様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　１５　日現在

機関番号：　１３６０１
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2016
課題番号：　２６３３０１５４

研究課題名（和文）階層型ブルームフィルタを用いた暗号化データに対する柔軟で効率的な検索法の開発

研究課題名（英文）On developing flexible and efficient search schemes on encrypted data using a hierarchical Bloom filter

研究代表者

山本　博章（YAMAMOTO, Hiroaki）

信州大学・学術研究院工学系・教授

研究者番号：１０１８２６４３

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：検索可能暗号とは，暗号データを暗号化したまま検索する技術である。本研究では，階層型ブルームフィルタを用いた新たな検索可能暗号として，以下の成果を得た。
(1)キーワードをベースとした階層型ブルームフィルタを導入し，従来法よりも時間及び空間において効率的な検索可能暗号を開発した。(2)通常の部分文字列検索のデータ構造を暗号化することにより，部分文字列検索可能な検索可能暗号を開発した。(3) ブーリアン検索への応用を示した。

研究成果の概要（英文）：Searchable symmetric encryption (SSE) is a search scheme to search encrypted documents without decrypting them. In this research, we gave the following results.
(1) We showed an efficient SSE using a Bloom filter based on keywords. The proposed scheme is efficient in time and space because it uses a simpler data structure than that of the existing schemes. (2) We showed a new substring SSE which can search for all substrings contained in a given text. The proposed substring SSE is constructed by an encrypted DAWG. (3) We showed a boolean SSE based on a Bloom filter.

研究分野：　情報工学

キーワード：　検索可能暗号　プライバシー保護　情報検索

１．研究開始当初の背景

情報通信技術の発展により、クラウドなどネットを利用したサービスが拡大している。そのため、情報セキュリティの強化が強く求められるようになってきた。データベースのサービスにおいても、ユーザが外部サーバにデータを保存し、データを利用する形態が増えてきている。しかし、セキュリティ上の問題点として、サーバの管理者がそこに保存されているデータへのアクセス権を持つとは限らないことがある。ユーザにとっては、データへのアクセス権限がない管理者にはその中身については一切知られたくない。

データの中身を秘匿する最も効果的な方法はデータの暗号化である。したがって、データを暗号化したまま検索を行うことができれば、プライバシーや機密データを保護した安全な検索システムを実現することができる（このような暗号化データの検索法を**検索可能暗号**と言う）。このため、効率的な検索可能暗号の開発に向けた研究が活発に行われるようになってきた。検索可能暗号では，ユーザ（クライアント）がデータの所有者で、サーバ上のデータはすべてユーザの秘密鍵で暗号化されており、サーバが暗号化データ上で検索を行う。図 1 に検索可能暗号の概略を示す。ここでは，ユーザが検索用のキーワード暗号化しサーバに送る。サーバは，暗号化キーワードと暗号化索引を使って検索し，結果をユーザに返す。

検索可能暗号では，安全性を維持したまま柔軟で効率的な検索を提供することが重要であるが，すべてを満足することはなかなか難しい。しかしながら，実用性の高い検索法を目指し，多くの研究がなされている。

２．研究の目的

前記の背景の中，本研究課題では、代表者が開発した階層型ブルームフィルタに基づいた検索可能暗号を発展させ，次の点について明らかにする。

(1)　階層型ブルームフィルタの空間効率の改善と暗号化ブーリアン検索への拡張： 階層型ブルームフィルタによる暗号化索引は、文書を木構造で管理するため、登録されるキーワードの個数が根に向かうほど少なくなる。本課題では、この性質を利用し、キーワード数に応じてブルームフィルタのサイズを変える「可変長階層型ブルームフィルタ」を設計することによって空間効率を改善する。さらに時間効率を改善するため、論理演算が利用できる暗号化ブーリアン検索へ拡張する。

(2)　階層型ブルームフィルタを用いたキーワードに基づく暗号化索引の開発：従来法として，代表者は、ドキュメントに基づいた暗号化索引の構成法を示した。本課題では、登録キーワード数の対数時間での検索を実現するため、キーワードに基づく暗号化索引を開発する。

