

## パーソナルデータを暗号化した状態で安全に匿名化する技術を開発 匿名化したパーソナルデータの利活用ニーズに対応

株式会社日立製作所(執行役社長兼 COO:東原 敏昭/以下、日立)は、このたび、パーソナルデータ\*1(個人に関する情報)を暗号化した状態で安全に匿名化(個人を特定できないように変換)する技術を開発しました。本技術によって、より安全にパーソナルデータの匿名化を行うことが可能になり、2015年9月の個人情報保護法の改正案の成立を受けて今後拡大が見込まれる、匿名化したパーソナルデータ(匿名加工情報)の利活用のニーズに対応していきます。

近年、モバイル端末やセンサー機器の急速な普及などにより、世の中で発生し、収集されるデータは増加と多様化を続けており、さまざまな分野でビッグデータの利活用が進められています。また、個人情報保護法の改正案が成立し、将来的に機器のデータだけでなく、個人が特定できないようにパーソナルデータを匿名化した匿名加工情報についても、第三者が利用できるようになります。これにより、例えば、移動履歴や購買履歴の匿名加工情報を高精度な市場調査に利用するなど、今後は、匿名加工情報の利活用が、大きく拡大することが見込まれています。

こうしたビッグデータの分析においては、計算処理能力をフレキシブルに利用できるクラウドの活用が広がっていますが、パーソナルデータのように機微なデータの取り扱いにおいては、より高い安全性が求められることから、クラウド上のデータを暗号化し、第三者が内容を解読できないようにした上で、暗号化した状態のまま検索や分析を行う技術の実用化が進んでいます。一方、個人が特定できないようにパーソナルデータを匿名化する技術としては、k-匿名化技術\*2 が注目されていますが、従来の技術では、暗号化したデータをそのまま匿名化することができず、一度暗号化を解除(復号)する必要があり、セキュリティ上の課題となっていました。

そこで、日立は、今後ニーズの高まるパーソナルデータの匿名化におけるセキュリティの向上を目的に、パーソナルデータを暗号化し、クラウド上では暗号化状態を維持しながら k-匿名化を実現する技術を開発しました。本技術の特長は以下の通りです。

### 1. 安全に暗号化データを集約する技術の開発

多くの k-匿名化技術では、値が異なるデータを同じデータに集約するツリー構造\*3を用い、下の階層から、上の階層へデータを集約します。具体的には、例えば、関東地方(10件)や東北地方(20件)といった階層(小区分)のデータを、東日本(30件)といった上の階層(中区分)のデータに集約する形で匿名化を行います。しかしながら、データを暗号化した場合、集約前の下の階層のデータが暗号化によって解読不能なため、このツリー構造が作成できませんでした。

日立は、暗号化状態を維持したままデータの比較を行う独自の検索可能暗号技術を用い、比較の結果、同じ値だと判定された暗号化データの数を集計し、この集計結果を用いてツリー構造を生成する技術を開発しました。このツリー構造は、暗号化データのうち、集計結果の件数が少ない方を下の階層に、件数が多い方を上の階層へ位置づけており、集約によってデータの情報量が減る影響を抑える工夫が施されています。

## 2. 実用的な処理速度と高い安全性を達成

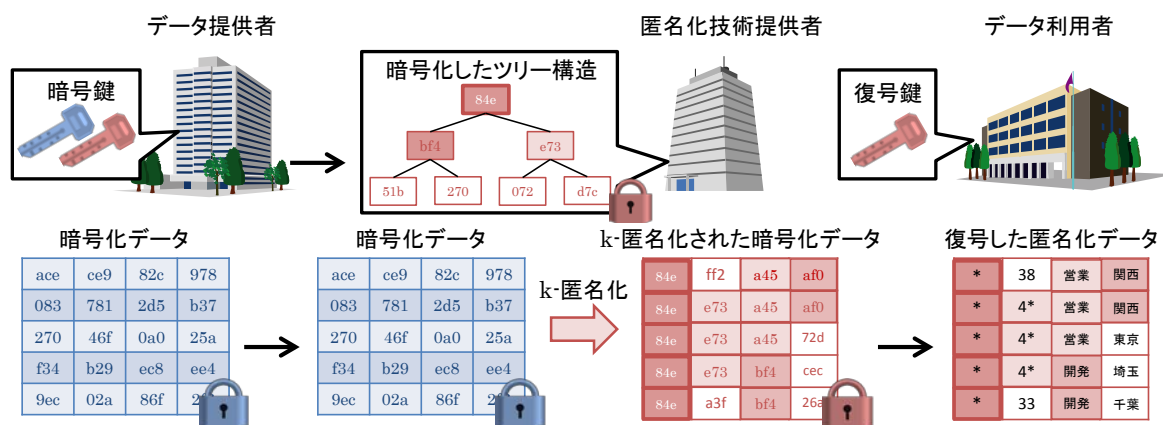
一般に、暗号化状態を維持したままのデータ処理は、暗号化しない場合と比べ、極端に処理速度が低下します。一方、日立がもつ検索可能暗号技術を用いることで、暗号化データ同士の比較処理を高速に行い、さらに、暗号化状態のデータ処理を必要最小限に留めることができます。これにより、データ処理におけるオーバーヘッドの増加を 30%<sup>\*4</sup> に抑え、実用的な処理速度を確保することに成功しました。

また、より高い安全性を確保するために、データを暗号化するための暗号鍵と、暗号化されたデータを匿名化するための暗号鍵は、それぞれ異なるものを用います。これにより、匿名化前の暗号化データが万一流出しても、暗号化を解除するための復号鍵はデータ提供者のみが保有するため、安全性を担保することができます。

日立は、パーソナルデータ利活用の拡大に向けて、開発した技術を用い、2018 年度中に今回開発した技術の実用化をめざします。

なお、本成果は、3 月 10 日、11 日に電気通信大学で開催される情報セキュリティ研究会にて発表予定です。

### ■今回開発した技術の概要イメージ図



- \*1 パーソナルデータ: 本ニュースリリースでは、個人情報保護法第2条第1項で規定された「生存個人の識別情報」よりも広く、位置情報や購買履歴などの個人識別性のない情報も含まれた「個人に関する情報」と定義します。
- \*2 k-匿名化技術: 特定の個人の識別を困難にするためのデータ加工方法の一つ。同じ属性を持つレコード(データベースの一行単位のデータの集まり)が k 件以上存在するようにデータを変換することで、個人が特定される確率を k 分の 1 以下に低減させる。
- \*3 ツリー構造: 親となるデータから子となる複数のデータに枝分かれし、さらに孫となる複数のデータに枝分かれするといった形で、階層が深くなるほど枝分かれしていくデータ構造の一種。
- \*4 測定結果は、k 値を 2 に設定し、3 万レコード、9 種類の属性に対し、暗号化されたツリー構造の作成から k-匿名化された暗号化データを作成するまでの、匿名化技術提供者(上記概要イメージ図に記載)の処理を測定したものの。

### ■本件に関するお問い合わせ先

株式会社日立製作所 研究開発グループ 技術統括センタ 情報企画部 [担当:湯本]

〒244-0817 神奈川県横浜市戸塚区吉田町 292 番地

TEL:050-3135-3409(直通)

以上

---

このニュースリリース記載の情報(製品価格、製品仕様、サービスの内容、発売日、お問い合わせ先、URL 等)は、発表日現在の情報です。予告なしに変更され、検索日と情報が異なる可能性もありますので、あらかじめご了承ください。

---