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　　７　日現在

機関番号：　１４４０１
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2016
課題番号：　２６３３００８５

研究課題名（和文）インタークラウド環境を用いたセンサーデータの分散解析手法の研究

研究課題名（英文）Distributed privacy preserving statistical computation for IoT in inter-cloud computing environment

研究代表者

中川　郁夫（Nakagawa, Ikuo）

大阪大学・サイバーメディアセンター・招へい准教授

研究者番号：７０６４７４３７

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：IoTにおける膨大な数のセンサーデータの処理に際し、プライバシ情報の漏洩リスクを抑えつつ、パブリッククラウドの計算機資源を有効に活用する秘匿分散統計解析手法の研究を行った。提案手法は複数クラウド環境に秘匿分散してデータを収集、分析することでデータ漏えいリスクを低減しつつ、正確な統計指標を得ることができる。
本研究では、同手法のプロトタイプ実装を行うとともに、実用化に向けて、DHTを用いた拡張性の実現や分散トランザクションの仕組みによる堅牢性の高いデータ処理の仕組みを提案した。また、複数のビジネスケースへの応用を検討し、スマートホームにおける応用例について試験実装しその有効性を確認した。

研究成果の概要（英文）：In this research, we proposed distributed privacy preserving statistical computation mechanism for processing a huge amount of IoT sensor data, where we reduce the risk of revealing privacy data while we would use scalable, elastic and cost effective computing power of public cloud. In the mechanism, IoT devices upload their data into multiple cloud computing environment with privacy preserving technique so that it reduce the risk of privacy data, while we would achieve accurate statistical indices.
We designed and implemented prototype system for the mechanism and proved its effectiveness. We also extended the mechanism to be scalable and resilient with DHT and distributed transaction technique.
We proposed to apply our mechanism for some business cases, such as smart home, healthcare or online voting. We designed and implemented the prototype for the case of smart home, as well.

研究分野：　Internet of Things

キーワード：　IoT　Intercloud　Statistical Computation

１．研究開始当初の背景

IoT (Internet of Things) の時代には膨大な数のセンサーや端末がネットに接続される。ネットにつながるセンサーの数は 2020 年までに数百億を超えるとの予想もある。これらのセンサーからは、例えば、位置情報、温湿度、音、個人の購買行動や行動履歴を含むあらゆる情報が収集され、気象、交通、防犯、医療・ヘルスケア、工業、農業はもちろん、ビジネスの現場においても実用化が進んでいる。

膨大な量のセンサーデータを蓄積・処理するには、安価 (cost effective) 、迅速 (elastic) 、拡張性 (scalability) を特徴とするパブリッククラウドの利用が有効である。一方、プライバシ情報が含まれる可能性のあるセンサーデータの処理においてパブリッククラウドを利用する場合には、例えば、特定機能に依存して、複数クラウドの有効活用が困難などの機能依存リスクや、第三者の計算資源からデータが漏洩するリスクなどが指摘された。

２．研究の目的

本研究では、膨大な数のセンサーから得られる大量データをパブリッククラウド上で処理する技術の研究を行った。本研究が提案する手法は、特に、以下の 2 点の特徴を有するデータ処理の仕組みを目的とした。

- これまで複数クラウドが協調してデータ処理をするモデルは確立されていなかったが、今回、複数の異なるクラウドが連携する分散型クラウドデータ処理モデルを提案する。

- 単一クラウドでは第三者リソースを利用する際の漏洩リスクが課題とされてきたが、今回、秘匿・分割したデータを複数クラウド上で分散処理する事で安全なデータ処理を可能にする。

提案手法は数億～数十億にも及ぶ多数のセンサーを入力として得られたデータを秘匿・分散した後に複数クラウドで保存・処理する。IoT における膨大なデータ処理に際して、パブリッククラウドの計算資源を有効に活用しつつ、データ漏えいリスクを押さえ、安全に解析を行うことを目指した。

３．研究の方法

本プロジェクトは次の 3 段階のステップで研究を推進した。
1. 理論的な研究と基本プロトタイプ実装
2. プロトタイプ実証と拡張性の実現
3. 実用化研究と実証実験

以下では、各ステップについての研究の方法についてまとめる。

3.1. 理論的な研究と基本プロトタイプ実装

センサーで取得されるデータをクラウド上で保存・処理する場合の課題について整理するとともに、複数の異なるクラウド上で処理する具体的な手法について検討・研究し、そのプロトタイプ実装を行った。課題の整理では、中川、樋地、菊池が参加する T クラウド研究会を中心に、センサーデータをクラウド上で保存・処理する際の、機能依存リスクやデータ漏洩リスクなどの課題の抽出を行った。

また、以下の 3 つの実施内容を含む、プロトタイプ実装を行い、動作確認を行った。
1. アーキテクチャ検討。センサー、クラウド、解析者の各ロールとデータの流れを定義。
2. アルゴリズムと API 設計。ロール間の受け渡し手順と API を定義。
3. 基本プロトタイプ実装。前述アルゴリズムと API に従い、各ロジックを実装。