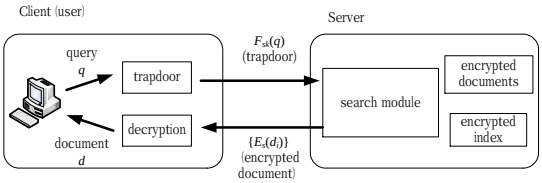

図 1 検索可能暗号の概略

(3)　階層型ブルームフィルタを用いた有限オートマトンの暗号化と全文検索アルゴリズムの開発：文字列パターンの検索では有限オートマトンは非常に有用である。本課題では、階層型ブルームフィルタを用いた有限オートマトンの暗号化法を設計し、それを用いた暗号化全文検索アルゴリズムを開発する。応用として、部分文字列の一致まで検索可能なシステムを構築する。

３．研究の方法

(1) 階層型ブルームフィルタの空間効率の改善と暗号化ブーリアン検索への拡張：

①　空間効率を改善するための可変長階層型ブルームフィルタの設計：階層型ブルームフィルタは、葉に相当する部分に各文書を配し、根に向かって順に文書を統合することによって文書の集合を木構造で管理する。このとき、ブルームフィルタは木構造の各階層に割り当てられ、暗号化されたキーワードが登録される。文書を統合していくため、根に向かうほど登録するキーワードの数が少なくなる可能性がある。したがって、この性質を利用するとブルームフィルタのサイズを削減できる。すでに、登録キーワードの数、ブルームフィルタのサイズ、ハッシュ関数の個数、誤り率に関する理論式が知られており、この理論式を用いることによって、登録キーワード数に応じて自動的に最適なブルームフィルタのサイズを決定できる。本課題では、登録キーワード数によってサイズを変える「可変長階層型ブルームフィルタ」を設計し、固定長の階層型ブルームフィルタとの性能比較を行う。さらに、偽陽性の発生確率を考慮した、より詳細な解析を行う。

②　暗号化ブーリアン検索への拡張：階層型ブルームフィルタは、正解の候補を絞り込むことができればできるほど高速に動作する。そのため、検索に複数キーワードの論理式が使えると非常に有用である。このような暗号化ブーリアン検索への拡張は、入力される論理式を構文解析し、各階層のブルームフィルタのチェック時に、構文解析した論理式を適用することで実現できる。本課題では、暗号化ブーリアン検索に向け検索アルゴリズムを拡張し、性能を評価する。

(2)　階層型ブルームフィルタによるキーワードに基づく暗号化索引の開発：従来法の暗号化索引はドキュメントに基づいて構成されている。そのため、ドキュメント数未満の時間(sub-linear time)で検索できる。本課

題では、キーワード数の対数時間での検索を可能にするため、葉にキーワードを配置することによって、キーワードに基づいた暗号化索引を開発する。従来法と比較し、その性能を明らかにする。

(3)　階層型ブルームフィルタを用いた暗号化有限オートマトンの構成：有限オートマトンの各状態と入力記号のペアに対し、そのペアの遷移先の状態を ID とし、暗号化した（ペア，ID）を階層型ブルームフィルタへ登録することで有限オートマトンの暗号化を実現する構成法を設計する。この構成法では、ちょうど、遷移先の状態が文書 ID に相当することになる。

(4)　暗号化有限オートマトンを用いた暗号化全文検索アルゴリズムの開発：全文検索とは、文字列パターン（以下、パターン）とテキストに対し、テキスト中に出現するすべてのパターンを検索することである。ここでは、以下で述べるテキストの暗号化とパターンに対する暗号化有限オートマトンを用いた全文検索アルゴリズムを開発する。

①　テキストの暗号化：各文字の出現頻度を隠すため、テキスト中の各文字を，数文字まとめてブロック化して暗号化を行う。暗号化にあたっては，文字列を互いに重ならないように完全に分割してブロックを構成するわけでなく，前後のブロックは半分文字列を共有することでブロックを構成する。これにより，検索の安全性を高める。

