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

令和　４ 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：2017～2021
課題番号：１７Ｋ００１８３

研究課題名（和文）オートマトンに基づく部分文字列検索に対応した検索可能暗号に関する研究

研究課題名（英文）Study on Substring Symmetric Searchable Encryption based on Finite Automata

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）　　3,400,000円

研究成果の概要（和文）：暗号化されたデータを暗号化したまま検索する技術は検索可能暗号と呼ばれ、安全に情報を管理するうえで重要な技術となっている。本研究課題では以下の成果を上げた。
(1)キーワード検索に向けては、階層型ブルームフィルタを用い、データの更新（追加・削除）可能な動的検索可能暗号、検索結果の検証機能を埋め込んだ検証可能動的検索可能暗号を開発した。(2)部分文字列検索可能暗号に関しては、拡張型のDAWG(directed acyclic word graph)やファクターオラクルを用いた安全で効率的な手法を設計した。(3)正規表現検索については、有限オートマトンの構造を漏らさずに検索する手法を提案した。

研究成果の学術的意義や社会的意義
クラウドサービスの発展により、ユーザが外部サーバにデータを保存し、利用する形態が拡大している。そのため、第3者だけでなく、サーバ管理者に対してもデータの内容を漏らさずに安全に利用できる仕組みの重要性が高まっている。情報化社会の発展において、企業等における機密情報や個人情報の漏洩は大きな妨げとなる。本研究成果は、外部サーバを利用した安全な検索を提供するもので、この分野の発展に寄与すると思われる。

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 obtained the following results.
(1) We showed an efficient and secure dynamic SSE scheme using a Bloom filter based on keywords. Furthermore, we improved the proposed SSE scheme to be a verifiable one.(2) We newly designed a augmented DAWG(directed acyclic word graph) and factor oracle and developed a new substring SSE which can search for all substrings contained in a given text.(3) For a regular expression matching problem, we showed a new SSE scheme.

研究分野：情報科学

キーワード：検索可能暗号　情報セキュリティ　情報検索　暗号プロトコル　プライバシー保護

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

情報通信技術の発展により、ユーザが外部サーバにデータを保存し、利用する形態が拡大している。そのため、第 3 者だけでなく、悪意のあるサーバ管理者に対してもデータを安全に利用できる仕組みの重要性が高まっている。これを受け、検索分野では、安全な検索法として、データを暗号化したまま検索する検索可能暗号の研究が活発に行われている。

一般に、検索可能暗号では、安全で効率的な検索を実現するため、暗号化索引を構成し、検索に用いる。したがって、検索システムの安全性、効率性、検索機能などは暗号化索引に大きく依存しており、いかに安全で効率処理可能な暗号化索引を設計するかが研究のポイントとなっている。また、基盤となる暗号方式には共通鍵暗号方式と公開鍵暗号方式があるが、本研究は共通鍵暗号方式に基づいた検索可能暗号を考える。

検索可能暗号における挑戦的課題の一つは、任意の部分文字列を安全かつ効率的に検索できるようにすることである。これに関し、2015 年に 2 つの手法が提案された。一つは、Chase と Shen(PET 2015)による接尾辞木を利用した手法、もう一つは、Strizov と Ray(IEEE IC2E 2015)による位置ヒープ木（position heap tree）を利用した手法である。これとは別に、申請者は、DAWG(Directed Acyclic Word Graph)と呼ばれるデータ構造を用いた新たな部分文字列検索可能暗号を開発した。DAWG は、決定性有限オートマトンの考えに基づき、文字列中に出現するすべての部分文字列をコンパクトに表現できるのが特徴である。申請者は、階層型ブルームフィルタを使って有限オートマトンを暗号化する手法を考案し、安全で効率的な部分文字列検索システムを実現した。しかしまだ、いくつかの課題が残されている。本研究では、申請者が提案した手法を改良し、より安全で効率的な部分文字列検索法を開発する。さらに、有限オートマトンを暗号化する手法を応用し、より高機能な検索として、正規表現に向けた検索可能暗号を実現する。