3.2.プロトタイプ実証と拡張性の実現

前述のプロトタイプ実装を用いて実証実験を行い基本機能の確認を行った。実証実験では、擬似的なサービスモデルを想定した。擬似的なセンサー(PC やスマホ)からの入力データを複数クラウド上で処理するモデルでの実証実験を行った。なお、クラウドに相当する機能として、中川、下條、菊池が研究プロジェクト distcloud として推進している大学間連携によるインタークラウドでのデータ処理基盤を実証実験のフィールドの一部に用いた。同基盤は、複数の大学のクラウド環境を JGN-X や SINET4 などの超高速ネットワークで結んで構築されていた。また、商用パブリッククラウドなどの第三者の計算機リソースも利用し、複数の(所有者も運用者も)異なるクラウドからなるインタークラウド環境を構築した。

拡張性の実現は理論・アーキテクチャ・アルゴリズムの研究を実施した。同拡張は新たな仕組みとしてプロトタイプ実装の拡張機能として実装し、その有効性を検証した。

3.3. 実用化研究と実証実験

提案手法を具体的なサービスモデルに適用するための実用化研究を行った。本研究では、T クラウド研究会のメンバ企業らと協調し、対象となるビジネス領域やビジネスモデルを想定した。同研究会ではセンサーデータとクラウドが連携することにより、新しいビジネスの創発を目指していた。本研究では、同研究会と共同で、本手法の応用領域の可能性と実用化の検討を行った。

また、実用化検討を行ったビジネスモデルの 1 領域にフォーカスし、試験実装と有効性の検証を行った。

## ４．研究成果

### 4.1. 秘匿分散統計解析手法の提案

本プロジェクトでは、秘匿分散統計解析を行うためのアーキテクチャ及び計算アルゴリズムを提案し、理論的な裏付けを行った。同手法では、センサーで取得されるデータを m 個のランダムシェアに分割し、それを m 個の独立なクラウドプラットフォームにアップロードする。各クラウドでは独立に統計解析指標を計算するが、最終的に計算された値を用いて本来の正確な統計指標を得ることが可能である。以下は合計と平均を計算する例である。

$$Sum(x) = \Sigma x = \Sigma_i (\Sigma x_i)$$

N ： センサーの数
x ： 個々のセンサーから取得される値
$x_i$ ： x のランダムシェアの一部 (i = 1..m)

$$Ave(x) = Sum(x) / N$$

同様に $x^2$ を秘匿分散することで、2 乗和、分散、標準偏差を得る。さらに、複数要素への応用も容易であり、内積、共分散、相関係数の計算や、推定、検定、統計分析への応用も可能である。

また、本統計解析を実現するためのアーキテクチャ及びプラットフォームの設計とプロトタイプ実装を行った。本手法は、デバイスでセンサー値を秘匿化（ランダムシェア生成）し、m 個のクラウドにデータをアップロードするクライアント部、クラウド上でセンサー値を収集・計算するプラットフォーム部、及び、計算された値を統合・統計指標を求めるマネージャ部からなる。本研究では、同アーキテクチャの設計とプロトタイプ実装を行った。なお、クライアント部は透過的クラウド型のプログラミングが可能な Dripcast を IoT 環境に応用し、簡単で分かりやすいデバイスプログラミングを可能とした。

### 4.2. プロトタイプ実証と拡張性の実現

上記、計算アルゴリズムに基づいたプロトタイプ実装を用いて、本アーキテクチャの動作について検証を行った。プロトタイプ実証では、大学間インタークラウド環境 distcloud を用いて、複数の異なるクラウド環境に統計解析プラットフォームを実装、デバイスから収集した秘匿データを元に分散環境での統計指標の計算が可能であることを示した。

また、以下の計算アルゴリズム拡張を行った。

1. DHT（分散ハッシュテーブル）を応用したスケールアウト型の分散アーキテクチャの設計を行った（図 1 参照）
2. 耐障害性を考慮した分散トランザクションの仕組みを導入した

上記拡張は、秘匿分散統計解析を IoT 環境で実用化する上で必須の機能である。DHT 拡張は、IoT における膨大な数のセンサー値を収集することを可能にし、また、デバイスとの通信における不慮の事故、エラーなどが発生しても統計の数値計算に誤差が発生しない仕組みを提案した。

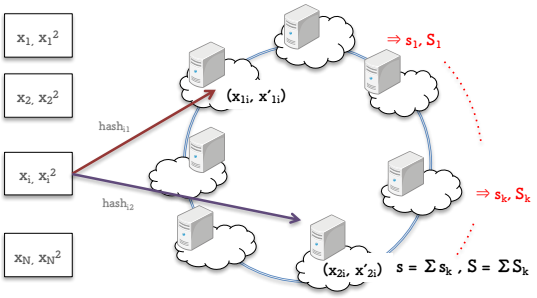


図 1

### 4.3. 実用化研究の成果

T クラウド研究会のビジネスケース研究プロジェクトと共同で、秘匿分散統計解析手法の応用領域について検討し、以下の 3 ケースにおける実装・実用化モデルについて検討した。

