# 3-2　安全なデータ利活用技術
## *3-2　Secure Data Utilization Technologies*
# 3-2-1　様々な検索可能暗号
## *3-2-1　Various Searchable Encryption Schemes*

小川 一人
OGAWA Kazuto

今日では、クラウドなどの外部インフラが盛んに使用されるようになってきている。外部のデータベースを利用することもその一つである。個人や個別の企業・団体で所有する PC やサーバにデータベースを置く場合とは異なり、データの保護の観点から、考慮すべき課題が多くなる。

この新たな課題に対処するために開発されたのが検索可能暗号である。検索可能暗号とは、サーバに保存されている暗号化データに対して、クライアントが検索したいデータだけでなく、検索キーワードの内容をも秘匿しながら、当該キーワードを含む暗号化データのみを効率的に検索できる暗号技術である。

本稿では、このような外部データベースを利用する場合であっても、データベースの内容を安全に秘匿する方法の一つである検索可能暗号についての我々の取組を紹介する。

Currently, public infrastructure, such as cloud service, is increasingly utilized, with public database systems being a notable feature. The security considerations for these public systems are markedly different from those associated with private databases hosted on personal computers or corporate servers. Users of public databases must address additional privacy concerns.

To mitigate these concerns, searchable encryption schemes have been developed. These schemes ensure the security of data stored in the database, protect the confidentiality of query keywords, and enable efficient searching of encrypted data using these keywords. This paper discusses our progress regarding searchable encryption schemes designed for securing data in public database systems.

## 1　まえがき

情報の電子化が進むにつれて、個人や企業などの団体が所有し、処理する電子化された情報（以下、データ）の量は増々多くなっている。このため、データを伝達するための通信の大容量化が要求されるだけではなく、データを処理するために処理能力が高いサーバや、データを保存するための大容量のストレージなどが必要となってきた。

個人や団体で持つ PC やサーバの処理能力や、データベース（DB）のサイズでは、これらの流れに対応できなくなってきた部分がある。外部クラウドのサーバや DB はこのような需要に適合して発展してきている。

個人でデータを保管する場合、自ら所有する PC 等に保管するのであれば、平文（ひらぶん）（暗号化などを行わずそのままの文章）で保管したとしても、それほど大きな問題にはならなかった。ただし、昨今では、個人の PC に対しても PC ウイルスなどの被害により、個人データが漏えいすることも考えられる。このため、平文でデータを保管することはお勧めできない。企業などの団体の場合で、団体の構成メンバ全員がデータを閲覧するような場合は、ヒューマンエラーによる情報漏えいの危険性は増すが、基本的には個人と同じ状況であり、平文で保管しても大きな問題にはならなかった。ただし、企業のように職群、職位があって、それらの属性に応じて閲覧／編集できるデータに差がある場合は、平文で保管することはお勧めできない。

さらに、各自が所有する PC やサーバに保管するのではなく、外部クラウドの DB に保管となると、個人や団体に関連する以外の第三者、例えばクラウドを管

理する管理者等がデータにアクセスすることも考慮しなくてはならない。

本稿では、これらの状況に鑑み、クラウドの DB を安全に利用する基盤技術の一つである、検索可能暗号について紹介する。**2** において検索可能暗号とはどのようなものであり、特に外部クラウドの DB を利用する場合、どのような留意が必要であるかを紹介する。次に **3** において、我々が取り組んできた検索可能暗号について紹介する。さらに、**4** において今後の課題について述べる。

## 2　検索可能暗号とは

検索可能暗号とは、主に、DB のデータの暗号化に使われる暗号方式である。DB では、データを保管するとともに、保管されたデータの検索・閲覧・追加・削除・編集などが可能である。検索可能暗号では、外部クラウドの DB を利用したとしても、これらの処理を安全に実行する機能が付与されている。