２．研究の目的

本研究課題では、次の点について明らかにする。

(1) 並列検索、動的データに向けた効率的検索可能暗号の開発

(2) 部分文字列検索可能暗号の安全性の強化と時間・空間効率の改善：従来の手法は、非適応的安全であることが示されている。本研究では、適応的安全性を満たすように手法を改良する。さらに、DAWG の性質を利用し、暗号化索引の時間・空間効率を改善する。

(3) 更新機能を備えた部分文字列検索可能暗号の開発：従来の手法は、データの更新（追加・削除）に対応していない。通常の DAWG に対しては、DAWG を動的に変化させるアルゴリズムが知られている。本研究では、そのアルゴリズムを暗号化データが扱えるように拡張することにより、更新機能を備えた部分文字列検索可能暗号を開発する。

(4) 正規表現に向けた検索可能暗号の開発：階層型ブルームフィルタによる有限オートマトンの暗号化を応用することにより、正規表現検索に向けた検索可能暗号を開発する。

３．研究の方法

研究目的を達成するため、以下の方法によって研究を進める。

(1) 並列化及び動的データに向けた検索可能暗号の改良：申請者は、キーワード検索に対し、階層型ブルームフィルタを用いた安全で効率的な手法を提案したが、これを並列化、動的データに向け改良する。並列化については、データを木構造で管理することで達成する。動的データに向けては更新可能なデータ構造を開発する。この時、データの追加に関しては前方秘匿性、削除に関しては後方秘匿性を満たす方法を考える。さらに、検索結果の検証機能を導入することにより、不正なサーバに対処できる手法の構築を目指す。

(2) 部分文字列検索可能暗号の構築：より安全で効率的な手法を開発するため、DAWG を改良し、拡張 DAWG を導入する。拡張 DAWG は決定性有限オートマトンで、そのすべての状態遷移は記号列による。さらにその記号列がすべて異なるという性質を持つ。拡張 DAWG をベースに暗号化索引を構成すると、従来法に比べ、情報漏洩の少ない、安全な索引を構成できる。本研究では拡張 DAWG を用いた手法を提案する。これに加え、さらに効率的な手法を目指すため、ファクターオラクルを用いた手法についても検討する。ファクターオラクルは DAWG と同様の性質を持ちかつそのサイズがより小さくなっている。欠点としては、検索結果に偽陽性が発生することである。

(3) 更新機能を備えた部分文字列検索可能暗号の開発：データの追加機能について考える。DAWG を用いた手法では、データの追加に対し暗号化索引の基礎となる DAWG を更新しなければならない。しかし、データが暗号化されているため、DAWG の更新作業は難しい。本研究では更新作業がより簡便なファクターオラクルをベースとした手法に対し、更新可能な部分文字列検索可能暗号を検討する。

(4) 正規表現検索に向けた検索可能暗号：正規表現を有限オートマトンに変換して検索を行う。したがって、有限オートマトンの暗号化とテキストの暗号化が必要となる。

(ア)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。階層型ブルームフィルタへの登録時に乱数を利用することにより、同じ正規表現でも異なるパターンでブルームフィルタを構成することができる。

(イ)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、k 個の文字単位(k-gram)で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、前後のブロックは文字列を共有することでブロックを構成する。

４．研究成果

本研究課題で得られた成果を以下に示す。

(1) 並列化および動的データに向けた省スペースな検索可能暗号の開発：並列化および動的データに向け、階層型ブルームフィルタと転置索引を組み合わせた手法の改良を行った。並列化に関しては、索引の構造をリストから木構造に変えることにより、並列検索を可能とした。動的データとは、データの更新（追加、削除）が行われるデータである。動的データに対する検索可能の安全性に関し、前方秘匿性および後方秘匿性が考えられているが、提案法はこれら両方を満たすように従来法の改良を行った。さらに、メッセージ認証の機能を活用し、検索結果の検証が可能な機能を備えた方式を設計した。

(2) 拡張型 DAWG を用いた部分文字列検索可能暗号の開発：文字列に対する DAWG(directed acyclic word graph)とは、その文字列のすべての部分文字列を受理する決定性有限オートマトンである。申請者が開発した従来の DAWG を用いた部分文字列検索可能暗号は、構造が複雑であり、弱い安全性しか満たしていない。より安全性の高い手法を開発するため、本研究では、まず DAWG をその状態遷移の記号がすべて異なるように改良した拡張型 DAWG を開発し、その基本的性質を明らかにした。さらに、それを用いた新たな部分文字列検索可能暗号を開発した。状態遷移の記号がすべて異なることにより、ランダム性の高い暗号化索引を作成できる。したがって、拡張型 DAWG を用いた手法は、安全性が高く、その暗号化索引は単純な構造をしている。本研究課題では、拡張型 DAWG を用いた部分文字列検索暗号として、次の 2 つの方式を設計した。