②　テキストに対する暗号化有限オートマトンの作成：テキストをブロック化暗号で暗号化した暗号化テキストを作成する。テキストのすべての部分文字列を検索できるようにするため，暗号化テキストから，暗号化 DAWG(Directed Acyclic Word Graph)を作成する。DAWG は，文字列から構成される決定性有限オートマトンであり，元の文字列のすべての部分文字列を受理する。DAWG を暗号化索引とし，検索に用いる。

③　安全性の証明：DAWG に基づく検索可能暗号の安全性を，Curtmola らの手法（Searchable Symmetric Encryption : Improved definition and efficient construction, J. Comput. Security, 2011）を用いて証明する。

④　暗号化全文検索アルゴリズムを用いた検索システムの構築と評価：上記で開発する暗号化全文検索アルゴリズムを利用し、部分一致まで検索できるシステムを次のように開発する。文書全体を暗号化すると効率が悪いため、各文書に対し、その文書から取り出したキーワードを連接し、暗号化することで文書ごとに暗号化索引を作成する。

４．研究成果

本研究は，代表者が考案したブルームフィルタを用いた検索可能暗号をさらに発展させ，より柔軟で効率的な手法を開発してきた。本課題で得られた成果を下記にまとめる。

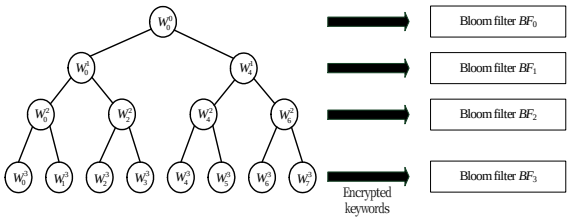

図 2　階層型ブルームフィルタ

(1)　キーワード型ブルームフィルタを用いた検索可能暗号：　従来法は，ドキュメントを基盤としたブルームフィルタを作成し，検索システムを構築した。本課題では，キーワードをベースとした新たな検索可能暗号を提案した。まず，非適応的安全性を満たす手法を開発し，その後，より安全性の高い手法として，適応的安全性を満たす手法を開発した。さらに，提案法は，従来法よりも単純なデータ構造を用いて暗号化索引を実現することにより，暗号化索引のサイズ，検索時間において効率的な手法となっている。また，ランダムオラクルを用いることにより，より効率的なシステムを構築できることを示した。提案法については，理論的に安全性を証明するだけでなく，検索時間及暗号化索引のサイズについて実験的に評価した。図 2 は，キーワードをベースとした階層型ブルームフィルタである。葉は 1 個のキーワードからなり，各ノードに対応するキーワード集合は，そのノードの子供のキーワード集合の和を取ったものとなっている。今後の課題としては，安全な更新機能を備えたシステムの開発があげられる。

(2)　部分文字列検索可能な検索可能暗号：従来法の多くは，あらかじめ文書からキーワードを抜き出し，暗号化索引を構成する。そのため，登録されているキーワードと完全一致する文字列しか探せない。キーワードの部分文字列まで検索できると，検索の幅が広がり，便利である。通常の部分文字列検索では，効率的な検索を実現するため，接尾辞オートマトン，接尾辞木，接尾辞配列，位置ヒープ木など多くのデータ構造が提案されている。暗号化検索の分野では，接尾辞木，位置ヒープ木を用いた手法が提案されている。本研究では，より安全で効率的な手法を目指し，暗号化 DAWG を用いることにより部分文字列まで検索可能な手法を提案した。文字列に対する DAWG とは，その文字列のすべての部分文字列を受理するコンパクトな決定有限オートマトンである。提案法は，暗号化 DAWG を階層型ブルームフィルタとして実現することにより，DAWG の構造まで隠すことができる安全性の高い手法である。DAWG を表す階層型ブルームフィルタは，図 2 で示した構造と同じ構造で構成される。提案法は，DAWG に加え，DAWG と密接に関係のある部分集合木と呼ばれる構造を利用することにより，さらに暗号化索引のサイズを縮小できることも示した。また，安全性に関しては，