**1** で述べたように、データが DB に平文で保管され、データに個人情報や、団体の機密情報などが保管されている場合、データの漏えいは個人情報や機密情報の漏えいに直結する。このため、データは暗号化して保管することが一般的である。

ただし、この暗号化方法にもいろいろな種類がある。シンプルな方法は、DB ソフトや、外部クラウドが提供する暗号方式と暗号鍵を用いて暗号化する方法である。利用者の利便性を考えると、最も安易な方法である。DB へのアクセスを高速にするためには、共通鍵暗号方式を使って、暗号化しておくことが、最も利便性が高い。データの検索・閲覧・追加・削除・編集を行う場合は、クラウド側で平文に戻し、該当する処理を行い、結果を DB に反映させ、利用者に伝える方式である。この方式では、外部クラウドの管理者には、暗号方式や暗号鍵が既知となり、DB 内のデータが管理者には閲覧でき、自由に扱うことが可能となる。しかし、データをクラウドに保管する利用者は、DB の管理者であってもその保管したデータは知られたくない場合が多い。

最も簡単な対策は、個人もしくは団体が独自に暗号方式を設定し、暗号鍵を持ち、データはクラウドに保管するが、暗号化･復号については、独自の PC やサーバで実施することである。これは、クラウドのストレージ部分だけを利用する方式であり、**1** にも述べたが、データ量が少ない場合はこの方式で問題はない。データ量が多大になると、その処理に時間がかかることになる。したがって、多大なデータをクラウドの DB で保管する場合は、その処理もクラウドに委託したほうが良い場合が多い。

上述の検索可能暗号は、このような課題を解決できる要素技術である。検索可能暗号では、クライアントが生成する鍵を用いて、保管するデータを暗号化することで、たとえクラウドの管理者であったとしても、データの内容を知ることができない構造になっている。さらに、データの検索では、キーワードの平文による検索ではなく、平文のキーワードからハッシュ関数等により変換された値（トラップドア）を使用して検索を行う。逆に述べると、平文のキーワードがわからないトラップドアが暗号化されたデータとともに保管されており、そのトラップドアを用いて検索が実施される。

## 3　NICT における検索可能暗号の研究開発

**1**、**2** においては、検索可能暗号と一言で述べているが、機能、必要とされる安全性の違いに応じて、様々な検索可能暗号が存在する。

まず、大きな 2 つの分類として、基本的な構造は同じであるが暗号化の処理に公開鍵暗号を使用する公開鍵暗号ベースの検索可能暗号（PEKS : Public-key Encryption with Keyword Search）と共通鍵暗号を使用する共通鍵暗号ベースの検索可能暗号（SSE : Searchable Symmetric Encryption）が存在する。PEKS は Boneh らによって提案された [1] アルゴリズムであるが、公開鍵を用いているため、任意のクライアントがサーバに暗号化データを保管することが可能となる。SSE は Song らにより概念が提案され [2]、Curtmola らによりモデルが定式化され [3]、秘密鍵を知るクライアントのみが暗号化データの格納をする方式であるが、処理は高速である。

機能の違いとしては、暗号鍵や DB に保管された暗号化データが更新可能な動的検索可能暗号方式、多数の利用者がいる場合を想定してデータへのアクセス制御を付与した放送型検索可能暗号方式などがある。

そして、必要とされる安全性の違いとしては、データが DB に追加された時、それ以前の検索処理からは追加されたデータに関する情報が漏えいしないことを定義したフォワード安全性、逆に、データが DB から削除された後、それ以後の検索処理から削除されたデータの情報が漏えいしないことを定義したバックワード安全性等がある。

本章では、これらの検索可能暗号の中で、セキュリティ基盤研究室が取り組んできた動的検索可能暗号と放送型検索可能暗号の研究を紹介する。さらに、検索可能暗号の安全性に関する再評価、安全性の強化、実装などの取組を紹介する。

### 3.1　動的検索可能暗号

本研究は、動的な SSE の研究である。動的検索可能暗号（DSSE：Dynamic SSE）とは、任意のタイミングで、暗号化データ、キーワード（キーワードに関連するトラップドア）を DB に登録可能な SSE タイプの検索可能暗号方式である。そして、DSSE では上述したフォワード安全性が不可欠となる。具体的に述べると、データの登録・追加を行う場合は、過去に行われた操作から、新たに追加されたデータが類推できないようにすることが、情報漏えいの観点から重要となる。定義としては、過去に検索されたキーワードが、新たに登録された暗号化データ、キーワードの中に含まれているかどうかが判別できない、ということになる。DSSE では、任意のタイミングでのデータ登録・追加を許容するため、この安全性は必須となる。