1. スマートホーム
   家電の電力使用量を収集し、街全体としての電力消費動向の分析を行う際に、本手法を応用するモデル。
2. ヘルスケア
   個人がヘルスケア情報を管理すると同時に、利用者全体での統計値を計算する際に本手法を応用するモデル
3. オンライン選挙
   選挙のデジタル化を見越して、投票者、及び投票先を秘匿したまま、投票結果を正しく得るために本手法を応用するモデル。

なお、上記のうち、特にスマートホームにおける応用については、実用化を想定したプロトタイプ実装を行い、その有効性を検証した。

## ５．主な発表論文等

〔雑誌論文〕（計 5 件）

I. Nakagawa, M. Hiji, Y. Kikuchi, M. Fukumoto, S. Shimojo: m-cloud - Distributed Statistical Computation Using Multiple Cloud Computers.
Proc. of IEEE COMPSAC, pp.301-305, Jul, 2014, Sweden,
DOI: 10.1109/COMPSACW.2014.53

I. Nakagawa, M. Hiji, et al.: Dripcast - Server-less Java Programming Framework for Billions of IoT Devices.
Proc of IEEE COMPSAC, pp. 186-191, Jul, 2014, Sweden,
DOI: 10.1109/COMPSACW.2014.35

I. Nakagawa, M. Hiji, Y. Kikuchi, M. Fukumoto, S. Shimojo: DHT extension of m-cloud - scalable and distributed privacy preserving statistical computation on public cloud.
Proc. of IEEE COMPSAC, pp. 682-683, Jul, 2015, Taiwan,
DOI: 10.1109/COMPSAC.2015.94

I. Nakagawa, M. Hiji, et al.: Dripcast - architecture and implementation of server-less Java Programming framework for billions of IoT devices,
IPSJ Journal Vol.23 (4), 2015, 458-464
DOI: 10.2197/ipsjjip.23.458

I. Nakagawa, M. Hiji, et al.: Design and Implementation of Global Reference and Indirect Method Invocation Mechanisms in the Dripcast.
Proc. Of IEEE COMPSAC, pp.338-343, Jun, 2016, Atlanta,
DOI: 10.1109/COMPSAC.2016.92

〔学会発表〕(計 7 件)

I. Nakagawa, M. Hiji, Y. Kikuchi, M. Fukumoto, S. Shimojo: Fault tolerant mechanism of m-cloud, distributed privacy preserving statistical computation on cloud, 電子情報通信学会 NS-IN 研究会, 信学技報 114(478), 267-272, 2015/3, Okinawa

中川郁夫, 樋地正浩, 菊池豊, 福本昌弘, 下條真司: 秘匿分散統計解析手法を応用したスマートホームシステムの設計と実装, ITRC/RICC Workshop 7th, 2016/2, Okinawa

中川郁夫, 樋地正浩, 菊池豊, 福本昌弘, 下條真司: 秘匿分散統計解析手法 ″m-cloud″ における分散トランザクションの設計と実装, 情報処理学会 第 32 回 IOT 研究会, 2016/3, Saga

中川郁夫, 樋地正浩, 菊池豊, 福本昌弘, 下條真司: インタークラウド環境での秘匿分散統計解析手法を応用したビジネスモデルの検討, ITRC/RICC, 2016/8, Sapporo,

I. Nakagawa, et al.: Cases - IoT/BigData changes your business, Cloud Expo 2016 West, 2016/11, Santa Clara

福本昌弘, 他: 秘密分散バックアップした医療データの部分復元システム, 電子情報通信学会 IA 研究会, 2016/12, Hiroshima

〔図書〕(計 0 件)

〔産業財産権〕

○出願状況 (計 1 件)

名称：データ秘匿型統計処理システム、統計処理結果提供サーバ装置及びデータ入力装置、並びに、これらのためのプログラム及び方法
発明者：中川 郁夫
権利者：中川 郁夫
種類：特許
番号：特願 2014-176590
出願年月日：2014/8/29
国内外の別：国内

○取得状況 (計 1 件)

名称：データ秘匿型統計処理システム、統計処理結果提供サーバ装置及びデータ入力装置、並びに、これらのためのプログラム及び方法
発明者：中川 郁夫
権利者：中川 郁夫
種類：特許
番号：特許第 5895080 号
取得年月日：2016/3/4
国内外の別： 国内

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
中川 郁夫 (NAKAGAWA, Ikuo)
大阪大学 サイバーメディセンター 招へい准教授
研究者番号：70647437

(2)研究分担者
下條真司 (SHIMOJO, Shinji)
大阪大学 サイバーメディアセンター 教授
研究者番号：00187478

(3)研究分担者
樋地正浩 (HIJI, Masahiro)
東北大学 経済学研究科 教授
研究者番号：40400212

(4)研究分担者
福本昌弘 (FUKUMOTO, Masahiro)
高知工科大学 情報学群 教授
研究者番号：70299387

(5)研究分担者
菊池豊 (KIKUCHI, Yutaka)
高知工科大学 地域連携機構 教授
研究者番号：80242288