提案法は非適応的安全性を満たすことを示した。今後の課題として，安全性の強化があげられる。すなわち，適応的安全性を満たす手法の開発である。

表 1　実験結果

一致数	3-key query		4-key query	
	queries	time	queries	time
1-100	5	0.13	7	0.13
101-1000	3	0.65	1	0.87
1001-10000	7	1.5	2	1.62
10001-100000	2	6.84	1	3.04
100001-400000	0	-	0	-
517431	1	47.79	0	-
平均		0.99		0.33

(3) ブーリアン検索可能な検索可能暗号：ブーリアン検索可能な手法について提案し，より詳細にその性能を評価した。ブーリアン検索とは，キーワードを論理演算子で結合したクエリに対し，そのクエリを満足する文書を探す検索である。本研究では，特に，AND 検索に対し，効率的に検索する手法を提案した。従来法は，個々のキーワードに対し，それを含むドキュメント ID を求め，それらの共通部分を取ることによって，AND 検索に対応したため，ドキュメント ID の集合の和集合のサイズに比例した時間となる。しかし，提案法は，共通集合のサイズに比例した時間で検索可能である。表 1 に実験結果を示す。3 個のキーワード，4 個のキーワードの AND 検索を実施した時の検索時間で，単位は秒である。「queries」の列は，ランダムに選んだクエリの内，実際に一致したドキュメントがあったクエリの数である。「一致数」の列はクエリに一致したドキュメント数を示しており，それが少ないほど高速であることが分かる。今後の課題としては，否定演算を含むクエリに対し，安全で効率的な手法の開発があげられる。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　2 件）

① H. Yamamoto, "Secure Automata-based Substring Search Scheme on Encrypted Data", Proc. of 11th International Workshop on Security (IWSEC 2016), LNCS 9836, pp.111--131 (2016) 査読有

② H. Yamamoto, T. Watanabe, and T. Miyazaki, "A Secure Conjunctive Keyword Search Using a Hierarchical Bloom Filter", Proc. of 5th International Conference on Informatics and Applications (ICIA 2016), pp.44--56(2016) 査読有

〔学会発表〕（計　6 件）

① 三好竜司，山本博章，"単純なデータ構造を用いた検索可能暗号"，コンピュータセキュリティシンポジウム (CSS2016), 2C3-2, pp.564--571 (2016), 学生論文賞，2016/10/12，秋田キャッスルホテル（秋田市）

② 北野峻平，山本博章，"否定検索可能な検索可能暗号について"，電子情報通信学会信越支部大会，1B-2 (2016)，2016/10/08，長岡技科大（長岡市）

③ 山本博章，宮嵜敬，"暗号データに対する部分文字列検索可能な安全な検索法"，暗号と情報セキュリティシンポジウム (SCIS2016), 2A1-1 (2016)，2016/01/20，ANA クラウンプラザホテル熊本ニュースカイ（熊本市）

④ 高野匠，山本博章，"キーワード型ブルームフィルタを用いた安全で効率的な検索法"，コンピュータセキュリティシンポジウム (CSS2015)，2015/10/23, 3D4-3, pp.1351--1358 (2015)，長崎ブリックホール（長崎市）

⑤ [招待講演]山本博章，"正規表現とその応用～有限オートマトンから文字列照合まで～"，電子情報通信学会技術報告，COMP2015-19，pp.21--26 (2015)，2015/09/01，信州大学工学部（長野市）

⑥ 渡邊尊司，山本博章，"階層型ブルームフィルタを用いた暗号化検索法の改良"，コンピュータセキュリティシンポジウム (CSS2014)，2E2-3，pp.543--550 (2014)，2014/10/23，札幌コンベンションセンター（札幌市）

６．研究組織

(1)研究代表者
山本　博章（YAMAMOTO Hiroaki）
信州大学・学術研究院工学系・教授
研究者番号：10182643

(3)連携研究者
宮嵜　敬（MIYAZAKI Takashi）
長野工業高等専門学校・　・教授
研究者番号：10141889