開発した具体的アルゴリズムを以下に示す [4]。

---

以下では、$\pi$を疑似ランダム置換の集合とする。

準備：最初にクライアントが秘密鍵$k$を選択する。簡単のため、ここでは$k$をデータの暗号化にも用いるとする。

データ保存：識別子$id$を持つファイル$f_{id}$に含まれるキーワードの集合$W_{id}$に対し、クライアントは全ての$\omega \in W_{id}$に対して$\pi_k(\omega, id)$を計算、$f_{id}$を$k$で暗号化し、$id$、$\pi_k(\omega, id)$、暗号文$cid$をサーバに送る。サーバはアドレス$\pi_k(\omega, id)$に$(id; cid)$を保存する。

データ削除：クライアントは削除対象データの$id$をサーバに送付、サーバは$(id; cid)$を削除する。

データ検索：あるキーワード$\omega$を含むデータを取得したい場合、クライアントは保持している$id$全てに対しトラップドア$\pi_k(\omega, id)$を計算、サーバに送付する。サーバはアドレス$\pi_k(\omega, id)$に値$(id; cid)$が保存されていれば$cid$をクライアントに返す。クライアントは$k$を用いて$cid$を復号し、データ$f_{id}$を得る。

---

クライアントは$k$以外に$id$だけ保持しており、検索時にはクライアントからサーバにトラップドアとして疑似乱数$\pi_k(\omega, id)$のみが送付される。さらに、現在の DB に保管されているデータに関する疑似乱数を生成しており、その後追加されるデータに関する情報は一切含まれていない。このことからフォワード安全性が達成されている。

DSSE については、このほかの研究にも取り組んできた。まず、DSSE の場合、“些細（ささい）な情報の漏えい”を許容することが一般的である。些細な情報とは、その漏えいした情報を悪用したとしても、データの重要な情報が得られない情報を指している。例えば、データに含まれるキーワードの数などである。キーワードの数から、暗号化されたデータの内容を得ることは困難と考えられる。困難ではあるが情報漏えいの可能性を下げるために、キーワードのダミーを付加する方法、さらに、キーワードの確率分布を固定し、その分布にあったキーワードを付加することで、多少の検索エラーが生じるが、ダミーの数を少なくする等のキーワードの数を隠す研究 [5] を行った。さらに、バックワード安全性を有する DSSE 方式についても取り組んでいる [6]。

### 3.2　外部匿名性を満たす放送型検索可能暗号

放送局では、放送に対する視聴者の反応をしばしば収集しているが、この情報を視聴者間で共有することを考える。放送局から放送されているコンテンツは多数あるため、放送局から提供される番組名や演者名等のいわゆるメタデータと、視聴者の方々の反応はデータベース化されていることが望ましい。さらに、有料放送では、コンテンツに対し視聴制御がかけられている。このため、有料放送の番組に対する場合は、このようなデータベースであっても、閲覧に対するアクセス制御をかけることは、自然な対応である。

放送型検索可能暗号（BEKS：Broadcast Encryption with Keyword Search）は、このようなデータベースのための検索可能暗号であり、Attrapadung らによって提案されて [7] 以来、研究開発が進んでいる。BEKS では、通常の検索可能暗号と同様に、暗号化されたデータと関連するトラップドアが DB に保存される。

上述の例では、複数の視聴者と放送局がクライアントであり、番組のメタデータと視聴者の反応が保管されるデータとなる。放送局としては、データの分析を行い、将来の番組に活用することを考えている。このため、視聴者についてもできるだけ細かい情報を望み、例えば、年齢や性別、趣味や職業などが含まれていることが望ましい。これに対し、放送局が所望する情報は、視聴者にとっての個人情報であり、開示を望まない視聴者もいる。

そして、既存の BEKS では、検索のアルゴリズムにクライアントの集合を入力とすることから、暗号化データにクライアント集合を含める必要があった。検索結果を得たクライアントは、それがどのクライアントのデータであるかを知ることになる。すなわち、データを預けたクライアントの匿名性（受信者匿名性）が確保されていなかった。

また、別の既存の BEKS では、受信者匿名性については、改善され、完全匿名性を達成されているが、暗号文長がクライアント数に対して、線形サイズ以下にならないことが示されていた。