(ア)拡張型 DAWG の遷移ラベルだけを利用した方式：拡張型 DAWG は状態遷移に記号列が割り当てられ、かつすべての記号列が異なる。そこで、状態遷移につけられた記号列だけを使って検索するための手法を開発した。本手法は、索引サイズ、検索時間に関して効率的であるが、DAWG の模倣にラベルだけを使っているため検索結果に偽陽性が発生する。ただし、偽陽性は簡単に識別できる。本手法は適応的安全性を満たす。

(イ)拡張型 DAWG の状態遷移を利用した方式：状態遷移のラベルだけでなく、状態遷移自身の情報をランダム化して格納し、それを使って DAWG を模倣する。偽陽性の発生を防ぐことができるが、暗号化索引に状態遷移の情報も格納するため、索引サイズは大きくなる。さらに、格納する情報が多くなるため、遷移ラベルだけと比べて、安全性の強度が落ちる可能性がある。

(3) ファクターオラクルを用いた部分文字列検索可能暗号の開発：申請者は、拡張型 DAWG を用いた部分文字列検索暗号について、その暗号化索引をより省スペースにするため、ファクターオラクルを導入した手法を設計した。ファクターオラクルは、DAWG と同様に文字列から作られる有限オートマトンで、文字列のすべての部分文字列を受理する。これは、DAWG よりサイズが小さくなるが、欠点として、認識において偽陽性が発生する。本研究では、ファクターオラクルを用い、検索での処理を少し増やすことで、索引サイズを小さくする手法を提案した。提案法は、DAWG のように拡張型ファクターオラクルがベースとなっている。本研究ではさらに、この手法を動的データに対応できるように改良した。

(4) 正規表現に向けた検索可能暗号：正規表現検索では、正規表現を有限オートマトンに変換して検索を行う。申請者はこれを暗号化して行う新たな手法を提案した。提案法は、文字列を k-gram に分割して行う。実際、提案法は次の（ア）、（イ）で示す暗号化 DFA と暗号化文字列を使って検索する。加えて、正規表現の暗号化検索の基礎となる有限オートマトン、検索アルゴリズムについて新たなアルゴリズムを開発した。

(ア)正規表現から k-gram に対応した暗号化有限オートマトンの作成：まず、正規表現から決定性有限オートマトン（DFA）を作成し、それを k-gram 対応 DFA に変換する。さらに、疑似ランダム関数と乱数を用いて k-gram 対応 DFA を暗号化することにより、同じ正規表現でも異なる暗号化 DFA を生成するようにする。暗号化 DFA はブルームフィルタに格納される。

(イ)テキストから検索用暗号化文字列の作成：暗号化 DFA を使って検索できるようにするため、テキストを k-gram を単位とするブロック文字列に変換し、各ブロックを疑似ランダム関数で暗号化することで検索用の暗号化文字列を作成する。

５．主な発表論文等

〔雑誌論文〕　計4件（うち査読付論文　4件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Hiroaki Yamamoto, Hiroshi Fujiwara	４．巻 104-D
２．論文標題 A New Finite Automata Construcion Using a Prefix and a Suffix of Regular Expressions	５．発行年 2021年
３．雑誌名 IEICE Trans. Inf. & Syst.	６．最初と最後の頁 381-388
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transinf.2020FCP0010	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Yamamoto Hiroaki、Wachi Yoshihiro、Fujiwara Hiroshi	４．巻 11821
２．論文標題 Space-Efficient and Secure Substring Searchable Symmetric Encryption Using an Improved DAWG	５．発行年 2019年
３．雑誌名 Lecture Notes in Computer Science	６．最初と最後の頁 130～148
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-31919-9_8	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Yamamoto Hiroaki	４．巻 143
２．論文標題 A faster algorithm for finding shortest substring matches of a regular expression	５．発行年 2019年
３．雑誌名 Information Processing Letters	６．最初と最後の頁 56～60
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.ipl.2018.12.001	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Miyoshi Ryuji、Yamamoto Hiroaki、Fujiwara Hiroshi、Miyazaki Takashi	４．巻 10674
２．論文標題 Practical and Secure Searchable Symmetric Encryption with a Small Index	５．発行年 2017年
３．雑誌名 Lecture Notes in Computer Science	６．最初と最後の頁 53～69
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-319-70290-2_4	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計16件（うち招待講演　0件／うち国際学会　0件）

