T} = \mathbb{F}_\tau$ とする．また，$\mathcal{I} = \prod_{i=1}^{n} \left(\{0,1\}^{\lfloor \log n \rfloor + 1} \times \mathcal{Y}_i\right)$ であり，$\mathcal{Y}_i$ は以下で定義する．また，$\Sigma = \{\sigma\}$ を $\mathbb{F}_\tau = \{0,1,\ldots,\tau-1\}$ 上の置換からなる集合とする．すなわち，

$$\sigma = \begin{pmatrix} 0 & 1 & \cdots & \tau-1 \\ \sigma(0) & \sigma(1) & \cdots & \sigma(\tau-1) \end{pmatrix}.$$

構成法は以下の通り．

- $k \leftarrow Gen(1^\lambda, \tau, n)$: 一般性を失わずに，$\mathcal{W} = \{w_1,\ldots,w_\tau\}$ の各元はそれぞれ $\mathbb{F}_\tau = \{0,1,\ldots,\tau-1\}$ の各元に $w_i \mapsto i-1$ の対応で符号化されているとする．この対応で両者を同一視し，$w_i \in \mathbb{F}_\tau$ と書く．集合 Σ よりランダムに置換 σ を選ぶ．任意の $i \in \{1,\ldots,\tau\}$ に対して，$a_i = \sigma(w_i)$ とし，$(b_1,\ldots,b_n) \in \mathbb{F}_{2^\lambda}^n$ を一様ランダムに選ぶ．
  $k = (a_1,\ldots,a_\tau,b_1,\ldots,b_n)$ を出力する．
- $(\mathcal{I}, ED) \leftarrow Enc(k, m_1,\ldots,m_n)$: 次に，各 $j \in \{1,\ldots,n\}$ に対して，以下を行う．
  (a) m_j の暗号文 $CT_j = m_j + b_j$ を計算する．
  (b) $\mathcal{Y}_j = \{a_i \mid w_i \in \mathcal{W}(m_j)\}$ を計算する[2]．

[2]鍵を持つユーザならば，$(a_1,\ldots,a_\tau)$ の並びから，各 a_i がどのキーワードに対応しているか知ることができることに留意する．

(c) $c_j = (j, c_j)$，また $id(m_j) = (j, \mathcal{Y}_j)$ とする．ここで $j \in \{0,1\}^{\lfloor \log n \rfloor + 1}$ とする．

$\mathcal{I} = \{id(m_1), \ldots, id(m_n)\}$ と $ED = \{c_1, \ldots, c_n\}$ を出力する．

– $t \leftarrow Tag(k, w)$: w の $\mathcal{W}$ における順序を i 番目とし，$t = a_i$ を出力する．

– $\mathcal{X} \leftarrow Search(\mathcal{I}, t)$: $\mathcal{X} = \{id(m_\ell) \in \mathcal{I}\ (1 \leq \ell \leq n) \mid t \in \mathcal{Y}_\ell\}$ を出力する．

– $m_i \leftarrow Dec(k, c_i)$: $c_i = (i, CT_i)$ とする．$m_i = CT_i - b_i$ を出力する．

定理 2. 上記の Π の構成法は (τ, n)-*secure* であり，かつ最適である．

証明. まず, search correctness と decryption correctness を満たしていることを示す．上記の構成法では，*Gen* であらかじめキーワードに対応した $a_1, \ldots, a_\tau$ を生成しておき，*Enc* でタグをドキュメントに紐づけておく．*Search* ではキーワード w に対応した $\sigma(w)$ をタグとして検索する．σ は置換のため，異なるキーワードに対して同じタグが生成されることは無い．従って，search correctness を満たす．

ドキュメントの暗号化はワンタイムパッドを用いて行われるため，decryption correctness を満たす．

次に，上記構成法が定義 2 を満たすことを示す．攻撃者は n 個の暗号化されたドキュメント $ED = \{c_1, \ldots, c_n\}$，各インデックス $id(m_i) = (i, \mathcal{Y}_i)\ (1 \leq i \leq n)$，また高々$\tau$ 個のタグ $t_1, \ldots, t_\tau$ を知ることができる．しかし，$t_j\ (1 \leq j \leq \tau)$ はランダムに置換された値，$c_i\ (1 \leq i \leq n)$ はワンタイムパッドであるため，キーワード，ドキュメントに関する情報を得ることが出来ない．従って，$H(W_1, \ldots, W_\tau, M_1, \ldots, M_n) = H(W_1, \ldots, W_\tau, M_1, \ldots, M_n | Z)$.