そこで、セキュリティ基盤研究室は、受信者匿名性

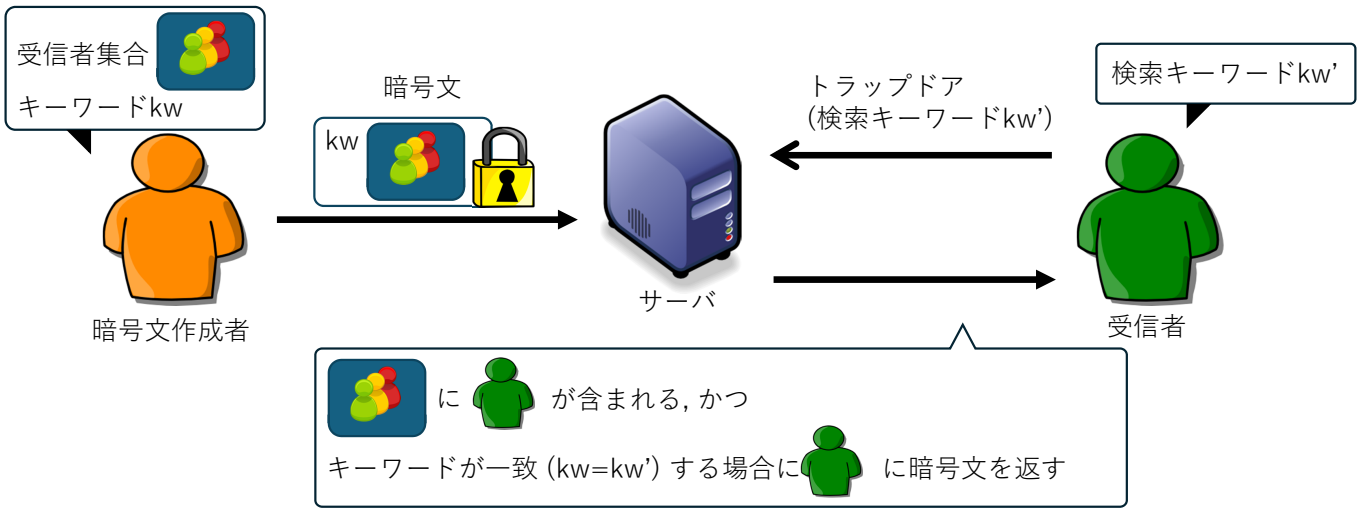


図 1　放送型検索可能暗号

を満たす放送型検索可能暗号であり、暗号文長がクライアント数に対して log サイズとなる一般的な構成方式を提案した [8]。具体的には、弱ロバスト性と呼ばれる 2 レベル階層型匿名 ID ベース暗号から、Fazio らによる受信者匿名性を満たす放送型暗号の一般的構成 [9] と、Boneh らによる PEKS の一般的構成 [10] を組み合わせた方法である。提案の構成方法では、文献 [9] の放送型暗号は暗号文長が受信者数に対して log サイズになる性質を保持し、暗号文サイズがクライアント数に対して log サイズとなった（図 1）。

### 3.3　検索可能暗号の性能評価

検索可能暗号に対し、攻撃方法に関する研究も盛んであり、種々の方法が提案されているが、それらはほぼすべて実験的な評価である。このため、攻撃者が本来持っていないはずの情報を持つことを仮定し、攻撃が行われている攻撃方法が多数存在する。

これらの攻撃が実際の運用においてどれほどの影響を及ぼすかを知るためには、パラメータを再設定し、非現実的な仮定を現実的な仮定に修正することが必要となる。そこで、パラメータを再設定、仮定を修正した上で、再実験を行い、検索可能暗号に対する攻撃の正確な性能評価を実施した [11]。実験では、攻撃により、暗号化データや、検索キーワードの復元率を求め、パラメータや仮定を修正する前後でどの程度変化するかを調査した。

結果の詳細については文献 [11] を参照していただきたいが、既存研究では、例えば、攻撃者が利用できるデータに現実とはかけ離れた設定が行われていた。このため不要な強度のセキュリティが要求されていた。このような、攻撃者が利用できるデータを現実的な設定に修正することで、復元率に大きな差が生じた。検索可能暗号を使用するためには、より現実的で、適切な設定が求められる結果となった。

### 3.4　セキュアチャネルフリー検索可能暗号

従来の PEKS では、もしトラップドアを入手することができれば誰であっても DB にそのトラップドアを送信することで、何らかの応答を DB から得ることが可能であった。言い方を変えれば、検索のテストアルゴリズムを実行できる状況になっていた。このため、検索のためのトラップドアを DB に送るためには安全な通信路（セキュアチャネル）が必要であった。