1．発表者名
伊藤天啓、山本博章、藤原洋志

2．発表標題
正規表現に対する検索可能暗号の改良

3．学会等名
SCIS2021

4．発表年
2021年

1．発表者名
藤村享平、山本博章、藤原洋志

2．発表標題
動的データに向けた部分文字列検索可能暗号

3．学会等名
SCIS2021

4．発表年
2021年

1．発表者名
小田亮輔、山本博章、藤原洋志

2．発表標題
DAWGに基づいた部分文字列検索可能暗号の改善

3．学会等名
情報処理学会全国大会

4．発表年
2021年

1．発表者名
小澤響平、山本博章、藤原洋志

2．発表標題
検証可能な機能に向けた検索可能暗号

3．学会等名
情報処理学会全国大会

4．発表年
2021年

１．発表者名
藤村享平、山本博章、藤原洋志

２．発表標題
ファクターオラクルを用いた部分文字列検索可能暗号

３．学会等名
SCIS 2020

４．発表年
2020年

１．発表者名
伊藤天啓、山本博章、藤原洋志

２．発表標題
正規表現検索に向けた検索可能暗号の提案

３．学会等名
SCIS 2020

４．発表年
2020年

１．発表者名
大井恒平，和智吉弘、山本博章，藤原洋志

２．発表標題
ファクターオラクルの拡張と文字列照合問題への応用

３．学会等名
冬のLAシンポジウム

４．発表年
2019年

１．発表者名
三好竜司，山本博章

２．発表標題
Backward安全に向けた検索可能暗号の改良

３．学会等名
SCIS2019

４．発表年
2019年

１．発表者名
三好竜司，山本博章

２．発表標題
並列処理かつ動的データに向けた検索可能暗号の改良

３．学会等名
CSS2018

４．発表年
2018年

１．発表者名
大井恒平，山本博章，藤原洋志

２．発表標題
ファクターオラクルの拡張と実験的評価

３．学会等名
情報科学技術フォーラム(FIT2018)

４．発表年
2018年

１．発表者名
三好竜司，山本博章

２．発表標題
並列処理に向けた検索可能暗号の改良

３．学会等名
平成30年度電子情報通信学会信越支部大会

４．発表年
2018年

１．発表者名
佐藤明幸，増田康行、山本博章，宮崎敬

２．発表標題
GPUに向けた並列マルチパターン照合アルゴリズム

３．学会等名
平成30年度電子情報通信学会信越支部大会

４．発表年
2018年

１．発表者名
大井恒平，山本博章，藤原洋志

２．発表標題
ファクターオラクルの拡張

３．学会等名
夏のLAシンポジウム

４．発表年
2018年

１．発表者名
和智吉弘，山本博章，藤原洋志，宮嵜敬

２．発表標題
オートマトンを用いた部分文字列検索可能暗号

３．学会等名
SCIS2018

４．発表年
2018年

１．発表者名
三好竜司，山本博章，藤原洋志

２．発表標題
省スペースに向けた検索可能暗号の改良

３．学会等名
電子情報通信学会信越支部大会

４．発表年
2017年

１．発表者名
大井恒平，山本博章，藤原洋志

２．発表標題
拡張位置ヒープ木を用いた文字列検索

３．学会等名
電子情報通信学会信越支部大会

４．発表年
2017年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
連携研究者	藤原　洋志 (Fujiwara Hiroshi) (80434893)	信州大学・工学部・准教授 (13601)	
連携研究者	宮嵜　敬 (Miyazaki Takashi) (10141889)	長野工業高等専門学校・電気電子工学科・嘱託教授 (53601)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関