最後に上記構成法が最適であることを示す．秘密鍵は $\tau \log \tau + n \log 2^\lambda = \tau \log \tau + n\lambda$ ビットであるため，鍵長の下界の等号を満たしており，最適である． □

5 まとめと今後の課題

情報理論的に安全な検索可能暗号を提案した．具体的には情報理論的に安全な検索可能暗号のモデル，安全性定義を提案し，秘密鍵のサイズの下界を導出した．加えて，鍵長の下界を満たす最適な構成法を提案した．

謝辞. 本研究は JSPS 科研費 15H02710 の助成，および，文部科学省国立大学改革強化推進事業の支援を受けたものです．第二著者は，JSPS 科研費 25·3998 の助成を受けています．

参考文献

[1] M. Abdalla, M. Bellare, D. Catalano, E Kiltz, T. Kohno, T. Lange, J. Malonelee, G. Neven, P. Paillier and H. Shi: Searchable encryption revisited: Consistency properties, reration to anonymous IBE, and extensions. In: J.Cryptology, vol. 21, no. 3, pp.350–391, Springer (2008)

[2] E. Ayday, E. De Cristofaro, J. Hubaux, and G. Tsudik: The chills and thrills of whole genome sequencing. In: Computer, 99, p.1, IEEE (2013) The full version is available at `http://arxiv.org/abs/1306.1264`.

[3] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano: Public key encryption with keyword search. In: EUROCRYPT 2004, LNCS 3027 pp.506–522, Springer (2004)

[4] E. Barker, W. Barker, W. Burr, W. Polk, and M. Smid: Recommendation for key management — part 1: General (revision 3). In: NIST Special Publication 800-57. (2012) Available at `http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57_part1_rev3_general.pdf`.

[5] D. Cash, S. Jarecki and C. Jutla: Highly-scalable searchable symmetric encry ption

with support for boolean queries. In: CRYPTO 2013, Lecture Notes in Computer Science Volume 8042, pp 353–373, Springer (2013)

[6] D. Cash and S Tessaro: The locality of searchable symmetric encryption. In: EUROCRYPT 2014, Lecture Notes in Computer Science Volume 8441, pp.351–368, Springer (2014)

[7] Q. Chai and G. Gong: Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers. In: IEEE International Conference on Communications (ICC), pp.917–922, IEEE (2012)

[8] Y. Chang and M. Mitzenmacher: Privacy preserving keyword searches on remote encrypted data. In: Applied Cryptography and Network Security(ACNS'05), volume 3531 of Lecture Notes in Computer Science, pp.442–455, Springer (2005)

[9] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky: Searchable symmetric encryption: improved definitions and efficient constructions. In: ACM CCS'06, pp.79–88, ACM (2006)

[10] S. Kamara and C. Papamanthou: Parallel and dynamic searchable symmetric encryption. In: Financial Cryptography and Data Security, Lecture Notes in Computer Science Volume 7859, pp.258–274, Springer (2013)

[11] D. Song, D. Wagner and A. Perrig: Practical techniques for searching on encrypted data. In: IEEE Symposium on Research in Security and Privacy, pp.44–55, IEEE (2000)

[12] The Presdental Commission for the Study of Bioethical Issues: Privacy and progress in whole genome sequencing. In: President 's Bioethics Commission Releases Report on Genomics and Privacy. (2012) Available at `http://bioethics.gov/sites/default/files/PrivacyProgress508.pdf`.