セキュアチャネルを必要としない方式として、セキュアフリー PEKS（SCF-PEKS: Secure-Channel Free PEKS）がある。SCF-PEKS では、DB を管理するサーバとクライアントの両方が公開鍵と秘密鍵のペアを持ち、キーワードをそれぞれの公開鍵を用いて暗号化して、検索するために利用する方式となっている。従来、トラップドアはクライアントの公開鍵だけを利用して作成されていたが、この方式によるとテストアルゴリズムの実行には、サーバの秘密鍵を必要とするため実行困難となる。

セキュリティ基盤研究室の研究 [12] では、SCF-PEKS の安全性定義をより安全な定義に変更した。例えば、アルゴリズムの一貫性を担保するための攻撃者の条件として、実際にテストアルゴリズムに入力するトラップドアとは異なるトラップドアであれば、任意のト

ラップドアを取得できるとした。これは、セキュアチャネルが確保されていない状況を安全性に加味したものである。また、サーバが、トラップドアの一貫性を壊すことができないことを示すため、攻撃者はサーバの秘密鍵を取得できるとした。これらの条件の上で、アルゴリズムの一貫性を担保した。さらに、キーワードの安全性については、攻撃者には、サーバの秘密鍵もしくは、クライアントの秘密鍵を取得できるとした上での安全性を示した。さらに、既存の暗号要素技術である、匿名 ID ベース暗号、タグベース暗号、ワンタイム署名などを基に、SCF-PKES を一般的に構成できることを示した。

### 3.5　応用：検索可能暗号を用いた暗号化ストレージ・チャットシステムの実装評価

検索可能暗号をどのように用いてセキュアシステムを構成すればよいのかは自明ではない。そこで、本研究では [13]、渡邉らによって提案された DSSE [4] を用いて、ストレージ及びチャットシステムを構築し、その実装評価を行った。さらに、実際のサービスとして使用する際の安全に対する課題抽出を行った。なお、一般の利用者に対し、検索可能暗号の使用を意識させない作りがサービスとして求められる。一方で、どのようなデータがどのように暗号化され、外部にどのように保存されているのかを一般利用者にも容易に説明可能であることが望ましい。そこで、暗号化されたデータや使用する秘密鍵が閲覧可能なコンシェルジュ機能を実装することで、説明容易性を向上させる試みも実施した。

実装方式について、模式図を図 2 に記載する。使用したサーバの仕様等の詳細は文献 [13] を参照いただきたい。

この実装評価においては、認証サーバとアプリケーションサーバの 2 つのサーバを使用した。2 つのサーバは結託しないことを前提とし、認証サーバに秘密鍵を保存し、クライアントの認証時に秘密鍵を取り出す方式としている。エンドツーエンドの暗号化の観点からは、クライアントだけが秘密鍵を保持するほうがよい。特に、シングルユーザ設定であるストレージシステムであれば、この設定は可能である。ただし、サーバによる鍵管理ではなく、クライアント本人による秘密鍵管理となることに注意が必要となる。

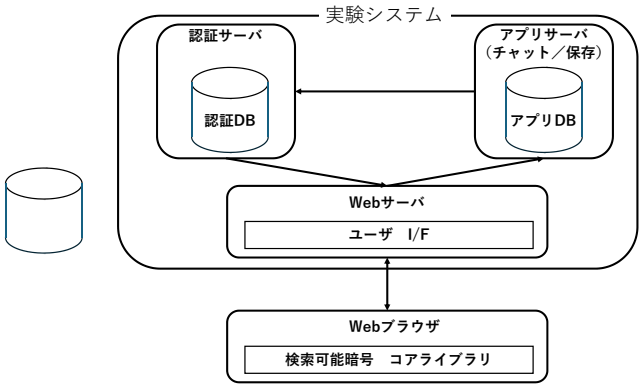


図 2　実験システム

一方、チャットシステムは複数クライアントが存在するマルチユーザ設定となる。このため、クライアント間の鍵交換が必要となる。結果として、公開鍵ペアを各クライアントが作成する必要が生じる。サーバが鍵管理をする場合は、クライアントによる公開鍵ペアを逐次作成するなどの必要はない。クライアント管理にした場合、公開鍵ペアを逐次作成しなければならないが、その保証の問題が生じる。すなわち、公開鍵所有者の正当性を確認しなければならず、別途 PKI（公開鍵インフラストラクチャ）を用いて証明書を発行するなど、公開鍵の妥当性検証、安全性の確保に留意しなければならない。

さらに、キーワード作成には形態素解析を利用した。使用する形態素解析ライブラリによっては、半角、全角の扱いが異なる場合や、ストップワード（“a”, “the”, “is” 等）の扱いに差がある。この差は、トラップドアの作成に影響を与え、ひいては、検索時間などに影響を与えることになった。形態素解析を利用する・しない、にかかわらず、キーワード集合の適切な定義は検索可能暗号を使う上で、大きな課題となる。

また、この実験では、暗号データを入力順（時系列）にそのまま並べて DB 化している。暗号化データ、暗号化キーワードの入力であるため、並べる順番を定めることは困難な部分はあるが、検索に適した順番にソートすることが、検索時間短縮には重要である。

## 4　むすび

個人の PC や、個々の団体のサーバの能力だけでは処理が遅い場合があり、保管するデータの量が多大になる場合、クラウドに置かれた CPU やストレージを利用することが最もコストパーフォーマンスの優れた方法となることは、容易に考えられる。そして、クラウド利用においてもデータのプライバシーを担保することは重要となる。我々も含め世界の多くの研究者による検索可能暗号の研究開発は、これらの社会の進展に貢献する成果となっている。そして、検索可能暗号はまだ発展途上である。すなわち、“些細な情報の漏えい” を許容している部分がある。この些細な情報が悪用される可能性が完全に排除されているわけではない。許容している些細な情報を削除する若しくは、これらが悪用される可能性が無い情報だけにすることが、今

後の課題となる。

## 【参考文献】


1 D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public Key Encryption with Keyword Search,” Eurocrypt 2004, pp.506–522, 2004.
2 D. X. Song, D. Wagner, and A. Perrig, “Practical Techniques for Searches on Encrypted Data,” IEEE S&P 2000, pp.44–55, 2000.
3 R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,” ACM CCS 2006, pp.79–88, 2006.
4 渡邉 洋平，岩本 貢，太田 和夫，“効率的でフォワード安全な動的検索可能暗号，” SCIS2019，3 C1-3，2019.
5 Y. Watanabe, K. Ohara, M. Iwamoto, and K. Ohta, “Efficient Dynamic Searchable Encryption with Forward Privacy under the Decent Leakage,” CODASPY2022, pp.312–323, 2022.
6 甘田 拓海，岩本 貢，渡邉 洋平，“効率的かつ安全な更新処理を備えた結果秘匿可能な検索可能暗号，” SCIS2023，3 A3-5，2023.
7 N. Attrapadung, J. Furukawa, and H. Imai, “Forward Secure and Searchable Broadcast Encryption with Short Ciphertexts and Private Keys,” Asiacrypt 2006, pp.161–177, 2006.
8 K. Emura, K. Kajita, and G. Ohtake, “Outsider-Anonymous Broadcast Encryption with Keyword Search: Generic Construction, CCA Security, and with Sublinear Ciphertexts,” IEICE Trans. Fundam. Electron. Commun. Comput. Sci.vol/106, no.3, pp.193–202, 2023.
9 N. Fazio and I. M. Perera, “Outsider-Anonymous Broadcast Encryption with Sublinear Ciphertexts,” PKC2012, pp.225–242, 2012.
10 D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public Key Encryption with Keyword Search,” Eurocrypt 2004, pp.506–522, 2004.
11 甘田 拓海，並木 拓海，岩本 貢，渡邉 洋平，“検索可能暗号に対する漏洩悪用攻撃の正確な性能評価に向けて，” SCIS2024，2 D4-3，2024.
12 T. Suzuki, K. Emura, and T. Ohigashi, “A Generic Construction of Integrated Secure-Channel Free PEKS and PKE,” ISPEC2018, pp.69–86, 2018.
13 江村 恵太，金森 祥子，野島 良，渡邉 洋平，“検索可能暗号を用いた暗号化ストレージ・チャットシステムの実装評価，” 信学技報，ISEC2021-5，2021.



**小川 一人**（おがわ かずと）
サイバーセキュリティ研究所
セキュリティ基盤研究室
上席研究員
博士（情報理工学）
高機能暗号